The Plant Cell, Vol. 17, 1598–1611, May 2005, www.plantcell.org

Structural and Kinetic Basis for Substrate Selectivity in *Populus tremuloides* Sinapyl Alcohol Dehydrogenase

Erin K. Bomati[a,b] and Joseph P. Noel[a,b,1]

[a] Jack Skirball Chemical Biology and Proteomics Laboratory, Salk Institute for Biological Studies, La Jolla, California 92037
[b] Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92037

We describe the three-dimensional structure of sinapyl alcohol dehydrogenase (SAD) from *Populus tremuloides* (aspen), a member of the NADP(H)-dependent dehydrogenase family that catalyzes the last reductive step in the formation of monolignols. The active site topology revealed by the crystal structure substantiates kinetic results indicating that SAD maintains highest specificity for the substrate sinapaldehyde. We also report substantial substrate inhibition kinetics for the SAD-catalyzed reduction of hydroxycinnamaldehydes. Although SAD and classical cinnamyl alcohol dehydrogenases (CADs) catalyze the same reaction and share some sequence identity, the active site topology of SAD is strikingly different from that predicted for classical CADs. Kinetic analyses of wild-type SAD and several active site mutants demonstrate the complexity of defining determinants of substrate specificity in these enzymes. These results, along with a phylogenetic analysis, support the inclusion of SAD in a plant alcohol dehydrogenase subfamily that includes cinnamaldehyde and benzaldehyde dehydrogenases. We used the SAD three-dimensional structure to model several of these SAD-like enzymes, and although their active site topologies largely mirror that of SAD, we describe a correlation between substrate specificity and amino acid substitution patterns in their active sites. The SAD structure thus provides a framework for understanding substrate specificity in this family of enzymes and for engineering new enzyme specificities.

INTRODUCTION

In terrestrial vascular plants, lignification imparts structural support, hydrophobicity to vascular elements, and mechanical defense against biodegradation. Upon pathogen attack or wounding, lignification and/or deposition of lignin-like phenolics also occur as an integral part of the plant defense response (Lewis and Yamamoto, 1990). The lignin polymer is derived from the oxidative condensation of phenolic hydroxycinnamyl alcohols, termed monolignols (Figure 1A), which are derived from Phe via the general phenylpropanoid biosynthetic pathway, with specific transformations catalyzed by two monolignol pathway–specific enzymes, cinnamoyl CoA reductase and cinnamyl alcohol dehydrogenase (CAD). Cinnamoyl CoA reductase catalyzes the reduction of the hydroxycinnamoyl–CoA thioester bond, yielding hydroxycinnamaldehydes, whereas CAD catalyzes the final reductive step in the conversion of hydroxycinnamaldehydes to hydroxycinnamyl alcohols (monolignols) (Figure 1B).

Recently, Li et al. (2001) reported the identification of an enzyme from *Populus tremuloides* (aspen) with high sequence similarity to previously characterized CADs and reported specificity for the substrate sinapaldehyde. They proposed that this enzyme, called sinapyl alcohol dehydrogenase (SAD), is required for the biosynthesis of syringyl lignin in angiosperms, whereas distinct coniferaldehyde-specific enzymes are required for the biosynthesis of guiacyl lignin. This hypothesis met with opposition because it was previously proposed that the control of lignin composition (i.e., the ratio of syringyl and guiacyl units) originates from the regulated expression of tissue-specific classical CAD isoforms. In short, a single tissue-specific CAD isoform would catalyze the reduction of both coniferaldehyde and sinapaldehyde in a specific ratio, depending on the concentration, availability, and kinetic preferences of the expressed isoforms that are selected for based on the type of cell wall tissue required by the plant during a particular developmental stage.

To directly address questions regarding the nature of substrate specificity determination in monolignol biosynthesis, we have undertaken a structural and structurally guided kinetic examination of SAD. Primary sequence analysis reveals that SAD is a zinc-dependent alcohol dehydrogenase and a member of the medium-chain dehydrogenase/reductase (MDR) superfamily (McKie et al., 1993; Persson et al., 1994; Riveros-Rosas et al., 2003). We report here the high-resolution crystal structure of recombinant SAD bound to the $NADP^+$ cofactor and present kinetic data for the *Escherichia coli*–expressed aspen SAD and several active-site mutants describing their kinetic preferences for hydroxycinnamaldehyde substrates. We also present the computer-assisted docking of SAD substrates into the experimentally determined crystal structure, which enables us to propose a chemically reasonable model describing the structural basis of substrate specificity determination.

This SAD crystal structure provides a starting point for modeling active site variations between intraspecies and

[1] To whom correspondence should be addressed. E-mail noel@salk.edu; fax 858-597-0855.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Joseph P. Noel (noel@salk.edu).
Article, publication date, and citation information can be found at www.plantcell.org/cgi/doi/10.1105/tpc.104.029983.

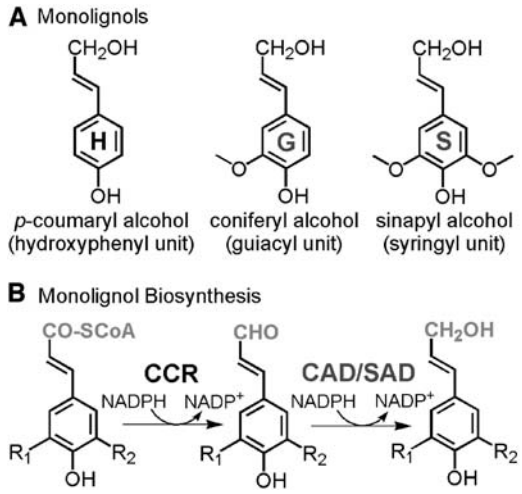

Figure 1. Biosynthesis of Monolignols.

(A) Monolignols.
(B) Monolignol biosynthesis.
p-Coumaryl alcohol: $R_1 = R_2 = H$; coniferyl alcohol: $R_1 = OCH_3$, $R_2 = H$; sinapyl alcohol: $R_1 = R_2 = OCH_3$.

interspecies CAD/SAD enzymes. Use of the SAD structure for homology modeling facilitates our understanding of the evolutionary origin of this sinapaldehyde-specific enzyme and its relationship to two subfamilies of plant alcohol dehydrogenases: the classical CADs, represented by *Eucalyptus gunni* CAD2 (EuCAD2) (Grima-Pettenati et al., 1993), and the defense-related dehydrogenases, represented by the *Arabidopsis thaliana* elicitor-responsive (ELI3-2), also called benzyl alcohol dehydrogenase (BAD) (Somssich et al., 1996). A greater understanding of the biosynthetic basis of lignin polymer composition is of both academic and industrial interest. Altering lignin polymer composition has implications for forage crop digestibility and paper pulping. This structural work, coupled with previously published structural studies of lignin-related methyltransferases such as caffeic acid 3-*O*-methyltransferase (Zubieta et al., 2002) and caffeoyl coenzyme A 3-*O*-methyltransferase (Ferrer et al., 2005), provides possible inroads to cogent engineering of lignin composition and type.

RESULTS

Purification and Crystallization

Aspen SAD was expressed in *E. coli* as an N-terminal octahistidine-tagged fusion protein and purified via nickel-nitrilotriacetic acid agarose (Ni^{2+}-NTA) affinity and size-exclusion chromatography. SAD was crystallized in the presence of a fourfold molar excess of both the $NADP^+$ cofactor and the substrate sinapaldehyde. Initial preparations of SAD protein, as well as initial crystallization conditions, included 2 mM DTT. Initial SAD crystals diffracted to 2.0 Å. Subsequent multiple anomalous dispersion phasing using the endogenous Zn^{2+} ions and density modification yielded excellent electron density maps. Model building and refinement against high-resolution data (see Methods) yielded a final SAD structure exhibiting well-defined electron density for all but the first three residues (Table 1). Despite our concerted efforts and the presence of sinapaldehyde in the crystallization solution, there was no electron density visible for sinapaldehyde. In previously reported structures of zinc-dependent alcohol dehydrogenases determined in the absence of substrate, a water molecule completes the coordination sphere of the catalytic Zn^{2+} ion at the active site (Clarke and Daffron, 1998). This coordinating water molecule is displaced as the substrate binds with the aldehyde/alcohol oxygen directly coordinating the Zn^{2+} ion. Analysis of the electron density around the catalytic Zn^{2+} ion reveals that a molecule of DTT (present during purification and in crystallization solutions) was bound in the active site of each monomer. To obtain a DTT- free crystal form of SAD, subsequent protein preparations and crystallization conditions contained 20 mM β-mercaptoethanol (βME) in place of DTT. SAD crystals grown in the presence of βME diffracted to 2.5 Å and lacked electron density for βME in the active site yet retained the zinc-coordinating water molecule and $NADP^+$. The structures of SAD crystallized with DTT or βME are nearly identical, overlaying with an overall backbone root mean square deviation (RMSD) of 0.25 Å and identical positioning of all active site side chains. The only observed difference is the torsion angle of Cys-98, which, in the case of the DTT-bound structure, is rotated to alleviate a steric clash with the DTT molecule. Because of the higher resolution of the refined SAD structure with a bound DTT molecule and the lack of active site changes, except in the case of the torsion angle of Cys-98, we will discuss only this structure in the remainder of this article.

Overall SAD Structure

The SAD monomer is composed of two domains: a dinucleotide binding domain (residues 170 to 302) that adopts a classic Rossmann fold (Rao and Rossmann, 1973), and a substrate binding domain (residues 1 to 169 and 303 to 362). The active site is defined by the deep cleft between the two domains, with a well-coordinated catalytic Zn^{2+} ion and NADP(H) cofactor bound at the active site (Figure 2A). The asymmetric unit contains two polypeptide chains, constituting the physiological SAD dimer. Dimerization occurs through the nucleotide binding domain (residues 282 to 302) of each monomer. The outermost strand (β16) of the parallel six-stranded β-sheet of monomer A forms hydrogen bonds with the corresponding β-strand of monomer B, exhibiting an antiparallel arrangement. The resulting quaternary structure contains a 12-stranded twisted β-sheet that buries 17.5% of the monomer surface area (2609 A^2). Additional intermonomer contacts are made through the α9 helix and the α9/β16 loop preceding the outermost β-strand.

$NADP^+$/NADPH Binding Site

Electron density for $NADP^+$ is well defined in both molecules located in the asymmetric unit (Figure 2B). The nicotinamide cofactor goes from a nonplanar, nonaromatic state in the reduced form (NADPH) to a planar, aromatic state in the oxidized form ($NADP^+$). The excellent quality of the electron density map

Table 1. Crystallographic Data, Phasing, and Refinement Statistics for Aspen SAD

	DTT SAD (Zn^{2+})				
	λ_1	λ_2	λ_3	HIGHRES	βME SAD
Wavelength	1.2822	1.2827	1.1808	1.1169	1.0093
Resolution (Å)	2.0	2.0	2.0	1.65	2.5
Total observations	740,853	703,510	669,126	1,649,776	224,963
Unique reflections	50,614	50,567	50,616	89,333	25,521
Completeness[a]	81 (82)	80 (76)	80 (79)	97 (94)	98 (99)
R_{sym} (%)[b]	8.6 (46.8)	8.2 (42.3)	8.5 (51.8)	7.8(76.8)	13.7(62.4)
No. of Zn^{2+} sites	4	4	4		
Overall FOM[c]			0.50		
Refinement statistics					
No. of protein atoms				5,432	5,434
No. of water molecules				477	43
No. of $NADP^+$ atoms				95	95
No. of DTT atoms				16	—
No. of glycerol atoms				6	—
R_{cryst}[d]				19.5	23.1
R_{free}[e]				21.3	28.9
RMSD bonds (Å)				0.0108	0.0114
RMSD angles (°)				1.622	1.702
Average B factors					
Protein (Å^2)				17.7	30.0
Water (Å^2)				26.7	23.6
$NADP^+$ (Å^2)				11.98	20.10
DTT (Å^2)				26.35	—
Glycerol (Å^2)				28.03	—

[a] Numbers in parentheses refer to the highest resolution shell.
[b] $R_{sym} = |Ih - \langle Ih \rangle|/Ih$, where $\langle Ih \rangle$ is the average intensity over symmetry equivalent reflections.
[c] FOM, figure of merit.
[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 using 5% of the data excluded from refinement.

depicts this planar state very clearly. The A-side of the nicotinamide ring is exposed to the solvent, confirming that SAD catalyzes A-side–specific hydride transfer, as reported for soybean CAD (Wyrambik and Grisebach, 1979).

Like many zinc-dependent ADHs, the SAD reaction is proposed to use a proton shuttle for donation of protons to or abstraction from the substrate (Clarke and Daffron, 1998). Specifically, in the reductive direction, the dihydronicotinamide C4 hydride attacks the C1 carbonyl carbon of sinapaldehyde, resulting in a highly polarized carbonyl oxygen that is stabilized by coordination to the catalytic Zn^{2+}. Ser-52 is proposed to shuttle its proton to the anionic carbonyl oxygen of the Zn^{2+}-stabilized alkoxide intermediate. Ser-52, in turn, receives a proton from the 2′-OH of the nicotinamide ribose, which then receives a proton from the neighboring nicotinamide ribose 3′-OH. His-55 completes the shuttle by donating a proton to the 3′-OH of the nicotinamide ribose. This proton relay necessitates that the nicotinamide ribose 2′ and 3′ hydroxyl moieties form hydrogen bonds with both Ser-52 and His-55, respectively, and with one another. Close examination of the density surrounding $NADP^+$ demonstrates that these hydrogen bonds form in this structure, supporting the contention that the static crystal structure represents a catalytically competent form of the SAD enzyme (Figure 2C).

Many Zn^{2+}-dependent ADHs have been structurally characterized; however, most such structures have been determined for NADH-dependent enzymes (Baker et al., 1992). The switch from the use of NADH to NADPH requires the substitution of an Asp residue, which hydrogen bonds with the 2′-OH of the adenine ribosyl of NADH, to something complementary with the 2′ phosphate of NADP(H). In SAD, this Asp residue is replaced by Thr-215, which, in conjunction with Ser-214, Ser-216, Lys-219, Asn-343, the backbone of Asn-192, and a water molecule, stabilizes the 2′ phosphate through hydrogen bonds and electrostatic interactions (Figure 2D). In classical CADs, the equivalent of Thr-215 in SAD is a Ser (Ser-212 in EuCAD2), and it has been implicated in determining cofactor specificity (Lauvergeat et al., 1995).

Zn^{2+} Coordination

Although originally reported to bind only one Zn^{2+} ion per monomer (Wyrambik and Grisebach, 1979), the SAD structure contains both catalytic and structural Zn^{2+} ions, which are characteristically coordinated to zinc-dependent MDR family enzymes. The structural zinc is coordinated by four Cys residues (Cys-103, Cys-106, Cys-109, and Cys-117) and localized to a protrusion (residues 102 to 120) from the core of the substrate

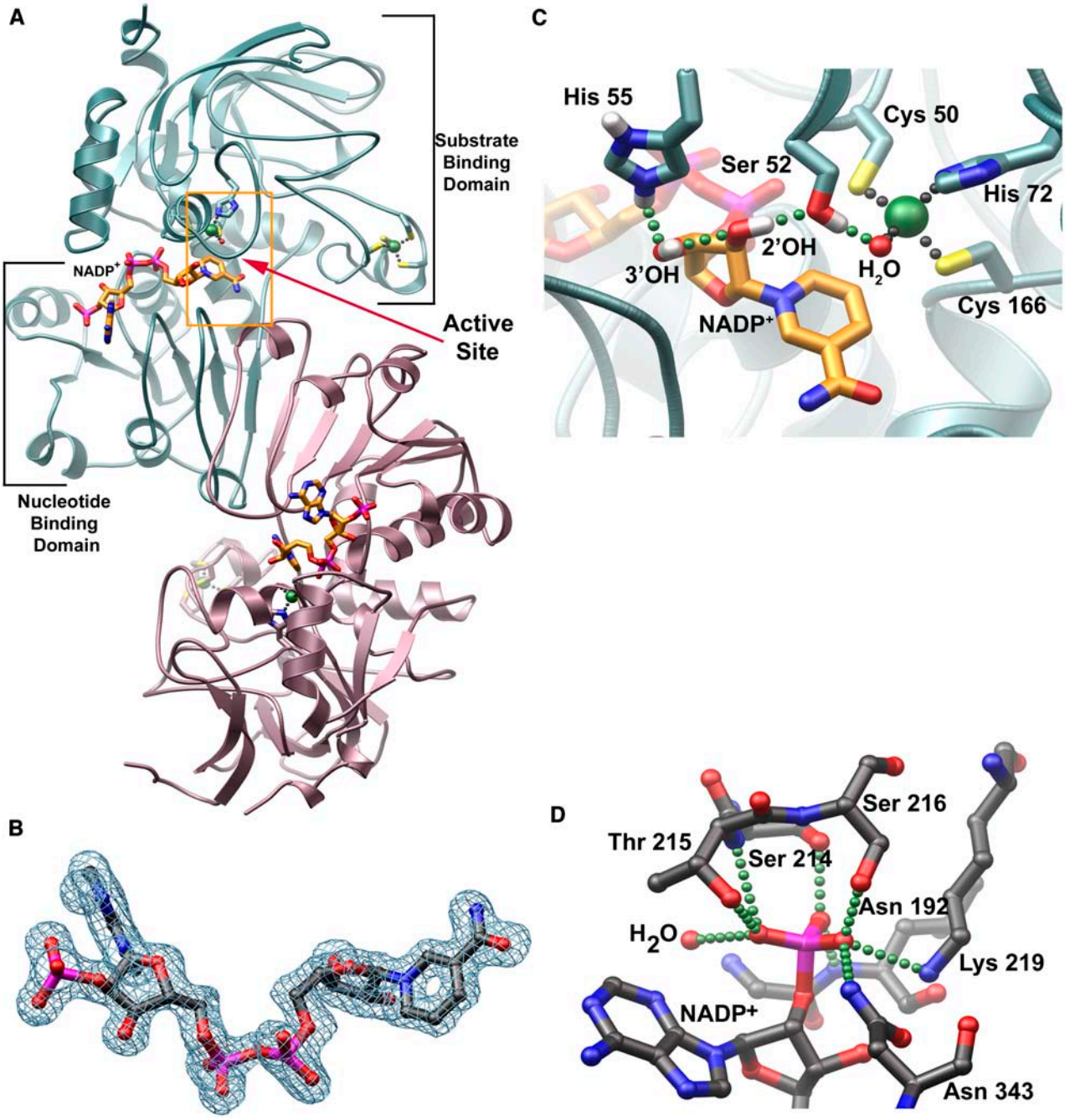

Figure 2. Three-Dimensional Architecture of the Aspen SAD/NADP$^+$ Complex.

(A) Ribbon representation of the aspen SAD dimer. Monomer A is colored blue, monomer B is colored rose, and NADP$^+$ is bound in the nucleotide binding domain. The active site cavity is boxed in orange. Each monomer coordinates one structural and one catalytic Zn^{2+} ion, each represented as green spheres with small black spheres indicating coordination bonds.

(B) 2Fo-Fc electron density map of NADP$^+$ contoured at 1.1 σ.

(C) Hydrogen bond formation (small green spheres) establishes a proton shuttle.

(D) Ball-and-stick representation of NADP$^+$ phosphate-stabilizing residues.

binding domain. In several tetrameric zinc-dependent ADHs, the structural zinc binding loops are involved in oligomerization, as dimers associate via these protruding loops, forming the functional tetramers (Banfield et al., 2001). Because SAD is dimeric, these loops are not in contact with other SAD molecules, and the functional role of the structural zinc remains unknown. The catalytic Zn^{2+} ion is located at the base of the active site cleft coordinated by the side chains of Cys-50, His-70, and Cys-166. A water molecule completes the tetrahedral coordination shell (Figure 3).

Substrate Binding Cavity

Although sinapaldehyde was included in the crystallization solution, the absence of extra electron density in the active site and the presence of an ordered water molecule coordinating the catalytic Zn^{2+} ion indicates that despite concerted efforts, sinapaldehyde was not bound in either the βME- or DTT-containing SAD crystal. However, well-defined density for $NADP^+$ and the overall high resolution of the refined structure enabled a precise description of the substrate binding cleft. Aliphatic and aromatic residues line the SAD active site and form a pocket in which the hydroxycinnamaldehyde substrates bind via largely shape-dependent van der Waals interactions (Figure 3). Residues of monomer B, in part, define the active site pocket of monomer A. The correct orientation of the aldehyde substrate would be facilitated by Zn^{2+} coordination to the aldehyde carbonyl oxygen and hydrogen bond formation with the Ser-52 hydroxyl side chain.

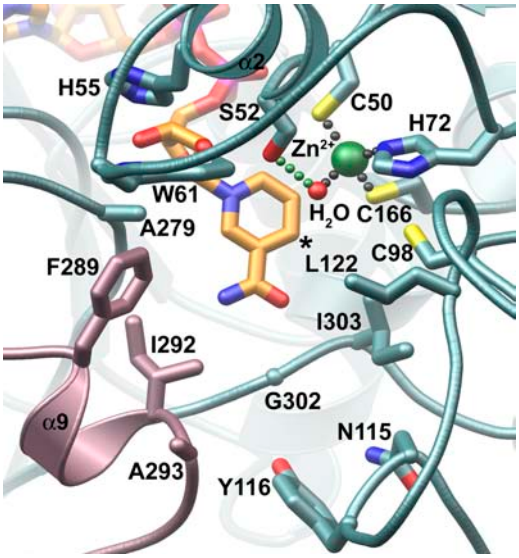

Figure 3. The Aspen SAD Active Site.

SAD active site architecture is defined by residues from both monomers. The catalytic Zn^{2+} ion is tetrahedrally coordinated by Cys-50, His-72, Cys-166, and a water molecule, represented as a red sphere. Small green spheres indicate a hydrogen bond formation between the Zn^{2+} ion coordinating water and Ser-52. A ball-and-stick model of $NADP^+$ is shown. The site of hydride transfer is labeled with an asterisk.

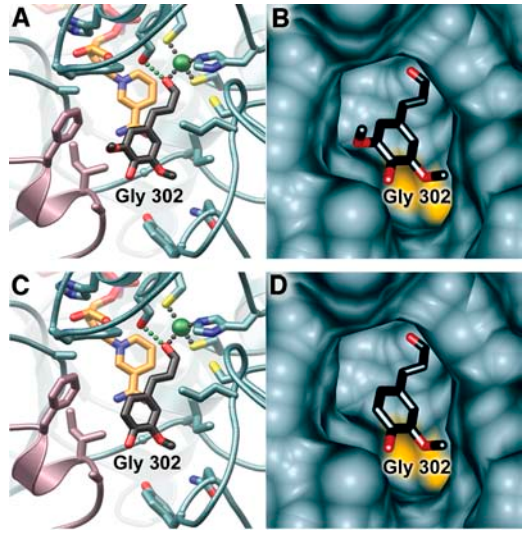

Figure 4. The SAD Substrate Binding Pocket.

(A) and **(C)** Docking solutions for sinapaldehyde and coniferaldehyde (shown with black carbon atoms), respectively. The phenyl rings pack flat against the base of the active site, with the aldehyde carbonyls of the substrates directly coordinating the Zn^{2+} ion.
(B) and **(D)** Surface representation of the SAD active site cavity. The surface corresponding to Gly-302 is shaded yellow.

The left active site wall of monomer A (defined based on the central location of the catalytic Zn^{2+} and $NADP^+$ on the top) is partially defined by two aromatic residues, Phe-289, which protrudes into the active site from α9 of monomer B, and Trp-61, which is localized to the α2/β4 loop of monomer A (Figure 3). Ala-279, from the β14/α9 loop, also contributes to the left wall of the active site. Ala-293, also from α9 of monomer B, completes the left wall of the active site. In addition to the nicotinamide ring, Gly-302, from the β15/α10 loop of monomer A, and Ile-292, from α9 of monomer B, define the floor of the active site. Tyr-116, the backbone of Asn-115 from monomer A, and Ile-303, from the loop connecting β15 and α10 of monomer A, define the right wall of the active site. The remainder of the upper right and left walls of the active site is molded by the polar residues coordinating the catalytic Zn^{2+} ion and include Cys-50, His-72, Cys-166, the critical hydrogen bond donor Ser-52, and Cys-98.

Computer-Aided Substrate Docking

To probe the structural basis of substrate binding in SAD, we used the docking software GOLD (Genetic Optimization for Ligand Docking) to determine the most favorable orientation of the substrate molecules in the SAD active site (Jones et al., 1997). The substrate sinapaldehyde was docked to the SAD active site, and the resulting solution (Figures 4A and 4B) was mechanistically consistent with sinapaldehyde reduction. The cinnamyl ring of sinapaldehyde resided comfortably in the active site, with both methoxy substituents cradled in separate hydrophobic clefts. The cinnamyl ring is predicted to bind directly over Gly-302. It follows that any substitution at the 302 position would

disrupt the binding of sinapaldehyde in the active site pocket, as calculated by GOLD. The phenolic hydroxyl moiety is oriented toward the backbone carbonyl of Ala-293 of monomer B. Although the distance in the calculated complex is just out of hydrogen bonding range (3.67 Å), a slight hinging motion at the interface between monomer A and monomer B may occur, resulting in hydrogen bond formation between the phenolic hydroxyl and the backbone carbonyl of Ala-293. Hydrogen bonding interactions at these sites would effectively tether the sinapaldehyde at both ends, through the aldehyde head, via Ser-52 hydrogen bonding and Zn^{2+} coordination, and at the phenolic tail, via hydrogen bonding with the Ala-293 backbone carbonyl oxygen. Chemical features of the docking solution and the fact that the top five solutions converged on this particular substrate orientation support the relevance of these calculated models.

SAD can also catalyze the reduction of coniferaldehyde. We used the same docking strategy, and the resulting solution was coniferaldehyde docked in the same orientation as sinapaldehyde, with the 3-methoxy substituent cradled in the hydrophobic cleft formed along the right wall (molded by Ile-303 and the backbone of Asn-115) of the active site, with the phenolic tail oriented toward the backbone carbonyl oxygen of Ala-293 (Figures 4C and 4D). Reflecting on the chemical structure of coniferaldehyde, we propose that SAD maintains its greatest specificity for sinapaldehyde, in part because of the greater hydrophobic surface area of the sinapaldehyde substrate, associated with the ''extra'' 5-methoxy group, and its complementarity to the available hydrophobic surface area of the SAD active site, as described above. Upon coniferaldehyde binding, the hydrophobic cleft along the left wall of the active site, formed largely by Phe-289 and Ile-292 of monomer B, is left exposed, providing a structural explanation for this experimentally determined specificity.

There has been significant debate surrounding the relationship of SAD to a family of enzymes that includes incorrectly annotated mannitol dehydrogenases (MTDs), defense-related ELI3 enzymes, and several nonclassical CADs (Raes et al., 2003; Kim et al., 2004). Recently, several ELI3s and nonclassical CADs were incorrectly reannotated as MTDs, despite published reports describing experimental preferences for cinnamaldehyde- and/or benzaldehyde-derived substrates. A careful reevaluation of the familial relationships is warranted in light of our structural and kinetic results. BLAST searches for short identical matches in key functional regions of SAD sequence space revealed by the crystal structure enabled us to focus on statistically significant matches that exhibit identity in regions of proposed functional importance, rather than on regions of ancestral structural importance. This database search revealed that SAD is most similar to a group of proteins annotated as MTDs/ELI3s/nonclassical CADs, all sharing key subfamily determinants, including a Trp at a location equivalent to aspen SAD position 61 (Figure 5). By contrast, classical CADs bear an invariant Trp at the position equivalent to Leu-122 of aspen SAD.

Although a vast number of these enzymes are simply annotated by similarity to each other, thus perpetuating functionally incorrect annotations in databases, several of these SAD-like enzymes have been functionally characterized. These characterized enzymes include FxaCAD (Blanco-Portales et al., 2002), recently incorrectly reannotated as MTD_FRAAN and bearing 78.9% identity to SAD, PcELI3 (Logemann et al., 1997), recently incorrectly reannotated as MTD_PETCR and bearing 71.6% identity to SAD, and AtBAD (Somssich et al., 1996), annotated as AtELI3-2 and bearing 71.3% identity to SAD. These enzymes use hydroxycinnamaldehydes and benzaldehydes with varying efficiency. Homology modeling reveals that the active site topologies of these enzymes largely mirror that of SAD, with interesting substitutions at the base of the active site. Like SAD, FxaCAD bears a Gly at the base of the active site, defining the flat surface against which the phenolic ring of the hydroxycinnamaldehyde substrates can pack. The docking of sinapaldehyde into the FxaCAD active site results in an orientation nearly identical to that of sinapaldehyde in SAD (Figures 6A and 6B). This in silico result correlates well with the published kinetics of FxaCAD, which indicate that, like SAD, FxaCAD displays high activity for sinapaldehyde (Blanco-Portales et al., 2002).

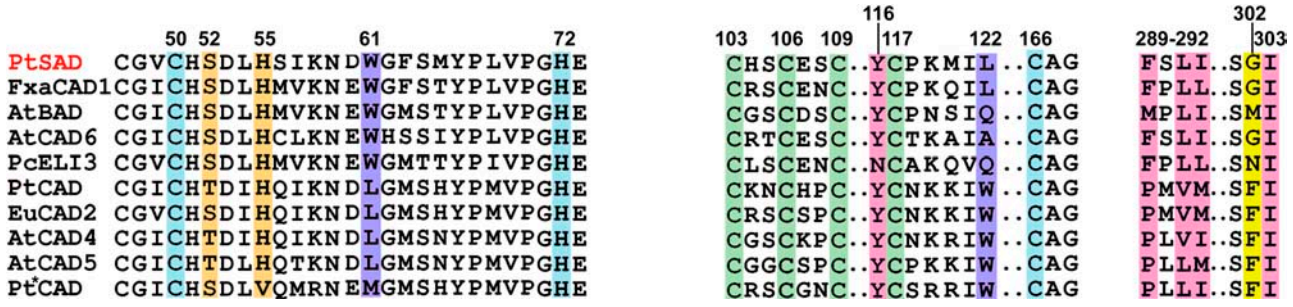

Figure 5. Partial Sequence Alignment of SAD-Like and CAD-Like Enzymes.

Alignment of key residues in SAD-like (*Populus tremuloides* SAD [PtSAD], *Fragaria ananassa* CAD1 [FxaCAD1], *Arabidopsis thaliana* BAD/ELI3-2 [AtBAD], *A. thaliana* CAD6 [AtCAD6], and *Petroselinum crispum* CAD [PcELI3]) and classical CAD-like (*P. tremuloides* CAD [PtCAD], *Eucalyptus gunnii* CAD2 [EuCAD2], *A. thaliana* CAD4 [AtCAD4], *A. thaliana* CAD5 [AtCAD5], and *Pinus taeda* CAD [Pt*CAD]) enzymes. Catalytic Zn^{2+} ion coordinating residues are shaded blue; proton shuttle residues are shaded orange; key bulky active site residues are shaded purple; structural Zn^{2+} ion coordinating residues are shaded green; residues defining active site topology are shaded pink; and the residue we propose to be the key determinant of substrate specificity is shaded yellow.

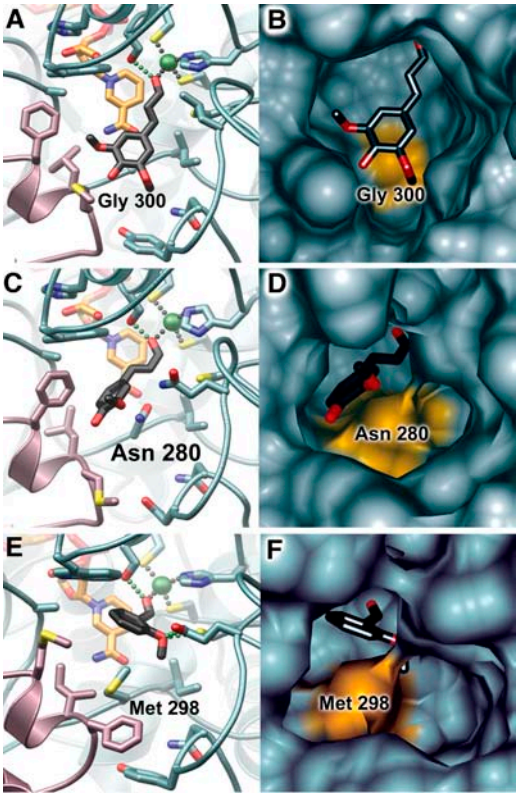

Figure 6. Substrate Docking in the Modeled Active Sites of SAD/CAD Family Enzymes.

(A) Docking solution for sinapaldehyde (shown with black carbon atoms) in the FxaCAD active site. The phenyl ring packs flat against the base of the active site in the same orientation as for sinapaldehyde docked in the SAD active site.
(B) Surface representation of the modeled FxaCAD active site cavity. The surface corresponding to Gly-300 is shaded yellow.
(C) Docking solutions for coniferaldehyde (shown with black carbon atoms) in the PcELI3 active site. Because of the bulk of the Asn side chain, the phenyl ring packs on its unsubstituted edge against the base of the active site.
(D) Surface representation of the modeled PcELI3 active site cavity. The surface corresponding to Asn-280 is shaded yellow.
(E) Docking solutions for 2-methoxybenzaldehyde (shown with black carbon atoms) in the AtBAD active site. Because of the bulk of the Met side chain, the active site is largely occluded.
(F) Surface representation of the modeled AtBAD active site cavity. The surface corresponding to Met-298 is shaded yellow.

Unlike SAD and FxaCAD, PcELI3 bears an Asn at the base of the active site cavity. It is clear from homology modeling that this substitution changes the size and shape of the cavity and necessitates a different substrate specificity than that of SAD and FxaCAD. In fact, Logemann et al. (1997) reported that PcELI3 maintains the smallest K_m for cinnamaldehyde, followed by coumaraldehyde and coniferaldehyde, and zero activity when assayed against sinapaldehyde. Analysis of the homology model of PcELI3 reveals a partially truncated pocket in which automated docking supports preferential recognition and binding of 2-methoxybenzaldehyde in the more spatially restricted active site. It is interesting that PcELI3 will accept cinnamaldehyde, coumaraldehyde, and coniferaldehyde as substrates but not sinapaldehyde. Automated docking of coniferaldehyde and cinnamylaldehyde revealed that these substrates can be accommodated in this smaller active site by binding in an orientation that is not flat and coplanar with NADPH, as in SAD. Specifically, cinnamylaldehyde, coniferaldehyde, and coumaraldehyde maintain a perpendicular arrangement with respect to the nicotinamide ring, with the unsubstituted edge of the phenyl ring of each wedged into the base of the active site (Figures 6C and 6D). It follows that given the bisubstitution pattern of sinapaldehyde, which bears two methoxy substituents on the benzyl ring, the active site arrangement of PcELI3 cannot accommodate this larger substrate.

Finally, AtBAD bears a Met at the base of the active site, and homology modeling again reveals a significant change in active site size and shape. The AtBAD active site is very sterically restricted relative to the active site of SAD, providing a structural explanation for published kinetic parameters that indicate that only the shorter benzaldehyde-derived substrates serve as acceptable substrates (Figures 6E and 6F) (Somssich et al., 1996).

Steady State Kinetic Analysis of Wild-Type SAD

Although kinetic parameters have been reported for aspen SAD (Li et al., 2001), we reevaluated these parameters because of unusual, previously unreported steady state kinetic behavior. SAD was incubated with sinapaldehyde or coniferaldehyde and the reactions acid-quenched at 1, 2, and 3 min, reaction times at which product formation is linear. Substrates and products were separated after quenching by HPLC, and the quantities of sinapyl or coniferyl alcohol products formed were used to calculate initial velocities (see Methods). It is clear from the plots of the initial reaction velocity versus substrate concentration that the SAD-catalyzed reduction of both sinapaldehyde and coniferaldehyde is inhibited with increasing substrate concentrations (Figures 7A and 7B, insets). This substrate inhibition phenomenon is not uncommon in enzymes exhibiting bisubstrate kinetics; in fact, substrate inhibition has been reported for both SAD-like (MsaCAD1) (Brill et al., 1999) and CAD-like (EuCAD2) (Lauvergeat et al., 1995) enzymes, although it has been overlooked and is inconsistent with recent reports of CAD family steady state kinetics (Kim et al., 2004).

As described previously for human estrogenic 17β-hydroxysteroid dehydrogenase (Gangloff et al., 2001), at low substrate concentrations, the SAD reaction was found to nearly obey Michaelis-Menten kinetics (Figures 7A and 7B). Therefore, we can use a low substrate concentration range to determine the apparent K_m and apparent maximum velocity (V_{max}). Under conditions of substrate inhibition, it is also appropriate to report K_i, the apparent inhibition constant for an inhibitory substrate molecule, which can be derived by fitting to the substrate inhibition equation

$$v = V_{max}[S]/(K_m + [S] + [S]^2/K_i) \text{ (Eszes et al., 1996)}$$

where v = initial reaction velocity. However, accurate fitting to this equation breaks down when the kinetic parameters K_m and K_i are

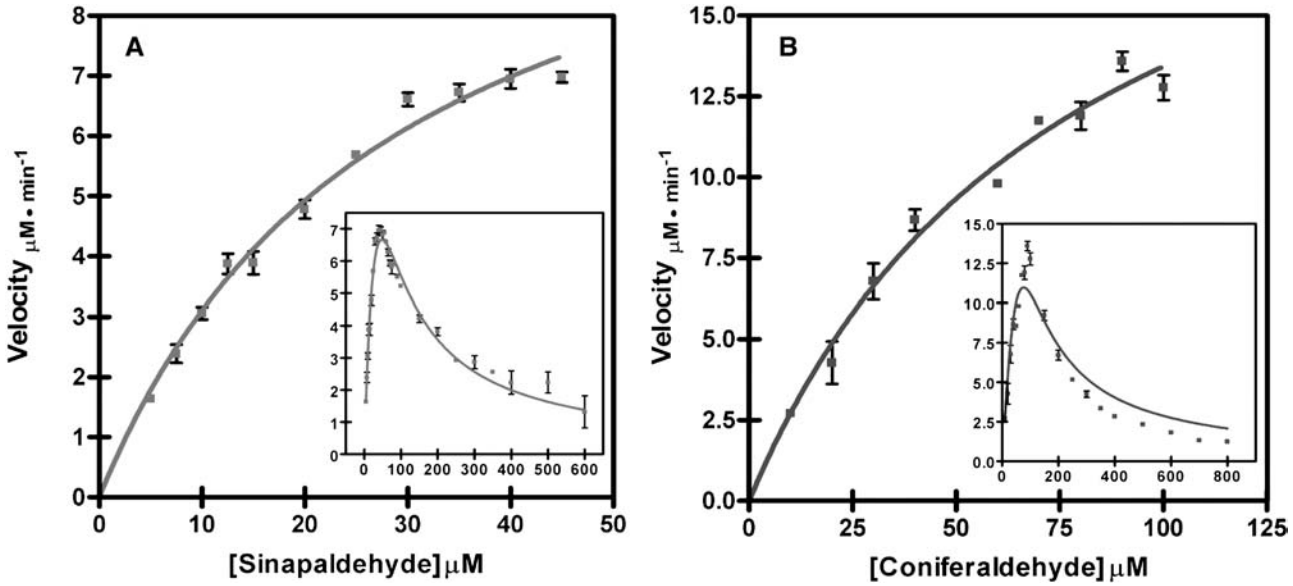

Figure 7. Initial Velocity versus Substrate Concentration Plots of Wild-Type SAD-Catalyzed Reactions.

(A) Wild-type SAD-catalyzed reduction of sinapaldehyde at low substrate concentrations. The inset shows substrate inhibition occurring at concentrations exceeding 60 μM.

(B) Wild-type SAD-catalyzed reduction of coniferaldehyde at low substrate concentrations. The inset shows substrate inhibition occurring at concentrations exceeding 100 μM.

tightly coupled (Eszes et al., 1996). From our results, it is apparent that the substrate inhibition constant K_i of the SAD-catalyzed reduction of coniferaldehyde is tightly coupled with K_m, resulting in a poor fit to the substrate inhibition equation (Table 2, Figure 7B). However, the coupling of K_i and K_m in the SAD-catalyzed reduction of sinapaldehyde is not as tight, allowing for a better fit to the substrate inhibition equation and thus affording determination of K_i (Table 2, Figure 7A). Further details regarding this inhibitory behavior and the relationship between substrate identity, enzyme structure, and architecture and inhibitory activity are currently under investigation.

Nevertheless, in agreement with previously reported data (Li et al., 2001), sinapaldehyde is preferentially reduced by the SAD enzyme, as suggested by the lower apparent K_m for the sinapaldehyde relative to coniferaldehyde (Table 2). The specificity constant, the ratio of the turnover number (k_{cat}) to the K_m (k_{cat}/K_m), provides an apparent second-order rate constant for the reaction of free SAD and free aldehyde substrate. k_{cat}/K_m is 6.8-fold greater for the reduction of sinapaldehyde than that of coniferaldehyde, supporting the contention that sinapaldehyde is the preferred in vitro substrate of recombinant aspen SAD.

Mutational Analysis of Aspen SAD versus Aspen CAD-Like Active Site Architecture

The use of the SAD structure for homology modeling of this group of MTD/ELI3/nonclassical CADs bearing high sequence identity to SAD has been informative for understanding the principles of substrate specificity in the SAD active site. However, on examination of active site residues of SAD and classical CAD-like enzymes, it is clear that major active site substitutions exist (Figure 5). A three-dimensional homology model of aspen CAD, a classical CAD, using the SAD structure as a template (55% sequence identity) reveals some striking active site differences. Notably, a swap of aromatic residues defining the active site cavity occurs across the two families. SAD bears bulky aromatic residues, including Trp-61 and Phe-289, on the left wall of the active site and small hydrophobic residues, including Leu-122 and Gly-302, at the base and on the right wall of the active site, respectively (Figure 3). On the other hand, classical CAD-like enzymes bear hydrophobic residues, including Leu-61 and Pro-289, on the left wall and bulky aromatic residues, such as Trp-122 and Phe-302, on the right wall and at the base of the active site, respectively. This complementary but opposite geometric substitution pattern substantially changes the topology of the active site.

As discussed above, the presence of Gly-302 is critical for the formation of the rather flat active site floor. Substitution of a Phe at this position in classical CADs necessitates that hydroxycinnamaldehyde substrates bind in classical CAD-like enzymes in a different orientation than predicted for SAD (Figures 4A and 4B). In fact, based on a homology model of EuCAD2, McKie and coworkers (1993) proposed that hydroxycinnamaldehyde substrates are sandwiched between Trp-119 (Leu-122 in SAD) and Phe-298 (Gly-302 in SAD), supporting our contention that the substrate binding mode is different between classical CADs and SAD-like enzymes. Overall sequence identity aside, given this large variation of active site topologies, it is unlikely that SAD has evolved most recently from the subfamily of ADHs defined by classical CADs; instead, it is more closely related to the MTD/ELI3/nonclassical CAD subfamily.

Table 2. Kinetic Constants Calculated for Wild-Type and Mutant Aspen SAD

	SAD		W61L/F289P		L122W/G302F		W61L/L122W/F289P/G302F	
Apparent Kinetic Parameter	SLD	CLD	SLD	CLD	SLD	CLD	SLD	CLD
V_{max} ($\mu M \cdot min^{-1}$)	12.0 ± 0.8	23.9 ± 2.6	>24[a]	110.8 ± 17.8	7.8 ± 0.2	6.6 ± 0.6	>81[a]	31.6 ± 3.0
K_m (μM)	28.6 ± 3.6	78.0 ± 16.0	>4.5 mM[a]	1694.0 ± 473.3	399.8 ± 56.6	22.7 ± 6.2	>5 mM[a]	504.7 ± 132.5
k_{cat} (min^{-1})	2,434.1	975.5	n.a.[a]	150.7	787.9	666.7	n.a.[a]	1279.4
K_i (μM)	36.2 ± 8.6	n.d. coupled[b]	n.a.[a]	n.d. insufficient[c]	n.d.[d]	155.5 ± 29.3	n.a.[a]	n.d. insufficient[c]
k_{cat}/K_m ($\mu M^{-1} \cdot min^{-1}$)	85.1	12.5	n.a.[a]	0.09	2.0	29.4	n.a.[a]	2.5
Enzyme concentration (nM)	4.93	24.5	9.9	735.0	9.9	9.9	30.8	24.7

CLD, coniferaldehyde; SLD, sinapaldehyde.

[a] Enzyme not saturated; therefore, kinetic parameters could not be determined.

[b] n.d. coupled, not determined because of tight coupling.

[c] n.d. insufficient, substrate inhibition insufficient to derive inhibition constant.

[d] n.d., no substrate inhibition detected.

In an attempt to address the importance of these bulky residues in the SAD and CAD active sites, we generated and kinetically characterized a set of active site mutants that represent several combinations of the conserved residues described above (Figure 8). First, bulky residues lining the left wall of wild-type SAD (Trp-61 and Phe-289) were replaced by the smaller hydrophobic residues that are conserved along the left wall of classical CADs (W61L and F289P), resulting in a larger active site cavity. As expected, kinetic characterization of this ''voluminous active site'' mutant confirmed that the major kinetic parameter affected was the K_m for both sinapaldehyde (Figure 8A) and coniferaldehyde (Figure 8B). The reaction velocity for the reduction of sinapaldehyde increased linearly from 0 to 5 mM sinapaldehyde, indicating that the K_m for sinapaldehyde was at least greater than 5 mM. This mutant reduced coniferaldehyde only when large amounts of enzyme were used (30-fold greater than that used for the wild-type SAD reaction) and exhibited a 22-fold greater K_m for coniferaldehyde. The k_{cat} was reduced 6.5-fold, and substrate inhibition occurred at coniferaldehyde concentrations greater then 2.5 mM (Table 2).

A quadruple mutant was generated to simulate a SAD-to-CAD conversion. Bulky residues lining the left wall of the SAD active site were replaced by the smaller hydrophobic residues that are conserved along the left wall of classical CADs (W61L and F289P), whereas bulky residues that are conserved along the right wall in classical CADs were engineered into the SAD active site (L122W and G302F), yielding a classical CAD-like active site topology. Again, the reaction velocity of the mutant SAD-catalyzed reduction of sinapaldehyde (Figure 8C, Table 2) increased linearly from 0 to 5 mM sinapaldehyde, indicating that the K_m for sinapaldehyde was at least greater than 5 mM. The reaction with coniferaldehyde (Figure 8D, Table 2) exhibited a k_{cat} similar to that of wild-type SAD but a 6.5-fold increase in K_m. Substrate inhibition occurred at coniferaldehyde concentrations greater than 1.75 mM (Figure 8D). If this mutant were in fact an aspen CAD-like enzyme, then it would display specificity for coniferaldehyde but retain its ability to reduce sinapaldehyde with reasonable efficiency. It is clear from these results that CAD residues beyond those explored here play a role in conferring specificity for coniferaldehyde while maintaining reasonable K_m values for sinapaldehyde.

Most interestingly, bulky residues lining the left wall of the wild-type SAD (Trp-61 and Phe-289) active site were conserved, whereas bulky residues that are conserved along the right wall in classical CADs were substituted in the SAD active site (L122W and G302F), resulting in a spatially restricted active site volume. Kinetic analysis of this mutant with sinapaldehyde revealed a 3-fold decrease in k_{cat} and a 14-fold increase in K_m, suggesting, as expected, that the bulkier substrate sinapaldehyde does not bind well in the spatially restricted active site (Figure 8E, Table 2). Furthermore, substrate inhibition was not observed in the concentration range tested (0 to 5 mM). Surprisingly, a 1.5-fold decrease in k_{cat} was seen in the coniferaldehyde-containing reaction, whereas a 3.4-fold decrease in K_m was observed, resulting in a 2.4-fold increase in the specificity constant k_{cat}/K_m (Figure 8F, Table 2). This finding suggests that coniferaldehyde is better accommodated in the restricted active site of the mutant than in that of the wild-type SAD. As in wild-type SAD, substantial substrate inhibition is observed when 90 μM coniferaldehyde concentrations are exceeded. Visual docking of coniferaldehyde in this spatially restricted active site reveals a new coniferaldehyde-specific binding pocket in which the cinnamyl ring of coniferaldehyde is enveloped by three aromatic residues, Trp-61, Trp-122, and Phe-302.

DISCUSSION

Along with other researchers, we maintain that SAD is most closely related to the group of reductases that maintain specificity for benzaldehyde- and hydroxycinnamaldehyde-derived substrates (Raes et al., 2003; Kim et al., 2004). We further propose that in SAD-like enzymes, the identity of the residue at a position equivalent to residue 302 in aspen SAD is a key determinant of substrate specificity. This hypothesis is supported by the correlation of substrate specificities, kinetic parameters, and automated substrate-docking experiments with the identity of residues at this position in both the SAD crystal structure and the SAD-like homology models.

Use of the SAD structure for automated substrate docking and homology modeling of related enzymes has revealed a structure-based rationalization of substrate specificities in SAD-like enzymes. This correlation will be of great value for understanding

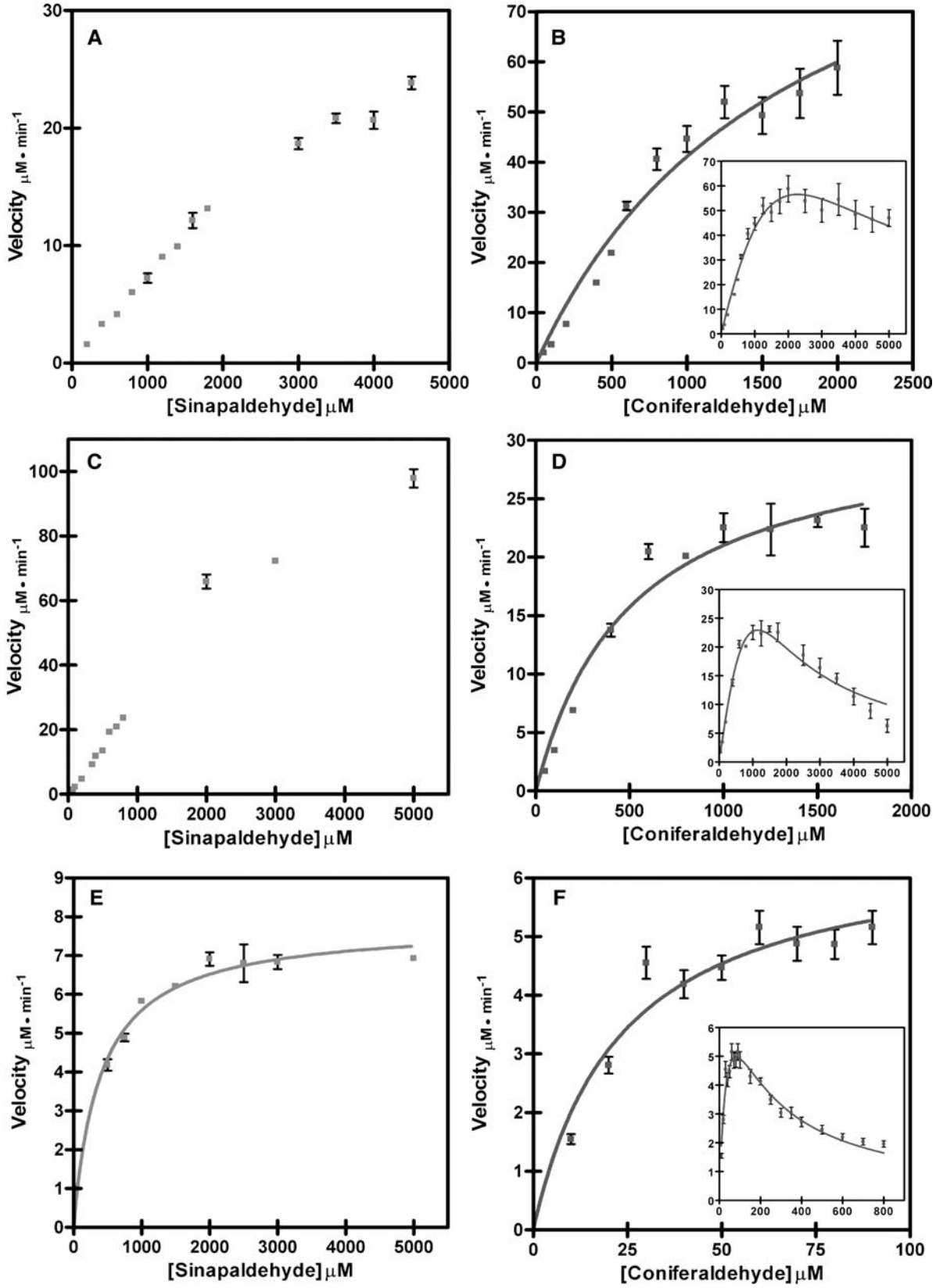

Figure 8. Initial Velocity versus Substrate Concentration Plots of Mutant SAD-Catalyzed Reactions.

and possibly predicting the substrate specificities of uncharacterized SAD-like enzymes and for providing a structure platform to begin the rational engineering of new specificities in this family of enzymes. Although SAD bears only 36% sequence identity to the recently characterized *Saccharomyces cerevisiae* CAD (ScCAD), the three-dimensional structures of these two enzymes share a similar backbone fold, with the alignment of Cα backbone atoms overlaying with an RMSD of 1.55 Å across 335 residues (Valencia et al., 2004). Unlike SAD, ScCAD does not reduce sinapaldehyde or coniferaldehyde, selectively using cinnamaldehyde as a substrate instead. Direct use of the SAD structure to understand this change in substrate specificity is not feasible, because ScCAD bears several major changes in the active site cavity. The equivalent of Gly-300 in SAD is replaced by a Tyr residue in ScCAD (Tyr-298). This Tyr points directly into the active site, blocking the putative substrate binding site predicted from the structure of SAD. Furthermore, the equivalent of the key α9 region of monomer B that contributes to the active site of monomer A contains a two-residue insertion in ScCAD, thus altering the active site topology compared with that of SAD. This insertion is unique to ScCAD, as no insertion occurs in either SAD-like or classical CAD enzymes. This insertion may be the key change in ScCAD that results in the half-of-the-sites reactivity mechanism proposed by Valencia and colleagues (2004). In the SAD structure presented here, $NADP^+$ is bound in both active sites of the functional dimer, further highlighting the large structural and functional divergence between these two distantly related enzymes.

Further work both in vitro and in vivo is necessary to address the possible physiological role of substrate inhibition in the SAD-like and CAD-like superfamilies of MDRs. Lauvergeat and coworkers (1995) suggested that CAD isoforms may exhibit different levels of substrate inhibition, necessitating the quantitative evaluation of kinetic parameters for family members with several putative substrates. Because different cell and tissue types require distinct compositions of lignin, this differential substrate inhibition may be the mechanism by which flux through the monolignol biosynthetic network is temporally and spatially regulated. Inhibition of CAD/SAD isoforms in the presence of a particular concentration of a potential substrate could effectively divert substrates to other isoforms bearing different specificities.

Structure-guided engineering of these hydroxycinnamaldehyde reductases will provide useful tools for probing both the functional role of substrate inhibition and the physiological role these enzymes play in monolignol formation. In particular, the L122W/G302F mutant can reduce sinapaldehyde (albeit with a lower turnover number) and yet does not exhibit sinapaldehyde substrate inhibition. Therefore, this mutant enzyme can be used to investigate the physiological role of substrate inhibition in monolignol formation in transgenic plants. Furthermore, unlike wild-type SAD, which is highly specific for sinapaldehyde, the L122W/G302F mutant displays its greatest specificity for coniferaldehyde. Thus, introduction of this coniferaldehyde-specific mutant into Δ*sad* aspen may directly address the hypothesis that aspen SAD is solely responsible for the formation of syringyl lignin.

This small suite of mutants demonstrates the complexity of sinapaldehyde versus coniferaldehyde specificity and the importance of addressing substrate-mediated kinetic inhibition in SAD-like and classical CAD-like enzymes. Discussion of substrate specificities and kinetic parameters of these enzymes will require the careful reevaluation of the experimental design used to derive steady state kinetic parameters. We urge caution in interpreting values derived from Lineweaver-Burk plots, which can return nearly linear plots even under conditions that do not satisfy the conditions used in the derivation of the Michaelis-Menten equation (Fersht, 1977). Our structural results, along with the kinetic characterization of a limited set of SAD mutants, demonstrates that major changes in specificities and catalytic efficiencies can occur as a result of minimal changes in active site residues. It is likely that a single point mutation in the SAD active site will alter the specificity and efficiency of the reduction reaction. Use of the SAD structure for predicting how minimal residue changes affect SAD-catalyzed reduction will facilitate the subsequent engineering of altered or novel substrate specificities in SAD-like enzymes. Engineering of SAD-like enzymes may have far-reaching effects on the commercial utility of plants as forage crops and for paper pulping. In addition, subtle changes in the substrate specificities of enzymes involved in modifying lignin content and chemistry may play a direct role in modulating defense responses in transgenic plants. The increased molecular understanding of enzymes involved in mounting a plant defense response will prove to be critical for understanding metabolic adaptation in the context of plant defense mechanisms that evolve under conditions of changing environmental challenges.

In a broader sense, the SAD structure presented here is one example of how structural biology combined with detailed kinetic analysis can generate new hypotheses concerning enzyme evolution in the context of secondary metabolism. Our structural and kinetic results suggest that SAD is structurally most closely

Figure 8. (continued).

(A) F289P/W61L SAD-catalyzed reduction of sinapaldehyde.

(B) F289P/W61L SAD-catalyzed reduction of coniferaldehyde at low substrate concentrations. The inset shows moderate substrate inhibition occurring at coniferaldehyde concentrations exceeding 2.0 mM.

(C) F289P/W61L/G302F/L122W SAD-catalyzed reduction of sinapaldehyde.

(D) F289P/W61L/G302F/L122W SAD-catalyzed reduction of coniferaldehyde at low substrate concentrations. The inset shows substrate inhibition occurring at coniferaldehyde concentrations exceeding 1.75 mM.

(E) G302F/L122W SAD-catalyzed reduction of sinapaldehyde.

(F) G302F/L122W SAD-catalyzed reduction of coniferaldehyde at low substrate concentrations. The inset shows substrate inhibition occurring at coniferaldehyde concentrations exceeding 90 μM.

related to dehydrogenases functioning in wound-induced plant defense. However, the physiological implications of these correlative structure–function relationships remain to be elaborated. Two alternative hypotheses can be drawn from these results. SAD may represent a defense-related dehydrogenase that was recruited to constitutive lignin biosynthesis upon the acquisition of specificity for monolignol precursors. Alternatively, SAD may represent a defense-related dehydrogenase that has acquired this unique specificity but uses traditional monolignol precursors as chemical building blocks in the toolbox of plant defense mechanisms. The complexity of both constitutive lignification and defense-induced deposition of monolignol-like phenolics in the cell wall renders this distinction experimentally challenging. Therefore, additional in planta experimentation will be critical to defining the true physiological function of SAD-like enzymes.

METHODS

Protein Expression, Purification, and Mutagenesis

Populus tremuloides (aspen) SAD (GenBank accession number AAK58693.1; a kind gift of Vincent Chiang) was subcloned into the pHIS8 *Escherichia coli* expression vector derived from pET28a(+). Transformed *E. coli* BL21(DE3) cells were incubated with shaking at 37°C in terrific broth (Tartof and Hobbs, 1987) containing 50 μg/mL kanamycin until $OD_{600} = 1.0$. Protein expression was induced with 0.5 mM isopropyl 1-thio-β-galactopyranoside, and the cultures were incubated with shaking at 22°C for 6 h. Cells were harvested by centrifugation at 9000*g*, and cell pellets were resuspended in lysis buffer (500 mM NaCl, 50 mM Tris-HCl, pH 8.0, 20 mM imidazole, 10% [v/v] glycerol, 1% [v/v] Tween 20, and 10 mM βME) supplemented with 4 mM benzamidine, 1 mM phenylmethylsulfonyl fluoride, 2 μg/mL leupeptin, and 0.5 mg/mL lysozyme. After sonication and centrifugation at 100,000*g*, supernatant was passed over a Ni^{2+}-NTA column (Qiagen, Valencia, CA) equilibrated in lysis buffer and washed with 10 bed volumes of wash buffer (500 mM NaCl, 50 mM Tris-HCl, pH 8.0, 20 mM imidazole, and 10 mM βME), and the His-tagged protein was eluted with 10 bed volumes of elution buffer (500 mM NaCl, 50 mM Tris-HCl, pH 8.0, 250 mM imidazole, and 10 mM βME). The N-terminal His tag was cleaved by thrombin digestion during a 24-h dialysis against digestion buffer (500 mM NaCl, 25 mM *N*-Tris(hydroxymethyl)methyl-3-aminopropane sulfonic acid [TAPS], pH 8.5, and 10 mM βME) at 4°C. Cleaved protein was isolated by running the dialyzed sample over another Ni^{2+}-NTA column equilibrated in digestion buffer to remove the His tag and uncleaved protein, followed by a benzamidine–Sepharose column to remove thrombin. A Superdex S200 gel filtration column (Amersham Biosciences, Piscataway, NJ) equilibrated in gel filtration buffer (500 mM NaCl, 25 mM TAPS, pH 8.5, and 2 mM DTT) was used to isolate homogeneous dimeric SAD. Peak fractions were collected and dialyzed against crystallization buffer (100 mM NaCl, 25 mM TAPS, pH 8.5, and 2 mM DTT), concentrated to 18 mg/mL, and stored at −80°C. All SAD mutants were generated using the QuikChange (Stratagene, San Diego, CA) PCR-based method. Mutant enzymes were expressed and purified as described above for wild-type SAD.

Crystallization

SAD crystals were grown overnight by vapor diffusion at 4°C in 2.3-μL drops, consisting of 1 μL of crystallization reservoir (21% [w/v] polyethylene glycol monomethyl ether 5000, 10% [v/v] polyethylene glycol 400, 100 mM 3-(*N*-morpholino)-2-hydroxypropanesulfonic acid [MOPSO], pH 6.0, and 2 mM DTT), 1 μL of protein solution (350 μM SAD, 1.5 mM $NADP^+$, and 1.5 mM sinapaldehyde [in 2-methoxyethanol]), and 0.3 μL of a 2% solution (w/v) of nonyl-β-D-glucopyranoside. Before freezing at 105K, crystals were soaked for 3 min in cryoprotectant (26% [w/v] polyethylene glycol monomethyl ether 5000, 10% [v/v] polyethylene glycol 400, 100 mM MOPSO, pH 6.0, 2 mM DTT, and 15% [v/v] glycerol). Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory beamline 9-2 on a Quantum 4 CCD detector. SAD crystallized in space group P2(1)2(1)2 (a = 76.5 Å, b = 137.8 Å, c = 69.5 Å), with one dimer in the asymmetric unit. Data were indexed, integrated, and scaled to 2.0 Å with HKL2000 (Otwinowski and Minor, 1997). Subsequent single-wavelength data were collected, indexed, integrated, and scaled to 1.65 Å. The space group was P2(1)2(1)2 (a = 76.5 Å, b = 137.8 Å, c = 69.5 Å). The identical procedure was repeated to produce DTT-free crystals by substituting 10 mM βME for 2 mM DTT throughout the purification and crystallization. βME SAD crystallized in space group P2(1)2(1)2 (a = 76.4 Å, b = 138.2 Å, c = 67.6 Å), with one dimer in the asymmetric unit. Data were indexed, integrated, and scaled to 2.5 Å with HKL2000 (Otwinowski and Minor, 1997).

Structure Elucidation

Phase determination was accomplished using SOLVE (Terwilliger and Berendzen, 1999) and multiple anomalous dispersion associated with the four Zn^{2+} ions bound per SAD dimer. After location of the four Zn^{2+} sites and initial multiple anomalous dispersion phasing using SOLVE, density modification to improve and extend the experimental phases allowed automated building using RESOLVE (Terwilliger, 2000). The initial RESOLVE model was manually adjusted and completed by model building in O (Jones et al., 1991) and refined using CNS (Brunger et al., 1998). After preliminary refinement of the initial model, subsequent stages of manual model building and refinement were refined against the high-resolution (1.65 Å) data. The final structure was evaluated with PROCHECK (Laskowski et al., 1996). The SAD/DTT/$NADP^+$ complex had 89.7, 9.9, 0.2, and 0.2% of residues in the most favored, allowed, generously allowed, and disallowed regions of the Ramachandran plot, respectively. The single residue in the disallowed region is Ile-303 of monomer B, which is at the base of the active site in strained contact with Gly-302; electron density at this residue is excellent. Main chain and side chain structural parameters were well above average, with a G factor of 0.34 (Laskowski et al., 1993). The final structural coordinates and structure factors were deposited into the Protein Data Bank under PDB ID 1YQD.

The βME SAD/$NADP^+$ complex had 87.4, 11.9, 0.5, and 0.2% of residues in the most favored, allowed, and generously allowed regions of the Ramachandran plot, respectively. The single residue in the disallowed region is, as described above, Ile-303 of monomer B. Main chain and side chain structural parameters were well above average, with a G factor of 0.26. The final structural coordinates and structure factors were deposited into the Protein Data Bank under PDB ID 1YQX.

Automated Substrate Docking

GOLD software was used for automated docking of ligands into the SAD active site (Jones et al., 1997). GOLD optimizes the fitness score of many possible solutions using a genetic algorithm. Parameters controlling the precise operation of the genetic algorithm were as follows: population size (100), selection pressure (1.100000), number of operations (100,000), number of islands (5), niche size (2), crossover weight (95), mutate weight (95), and migrate weight (10). The default parameter values for van der Waals and hydrogen bonding were used. Because of the expected catalytic mechanism (see Results), all docking calculations imposed a constraint requiring hydrogen bonding of the substrate aldehyde oxygen with the Ser-52 hydroxyl moiety. Fifty docking calculations were run for each ligand, and the GOLDscore was used to identify the lowest energy docking results.

Steady State Kinetic Analysis

SAD assays were performed as described by Kim et al. (2004), with modifications to account for substrate inhibition and strict adherence to the conditions required for Michaelis-Menten kinetic analysis. Reactions consisted of 40 mM 1,3-bis[tris(hydroxymethyl)methylamino]propane buffer, pH 6.25, varied amounts of enzyme in βME crystallization buffer (Table 2), and 400 μM NADPH in a total volume of 200 μL. To ensure adherence to the Michaelis-Menten conditions of steady state kinetic analysis, increasing substrate concentrations were tested to define the substrate concentration at which SAD is saturated. Concentrations of 0 to 800 μM substrate were used for wild-type SAD/sinapaldehyde, wild-type SAD/coniferaldehyde, and L122W/G302F SAD/coniferaldehyde. Concentrations of 0 to 5 mM substrate were used for all other reactions. Enzymatic reactions were initiated by the addition of SAD and incubated at 30°C. The SAD reaction was first evaluated to define the time frame during which product formation is linear with time. Reactions were quenched with 8 μL of glacial acetic acid at 1, 2, and 3 min. One hundred microliters of each assay mixture was injected onto an ODS-AQ reversed-phase column (YMC, Kyoto, Japan; 150 × 4.6 mm i.d.). Separations were performed with a flow rate of 1 mL/min, with detection at 280 nm and the following isocratic solvent system: 8% (v/v) A and 92% (v/v) B (A = acetonitrile, B = 3% [v/v] acetic acid in water) for the first 2 min, followed by a linear A:B gradient from 8% (v/v) A and 92% (v/v) B to 40% (v/v) A and 60% (v/v) B between 2 and 18 min. A product standard curve was generated by plotting the integrated peak area in absorbance units at 280 nm versus a range of authentic product concentrations in micromoles. The concentration of specific reaction products was calculated by dividing the product peak area by the slope (in arbitrary units per micromolar) of the product standard curve. The identity of the reaction products was confirmed by amination of both reaction products and authentic sinapyl and coniferyl alcohol standards followed by quadrupole time-of-flight liquid chromatography–mass spectrometry, to identify the derivatized product and derivatized standard molecular mass. Subsequent tandem mass spectrometry analysis was performed to further confirm that the fragmentation patterns of the reaction products and the corresponding authentic standards were correlated (data not shown). Initial velocities (micromolar per minute) determined from the slope of a product concentration versus time plot were plotted versus substrate concentration and the initial phase of the nonlinear curve directly fitted to the Michaelis-Menten equation using GraphPad Prism version 4.00 for Windows (GraphPad Software, San Diego, CA). The substrate inhibition constant can be estimated from the complete plot of velocity versus substrate concentration.

Sequence data from Figure 5 and this article have been deposited with the EMBL/GenBank data libraries under the following accession numbers: *Populus tremuloides* SAD, PtSAD, AAK58693; *Fragaria ananassa* (strawberry) CAD, FxaCAD, AF320110 (Q9ATW1); *Arabidopsis thaliana* BAD/ELI3-2, AtBAD, At4g37990; *Petroselinum crispum* (parsley) CAD, PcELI3, CAA48028; *Arabidopsis thaliana* CAD6, AtCAD6, At4g37970; *Arabidopsis thaliana* CAD4, AtCAD4, At3g19450; *Arabidopsis thaliana* CAD5, AtCAD5, At4g34230; *Eucalyptus gunnii* CAD2, EuCAD2, P31655; *Populus tremuloides* CAD, PtCAD, Q94G59; *Pinus taeda* CAD, Pt*CAD, P41637 (the asterisk is used to distinguish between *Pinus taeda* and *Populus tremuloides*). Coordinates and structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org) as PDB IDs 1YQD and 1YQX.

ACKNOWLEDGMENTS

We thank Vincent L. Chiang of North Carolina State University for kindly sharing the SAD clone with us. We thank Grace Tong and the Scripps Mass Spectrometry Facility for technical assistance with mass spectrometry analysis and insightful discussions. We thank members of the Noel group and Jean-Luc Ferrer of the European Synchrotron Radiation Facility for assistance and helpful discussions. The findings presented here are based on work supported by the National Science Foundation under Grant 0236027 to J.P.N. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The Stanford Synchrotron Radiation Laboratory Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program. We also acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities.

Received December 6, 2004; accepted March 3, 2005.

REFERENCES

Baker, P.J., Britton, K.L., Rice, D.W., Rob, A., and Stillman, T.J. (1992). Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold: Implications for nucleotide specificity. J. Mol. Biol. **228,** 662–671.

Banfield, M.J., Salvucci, M.E., Baker, E.N., and Smith, C.A. (2001). Crystal structure of the NADP(H)-dependent ketose reductase from *Bemisia argentifolii* at 2.3 A resolution. J. Mol. Biol. **306,** 239–250.

Blanco-Portales, R., Medina-Escobar, N., Lopez-Raez, J.A., Gonzalez-Reyes, J.A., Villalba, J.M., Moyano, E., Caballero, J.L., and Munoz-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.

Brill, E.M., Abrahams, S., Hayes, C.M., Jenkins, C.L., and Watson, J.M. (1999). Molecular characterisation and expression of a wound-inducible cDNA encoding a novel cinnamyl-alcohol dehydrogenase enzyme in lucerne (*Medicago sativa* L.). Plant Mol. Biol. **41,** 279–291.

Brunger, A.T., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. **54,** 905–921.

Clarke, A.R., and Daffron, T.R. (1998). Nicotinamide Cofactor-Dependent Enzymes. (San Diego: Academic Press).

Eszes, C.M., Sessions, R.B., Clarke, A.R., Moreton, K.M., and Holbrook, J.J. (1996). Removal of substrate inhibition in a lactate dehydrogenase from human muscle by a single residue change. FEBS Lett. **399,** 193–197.

Ferrer, J.-L., Zubieta, C., Dixon, R.A., and Noel, J.P. (2005). Crystal structures of alfalfa caffeoyl coenzyme A 3-*O*-methyltransferase. Plant Physiol. **137,** 1009–1017.

Fersht, A. (1977). The basic equations of enzyme kinetics. In Enzyme Structure and Mechanism (New York: W.H. Freeman), pp. 98–120.

Gangloff, A., Garneau, A., Huang, Y.W., Yang, F., and Lin, S.X. (2001). Human oestrogenic 17beta-hydroxysteroid dehydrogenase specificity: Enzyme regulation through an NADPH-dependent substrate inhibition towards the highly specific oestrone reduction. Biochem. J. **356,** 269–276.

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.

Jones, G., Willett, P., Glen, R.C., Leach, A.R., and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. **267,** 727–748.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991).

Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A **47** (Pt 2), 110–119.

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

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26,** 283–291.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR **8,** 477–486.

Lauvergeat, V., Kennedy, K., Feuillet, C., McKie, J.H., Gorrichon, L., Baltas, M., Boudet, A.M., Grima-Pettenati, J., and Douglas, K.T. (1995). Site-directed mutagenesis of a serine residue in cinnamyl alcohol dehydrogenase, a plant NADPH-dependent dehydrogenase, affects the specificity for the coenzyme. Biochemistry **34,** 12426–12434.

Lewis, N.G., and Yamamoto, E. (1990). Lignin: Occurrence, biogenesis and biodegradation. Annu. Rev. Plant Physiol. Plant Mol. Biol. **41,** 455–496.

Li, L., Cheng, X.F., Leshkevich, J., Umezawa, T., Harding, S.A., and Chiang, V.L. (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.

Logemann, E., Reinold, S., Somssich, I.E., and Hahlbrock, K. (1997). A novel type of pathogen defense-related cinnamyl alcohol dehydrogenase. Biol. Chem. **378,** 909–913.

McKie, J.H., Jaouhari, R., Douglas, K.T., Goffner, D., Feuillet, C., Grima-Pettenati, J., Boudet, A.M., Baltas, M., and Gorrichon, L. (1993). A molecular model for cinnamyl alcohol dehydrogenase, a plant aromatic alcohol dehydrogenase involved in lignification. Biochim. Biophys. Acta **1202,** 61–69.

Otwinowski, Z., and Minor, W. (1997). Processing of X-Ray Diffraction Data Collected in Oscillation Mode. (New York: Academic Press).

Persson, B., Zigler, J.S., Jr., and Jornvall, H. (1994). A super-family of medium-chain dehydrogenases/reductases (MDR): Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. Eur. J. Biochem. **226,** 15–22.

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

Rao, S.T., and Rossmann, M.G. (1973). Comparison of super-secondary structures in proteins. J. Mol. Biol. **76,** 241–256.

Riveros-Rosas, H., Julian-Sanchez, A., Villalobos-Molina, R., Pardo, J.P., and Pina, E. (2003). Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. Eur. J. Biochem. **270,** 3309–3334.

Somssich, I.E., Wernert, P., Kiedrowski, S., and Hahlbrock, K. (1996). *Arabidopsis thaliana* defense-related protein ELI3 is an aromatic alcohol:NADP(+) oxidoreductase. Proc. Natl. Acad. Sci. USA **93,** 14199–14203.

Tartof, K.D., and Hobbs, C.A. (1987). Improved media for growing plastid and cosmid clones. Focus **9,** 12.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr. D Biol. Crystallogr. **56,** 965–972.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. **55,** 849–861.

Valencia, E., Larroy, C., Ochoa, W.F., Pares, X., Fita, I., and Biosca, J.A. (2004). Apo and Holo structures of an NADPH-dependent cinnamyl alcohol dehydrogenase from *Saccharomyces cerevisiae*. J. Mol. Biol. **341,** 1049–1062.

Wyrambik, D., and Grisebach, H. (1979). Enzymic synthesis of lignin precursors: Further studies on cinnamyl-alcohol dehydrogenase from soybean-cell-suspension cultures. Eur. J. Biochem. **97,** 503–509.

Zubieta, C., Kota, P., Ferrer, J.L., Dixon, R.A., and Noel, J.P. (2002). Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-*O*-methyltransferase. Plant Cell **14,** 1265–1277.