

Structure and Engineering of L-Arabinitol 4-Dehydrogenase from *Neurospora crassa*

Brian Bae¹†, Ryan P. Sullivan²†, Huimin Zhao^{1,2,3*}
and Satish K. Nair^{1,3*}

¹Department of Biochemistry,
University of Illinois at
Urbana-Champaign, Urbana,
IL 61801, USA

²Department of Chemical and
Biomolecular Engineering,
University of Illinois at
Urbana-Champaign, Urbana,
IL 61801, USA

³Department of Chemistry and
Institute for Genomic Biology,
University of Illinois at
Urbana-Champaign, Urbana,
IL 61801, USA

Received 16 May 2010;
received in revised form
13 July 2010;
accepted 16 July 2010
Available online
22 July 2010

Edited by G. Schulz

L-Arabinitol 4-dehydrogenase (LAD) catalyzes the conversion of L-arabinitol into L-xylulose with concomitant NAD⁺ reduction. It is an essential enzyme in the development of recombinant organisms that convert L-arabinose into fuels and chemicals using the fungal L-arabinose catabolic pathway. Here we report the crystal structure of LAD from the filamentous fungus *Neurospora crassa* at 2.6 Å resolution. In addition, we created a number of site-directed variants of *N. crassa* LAD that are capable of utilizing NADP⁺ as cofactor, yielding the first example of LAD with an almost completely switched cofactor specificity. This work represents the first structural data on any LAD and provides a molecular basis for understanding the existing literature on the substrate specificity and cofactor specificity of this enzyme. The engineered LAD mutants with altered cofactor specificity should be useful for applications in industrial biotechnology.

© 2010 Elsevier Ltd. All rights reserved.

Keywords: L-arabinitol 4-dehydrogenase; crystal structure; substrate specificity; cofactor specificity; mutagenesis

Introduction

The abundance and inexpensive cost of lignocellulosic biomass have led to a surge in research efforts to develop technologies focused on its utility as a source of biofuels and industrial chemicals. Although the pentose sugars D-xylose and L-arabinose represent the most abundant components of hemicellulose in this biomass,^{1,2} the inability of production organisms to effectively harness these pentose sugars represents a significant biotechnological bottleneck.³ For example, the predominant ethanol producer *Saccharomyces cerevisiae* is not able to utilize these hemicellulosic pentose sugars without the introduction of exogenous pathway genes.^{4,5} In contrast, both L-arabinose and D-xylose can be catabolized through a common intermediate (D-xylulose 5-phosphate) via distinct pathways in bacteria⁶ and fungi.⁷

*Corresponding authors. H. Zhao is to be contacted at Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, USA; S. K. Nair, Department of Biochemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, USA. E-mail addresses: zhao5@illinois.edu; snair@illinois.edu.

† B.B. and R.P.S. contributed equally to this work.

Abbreviations used: LAD, L-arabinitol 4-dehydrogenase; ncLAD, *Neurospora crassa* LAD; PDB, Protein Data Bank; ncLAD-S, D211S; ncLAD-SR, D211S/I212R; ncLAD-SRN, D211S/I212R/D213N; ncLAD-3x, D211S/I212R/S348T; psXDH, *Pichia stipitis* xylitol dehydrogenase.

L-Arabinitol 4-dehydrogenase (LAD; EC 1.1.1.12) catalyzes the second step of L-arabinose assimilation by oxidation of L-arabinitol to L-xylulose, with concomitant reduction of NAD^+ to NADH (Fig. 1a).⁸ LAD and the L-xylulose reductase that produces xylitol in the successive step are the two unique enzymes required for L-arabinose utilization, while the other enzymes in the pathway are shared in the D-xylulose pathway.⁹ Due to their importance in cost-effective pentose sugar catabolism, several LADs from different fungal sources have been biochemically characterized, and *Neurospora crassa* LAD (*ncLAD*) was among the most active and stable LADs.¹⁰ *ncLAD* is predominantly NAD^+ dependent, as is the case with all LADs characterized to date, and is a member of the superfamily of medium-chain dehydrogenases/reductases¹¹ with homology to xylitol dehydrogenase and sorbitol dehydrogenase. The preferred substrate is L-arabinitol (Compound 1 in Fig. 1b), with oxidation activity also present with the epimeric sugar alcohols xylitol (Compound 2 in Fig. 1b) and adonitol (Compound 3 in Fig. 1b). In addition, *ncLAD* shows low yet detectable activity with the NADP^+ cofactor, suggesting an inherent relaxed cofactor specificity.¹⁰

Despite recent biochemical characterizations, the lack of structural data on any LAD hinders any description of molecular interaction in cofactor and substrate binding. Although structural data on the eukaryotic orthologue sorbitol dehydrogenase (48% identity over 306 residues) are available,^{12,13} *ncLAD* cannot utilize D-sorbitol as a substrate, suggesting notable differences in active-site architecture between these two enzymes.¹⁰ Furthermore, active-site variants of *ncLAD*, generated based on a homology model derived from the crystal structure of human sorbitol dehydrogenase, failed to demonstrate the expected kinetic parameters, further

reinforcing the distinctions in the active sites of the two enzymes.

To elucidate the determinants of substrate and cofactor recognition, we have determined the first high-resolution X-ray crystallographic structure of LAD determined to a resolution of 2.6 Å. The crystal structure of the *ncLAD*- NAD^+ complex has been successfully utilized for the engineering of variants with altered cofactor specificity for NADP^+ . These studies provide the framework for further engineering of variants with increased activity, which can overcome the bottleneck in pentose catabolism in yeast.

Results

Overall structure

The crystal structure of the *ncLAD*- NAD^+ complex was solved to a resolution of 2.6 Å by the molecular replacement method using the coordinates of human sorbitol dehydrogenase [Protein Data Bank (PDB) ID 1PL8]. The two copies of the complex in the crystallographic asymmetric unit provide independent views of the molecule and allow for a detailed analysis of the structure, independent of crystal packing. Each monomer contains a bidomain architecture composed of a large catalytic domain (residues Ala5 through Val167, and residues Arg308 through Leu362) and a smaller cofactor-binding domain (residues Ala168 through Tyr307), with a large cleft separating the two domains (Fig. 2a). The smaller domain bears the structural elements of a canonical Rossmann fold, and the requisite NAD^+ cofactor is housed within this domain. The larger catalytic domain consists of an α/β fold with an overall structure similar to that

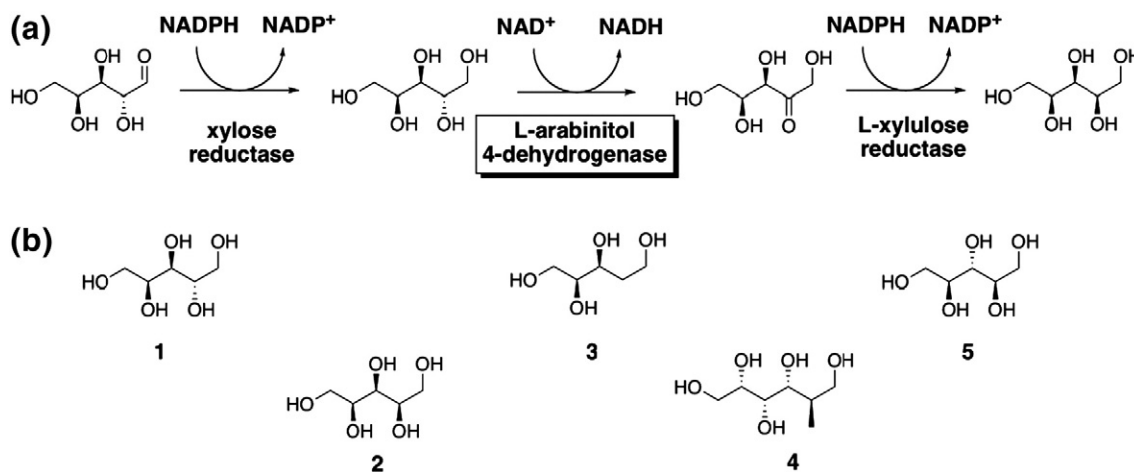


Fig. 1. (a) Pathway for the production of xylitol from L-arabinose in *N. crassa*. Note the cofactor imbalance in enzymes that perform the successive redox reactions along this pathway. (b) Chemical structures of the sugars L-arabinitol (1), xylitol (2), adonitol (3), D-sorbitol (4), and D-arabinitol (5).

found in members of the medium-chain dehydrogenase/reductase family.¹¹ Two bound metal ions (corresponding to zinc atoms as inferred from inductively coupled plasma–mass spectrometry analysis) are associated with the catalytic domain and correspond to the catalytic zinc ion located at

the bottom of the catalytic domain and adjacent to the NAD⁺ cofactor and to a second zinc ion that plays a structural role. Based on inferences from other related structures,¹¹ substrate binding and catalysis likely occur at the cleft between the two domains.

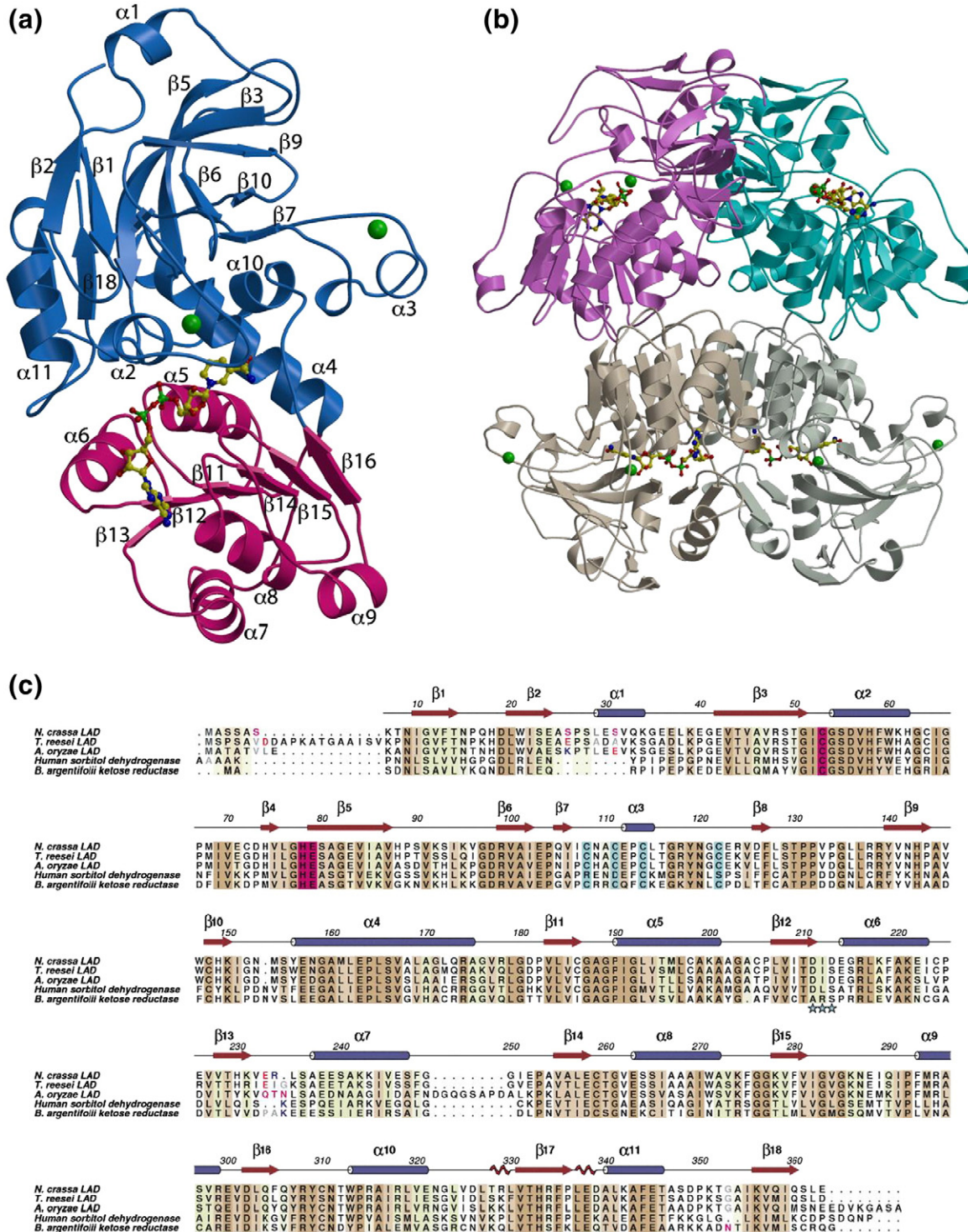


Fig. 2 (legend on next page)

Analytical size-exclusion chromatographic studies demonstrate that *ncLAD* forms a homotetrameric species in solution, consistent with the quaternary structures observed for other polypeptides of the alcohol dehydrogenase fold family.¹⁰ The dimer observed in the crystallographic asymmetric unit corresponds to one-half of the biological tetramer and is related to the other half by the crystallographic 2-fold axis. Consistent with the quaternary structures observed for other members of this structural family, the oligomeric structure may be considered a dimer of identical dimers (Fig. 2b). The two monomers in the crystallographic asymmetric unit interact with the larger interface and bury 2319 Å² of solvent-accessible surface area upon dimerization. In contrast, the interaction of each monomer through the interface created by the crystallographic 2-fold axis buries 1908 Å² of solvent-accessible surface area.

Metal ligands

Prior characterization of *ncLAD* established that the purified enzyme contained two zinc atoms per monomer, consistent with the presence of both a structural metal ion and a catalytic metal ion.¹⁰ The structural zinc ion is situated at a loop region located adjacent to the catalytic domain, where it is ligated by enzyme residues Cys108, Cys111, Cys114, and Cys122. The catalytically requisite zinc ion constitutes the second metal found in each monomer of *ncLAD*. This metal is coordinated by residues Cys53, His78, and Glu79, with a water molecule completing a near-tetrahedral coordination sphere. A conserved glutamic acid residue (Glu163) resides adjacent to the catalytic zinc ion and is hydrogen bonded to the zinc-bound water molecule.

Substrate specificity

Prior studies have established that *ncLAD* can utilize L-arabinitol, xylitol, or adonitol as substrates.¹⁰ Although the k_{cat} values for each of these substrates are fairly similar ($\sim 1000 \text{ min}^{-1}$), the affinity of the enzyme for these substrates varies considerably, with the highest affinity for L-arabinitol ($K_{\text{m}} = 12 \text{ mM}$) and considerably lower affinity for xylitol ($K_{\text{m}} = 280 \text{ mM}$) and adonitol ($K_{\text{m}} = 80 \text{ mM}$). No activity is observed with D-arabinitol (Compound 5 in Fig. 1b) as

substrate. The promiscuity of *ncLAD* towards different substrates is restricted to five-carbon sugars, and no activity is observed towards either D-sorbitol (Compound 4 in Fig. 1b) or D-mannitol. This pattern for substrate preference is similar to that observed for LADs from other sources.¹⁰

A comparison of the active sites of the *ncLAD*-NAD⁺ complex with those of human sorbitol dehydrogenase in complex with the inhibitor CP-166572¹³ provides a molecular rationale for understanding the observed substrate preference pattern (Fig. 3a). We generated a hypothetical model for the binding of L-arabinitol to the enzyme active site utilizing the cocrystal structure of the sorbitol dehydrogenase-CP-166572 complex. Briefly, a linear molecule of L-arabinitol was manually positioned such that the C1 and C2 carbonyl oxygen atoms directly coordinated to the zinc atom and overlapped with the hydroxymethyl nitrogen and oxygen of CP-166572. The rationale for this binding orientation is that polarization of the C2 carbonyl by the zinc is likely to facilitate hydride transfer from the 2-carbon to NAD⁺, facilitating the oxidation of L-arabinitol to L-xylulose. The carbon atoms were similarly aligned with the atoms in the pyrimidine and piperazine rings of the inhibitor. The resultant model suggests that hydrogen bonding between Ser55 and the C4 hydroxyl and electrostatic stabilization of the C3 hydroxyl by Arg308 additionally stabilize the binding of the substrate. These interactions are lost with either xylitol or adonitol, explaining the higher K_{m} values observed with these other five-carbon sugar substrates.

A number of residues, including Ile65, Tyr307, and Arg308, flank the putative substrate binding site above the cofactor, where steric clashes between these residues and the NAD⁺ cofactor preclude binding of D-arabinitol, thereby confining binding to only the L-enantiomers of five-carbon sugars (Fig. 3b). At the outer end of the active site, residues Ile65, Met68, and Val284 protrude inwards and cap the binding site. In enzymes that can utilize sugar substrates of length greater than five carbons, Met68 is typically a phenylalanine that is directed away from the active site (e.g., Phe59 in sorbitol dehydrogenase). Moreover, the loop region bearing Met68 is extended outwards in both human sorbitol dehydrogenase and whitefly ketose reductase, resulting in a large active-site cavity that can accommodate

Fig. 2. (a) Ribbon diagram for the overall structure of *ncLAD* with bound NAD⁺ cofactor. The catalytic domain housing both catalytic and structural zinc ions is shown in blue, and the cofactor-binding domain is shown in magenta. The two zinc atoms are shown as green spheres, and the bound NAD⁺ cofactor is shown as a yellow ball-and-stick model. (b) Structure of the biological *ncLAD* tetramer, with the two monomers comprising the crystallographic asymmetric unit shown in pink and cyan, and with the two monomers related by 2-fold symmetry shown in gray and tan. (c) Structure-based multiple-sequence alignments of LAD from various organisms, along with the structurally related human sorbitol dehydrogenase and whitefly ketose reductase. Sequence conservation is shown with identical residues in brown, highly conserved residues in tan, and similar residues in olive green. The residues comprising the catalytic zinc ion ligands are shown in magenta, and the residues binding the structural zinc ion are shown in cyan. The residues thought to confer cofactor specificity are shown with stars.

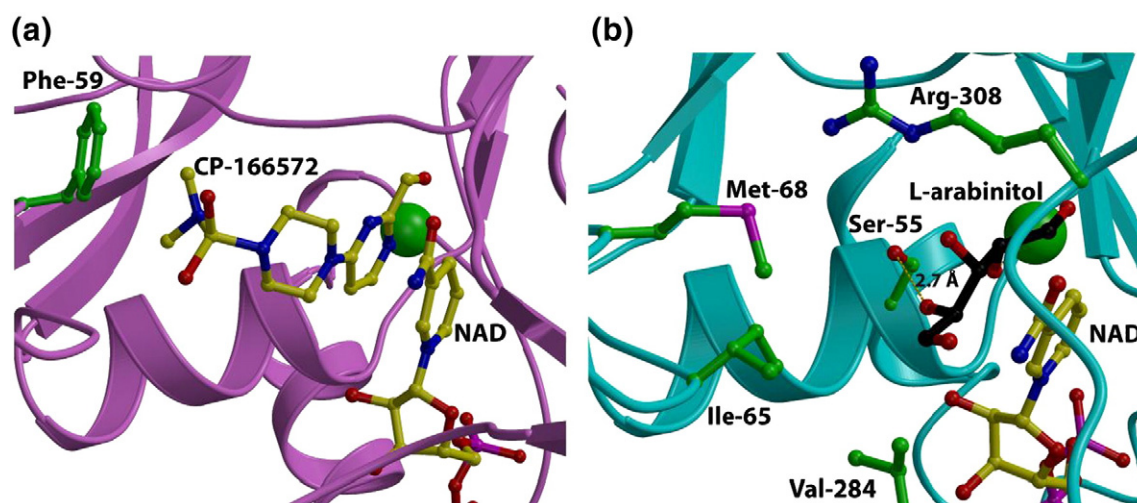


Fig. 3. (a) Crystal structure of human sorbitol dehydrogenase¹³ bound to the competitive inhibitor CP-166572. The cofactor and the inhibitor are shown as yellow sticks, and residue Phe59 capping the active-site pocket is shown in green sticks. (b) Structure of a hypothetical model of *ncLAD* with bound L-arabinitol. The cofactor is shown as yellow sticks, and the modeled substrate is shown in black sticks. Ser55 likely hydrogen bonds to the substrate, and Ile65, Met68, and Val284 occlude the active-site pocket to accommodate only five-carbon sugar substrates.

substrates such as sorbitol and the inhibitor CP-166572. The protrusion of the loop region bearing Met68, as well as that of the side chains of Ile65 and Val284, restricts the active-site cavity such that only smaller five-carbon sugar substrates can be accommodated without significant steric clashes (Fig. 3b).

Cofactor specificity

In each monomer of *ncLAD*, clear electron density corresponding to the bound NAD⁺ cofactor is observed within the smaller nucleotide-binding domain, along a crevice between the two domains (Fig. 4a). The adenine ring is nestled in a shallow pocket created by numerous hydrophobic residues, including Ile212, Val232, Thr260, and Val262. The base of this pocket is lined by strands β 11 and β 12, and is flanked by strand β 14. Both of the ribose sugars are in the 2' endo pucker, similar to that observed in the cocrystal structure of human sorbitol dehydrogenase, with the nicotinamide ring in *anti* conformation. The nicotinamide is adjacent to the catalytic zinc ion, where it is poised for hydride transfer to the C2 atom of the substrate.

The molecular determinants for cofactor specificity in *ncLAD* are consistent with those described for other members of the medium-chain dehydrogenase/reductase family. The most significant contribution to cofactor specificity is the presence of an aspartic acid residue (Asp211), located within the loop that connects β 12 and α 6, which forms a hydrogen bond to the 2' and 3' hydroxyls of the adenine ribose (Fig. 4b). In dehydrogenases that utilize NADP⁺ as cofactor, this residue is truncated to a small uncharged residue to accommodate the 2'-

phosphate. Concurrently, an arginine residue is also located adjacent to this residue, where it can form a positive binding pocket. For example, in whitefly ketose reductase, Asp211/Ile212 are replaced by an alanine (Ala199) and an arginine (Arg200), providing both a steric binding pocket and an electrostatic binding pocket for the 2'-phosphate of the preferred NADP⁺ substrate^{14,15} (Fig. 4c). Additional stabilization of the phosphate oxygens is provided by Arg204 and the backbone nitrogens of Gly178, Arg200, and Ser201.

Mutagenesis and kinetics

Based on similar attempts to alter the cofactor specificity of homologous enzymes, coupled with homology modeling¹⁰ and eventual determination of the 2.6-Å-resolution crystal structure of the *ncLAD*-NAD⁺ complex, the *ncLAD* enzyme was subjected to site-directed mutagenesis and site saturation mutagenesis, followed by a round of error-prone PCR library screening to identify variant NADP⁺-dependent LADs. The kinetic parameters of the parent wild-type and engineered mutant *ncLAD* enzymes for NAD⁺ and NADP⁺ are listed in Table 1. Three rounds of rational design were implemented to target residues Asp211, Ile212, and Asp213. The best first-round rational design mutant D211S (*ncLAD*-S) showed a dramatic decrease in activity towards NAD⁺, with a minimal yet detectable activity increase towards NADP⁺. The best second-round rational design mutant D211S/I212R (*ncLAD*-SR) displayed a significant reversal in cofactor specificity, although it still had ~5-fold lower k_{cat}/K_m than the wild type had

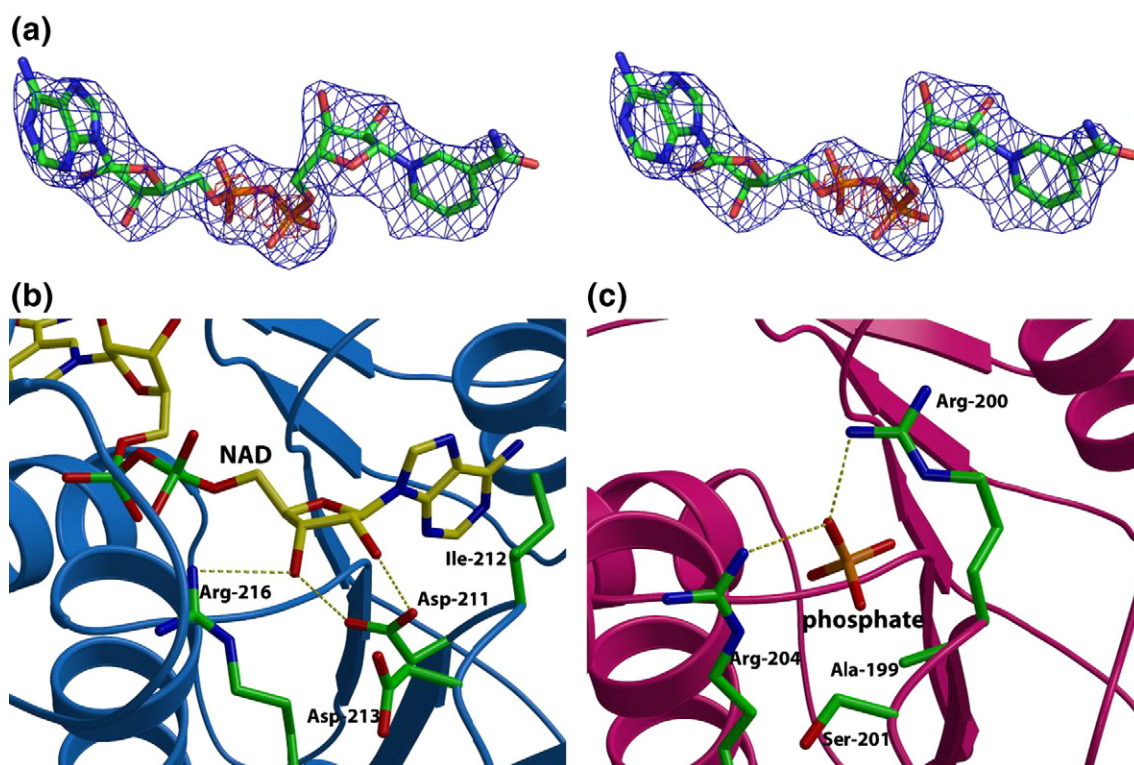


Fig. 4. (a) Difference Fourier electron density maps (blue, contoured at 3σ over background; red, contoured at 10σ over background) calculated with coefficients $F_{\text{obs}} - F_{\text{calc}}$ at the cofactor binding site, measured from phases in the final refined model with the coordinators of the cofactor deleted prior to one round of crystallographic refinement. The final coordinates of the NAD cofactor are superimposed. (b) The cofactor binding site of *ncLAD* with bound NAD⁺ is shown in yellow sticks, and residues that confer specificity are shown as green sticks. (c) Similar view of the active site of whitefly ketose reductase,¹⁴ which utilizes an NADP⁺ cofactor. This structure does not contain a bound ligand, but a phosphate ion is situated at the likely position of the 2'-phosphate of the cofactor.

with NAD⁺. The third round resulted in finding mutant D211S/I212R/D213N (*ncLAD*-SRN), but precipitation during purification hampered an accurate determination of kinetics. Saturation mutagenesis and screening of surrounding residues Thr210, Glu214, and Gly215 did not result in finding any further improved mutants (data not shown).

With rational design targets exhausted, a round of error-prone PCR on the *ncLAD*-SR parent was

attempted. The mutation rate was approximately one to two amino acids per mutant enzyme, and screening a library of ~ 5000 mutants via a 96-well plate format resulted in the discovery of D211S/I212R/S348T (*ncLAD*-3x), which contained a third mutation S348T, and showed improved affinity for NADP⁺. In addition, activity with NAD⁺ was kept low, being only slightly above background activity. This result was promising, as the goal of the engineering effort was not only to improve activity towards NADP⁺ but also to

Table 1. Kinetic parameters of LAD mutants for nicotinamide cofactors

Enzyme	NAD ⁺			NADP ⁺		
	k_{cat} (min ⁻¹)	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ (mM ⁻¹ min ⁻¹)	k_{cat} (min ⁻¹)	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ (mM ⁻¹ min ⁻¹)
<i>ncLAD</i> -wt	650±25	0.14±0.03	4740	— ^a	>8	<30
<i>ncLAD</i> -S	515±21	2.9±0.3	175	— ^a	>5	<40
<i>ncLAD</i> -SR	191±17	3.6±0.7	53	494±26	0.48±0.08	1030
<i>ncLAD</i> -SRN	— ^a	>5	<100	714±8	1.45±0.05	496
<i>ncLAD</i> -3x	— ^a	>5	<100	1210±44	0.55±0.09	2190

All assays were performed in triplicate at 25 °C in 50 mM Tris-HCl (pH 8.0).

All enzymes were purified and characterized with N-His₆ tag.

^a Not determined (saturation of cofactor not reached).

Table 2. Kinetic parameters of LAD mutants for L-arabinitol substrate

Enzyme	k_{cat} (min ⁻¹)	K_m (mM)	k_{cat}/K_m (mM ⁻¹ min ⁻¹)
<i>ncLAD</i> -wt	736±110	11.6±0.9	64
<i>ncLAD</i> -S	— ^a	>1000	— ^a
<i>ncLAD</i> -SR	— ^a	>500	— ^a
<i>ncLAD</i> -3x	1190±390	292±110	4

All assays were performed in triplicate at 25 °C in 50 mM Tris-HCl (pH 8.0).

All enzymes were purified and characterized with N-His₆ tag.

^a Not determined (saturation of cofactor not reached).

minimize activity with NAD⁺ in order to potentially alleviate the cofactor imbalance of the wild-type arabinose catabolic pathway. However, as shown in Table 2, an unfortunate result of the screening procedure was the subsequent loss of affinity for substrate L-arabinitol of the mutant *ncLAD*-3x ($K_{m/L\text{-arabinitol}} \sim 290$ mM) compared to the wild type ($K_{m/L\text{-arabinitol}} \sim 12$ mM). As a result, catalytic efficiency is also substantially reduced from 64 mM⁻¹ min⁻¹ for the wild type to 4 mM⁻¹ min⁻¹ for *ncLAD*-3x. This loss of affinity and catalytic efficiency came about despite keeping the L-arabinitol concentration of the screening assay at 25 mM, which is near the wild-type $K_{m/L\text{-arabinitol}}$ value. Intermediate mutants *ncLAD*-S and *ncLAD*-SR could not be saturated for sufficient substrate kinetic parameter analysis, and substrate kinetics were not attempted with *ncLAD*-SRN.

Discussion

The overall structure of *ncLAD* is similar to that of human sorbitol dehydrogenase (rmsd of 1.5 Å over 344 α carbon atoms; Z-score=49.1), as anticipated from the 48% sequence identity between the two polypeptides.^{12,13} The overall structure is similar to those of other members of the zinc-dependent alcohol dehydrogenase family, including *Bemisia argentifolii* (whitefly) ketose reductase (rmsd of 1.7 Å over 343 α carbon atoms; Z-score=47.3; 41% sequence identity)^{14,15} and L-threonine 3-dehydrogenase (rmsd of 2.1 Å over 335 α carbon atoms; Z-score=41.5; 30% sequence identity)¹⁶ (see Fig. 2c for a sequence alignment). Although the overall monomeric structure of *ncLAD* is similar to those of other members of the zinc-dependent alcohol dehydrogenase family, superposition of the three-dimensional structures is not precise. This is a consequence of deviations in the position of secondary structural elements within each domain and of changes in the relative orientations of each of the two domains between *ncLAD* and other members of the medium-chain dehydrogenase/reductase structural family. These notable changes

in orientation explain the failure of homology models derived from the human sorbitol dehydrogenase structure to serve as an accurate template for the selection of site-directed variants of *ncLAD*.¹⁰

The structural zinc ion is found only in a subset of the medium-chain dehydrogenase/reductase family members, including bacterial alcohol dehydrogenases and whitefly ketose reductase.¹¹ Although the orientation of the loop harboring the structural zinc ion in *ncLAD* is nearly identical with that in human sorbitol dehydrogenase,^{12,13} the latter enzyme lacks the structural metal, and three of the residues that bind the structural metal in *ncLAD* are not conserved in sorbitol dehydrogenase. Interestingly, this loop region is involved in intersubunit contacts that mediate oligomerization, and the presence of a stabilizing metal within this loop may play a role in the previously noted stability of *ncLAD*. For example, *Pichia stipitis* xylitol dehydrogenase (*psXDH*) does not contain a structural zinc ion, and introduction of cysteine residues into the corresponding loop region results in an enzyme that contains an additional zinc binding site and demonstrates improved thermostability over the wild type.¹⁷

The catalytic zinc ion is coordinated in a tetrahedral configuration by three residues (Cys53, His78, and Glu79) and a water molecule that is hydrogen bonded to a conserved glutamic acid residue, Glu163. A glutamic acid residue is analogously located in whitefly ketose reductase and human sorbitol dehydrogenase, and this residue is proposed to transiently coordinate to the zinc atom in order to facilitate product release following catalysis. Mutation of the similar Glu155 in rat sorbitol dehydrogenase results in a 500-fold increase in K_m for the substrate, consistent with the role of this residue in facilitating product release.

There is a large body of literature describing the alteration of nicotinamide cofactor specificity,^{17–26} including two review articles outlining the typical determinants and evolution of nicotinamide binding sites.^{27,28} Yet despite the prevalence of these attempts, altering cofactor specificity remains a challenge, as there are very few examples where catalytic efficiency for the initially disfavored cofactor has been improved such that it reaches the same level of efficiency as with the physiological cofactor. Comparison of the strategies required to achieve efficient use of the nonphysiological cofactor in these enzymes indicates that there is no clear formula for success, as mutations found in some enzymes do not necessarily translate to the same activity effect in another template enzyme. The cofactor specificity of *ncLAD* is due to the significant contributions of Asp211 and Ile212, which limit space and electrostatic interactions with the phosphate moiety of NADP⁺. These

structural rationales for cofactor specificity have been further borne out through studies of engineered variants of NAD⁺-dependent *ps*XDH, designed to utilize NADP⁺ as cofactor.¹⁷ A triple mutant of *ps*XDH, consisting of replacement of the negatively charged aspartate (Asp207→Ala), introduction of a proximal arginine residue (Ile208→Arg) to accommodate the 2'-phosphate, and an additional cavity creating mutation (Phe209→Ser), results in a reversal of cofactor preference. Likewise, an analogous double mutant of NAD⁺-dependent *Gluconobacter oxydans* xylitol dehydrogenase (with mutations Asp38→Ser/Met39→Arg) exhibits an exclusive preference for NADP⁺ and shows a higher cofactor affinity ($K_m=206\ \mu\text{M}$ for NADP⁺) than the wild-type enzyme has for the natural cofactor ($K_m=348\ \mu\text{M}$ for NAD⁺).²⁹ Although short-chain dehydrogenases, such as *G. oxydans* xylitol dehydrogenase, are structurally and mechanistically distinct from *nc*LAD and other medium-chain dehydrogenases, the determinants for cofactor specificity are highly conserved.

These studies with homologous enzymes, coupled with the 2.6-Å-resolution crystal structure of the *nc*LAD-NAD⁺ complex, provided a path for engineering *nc*LAD for reversed cofactor specificity. A triple mutant of *nc*LAD (Asp211→Ala/Ile212→Arg/Asp213→Ser) was constructed based on the *ps*XDH framework, but the resultant protein failed to express in appreciable quantities to permit biochemical analysis. Consequently, a systematic analysis of various single, double, and triple mutants was carried out in order to derive an appropriate variant with switched cofactor specificity. Site-directed mutagenesis and site saturation mutagenesis approaches were utilized to identify a double mutant, Asp211→Ser/Ile212→Arg (*nc*LAD-SR), with expression level and solubility comparable to those of the wild-type enzyme.

Modeling of NADP⁺ binding, based on the crystal structure of the *nc*LAD-NAD⁺ complex, suggests that the negatively charged Asp213 would cause electrostatic repulsion with the 2'-phosphate. Mutational analysis at this residue was pursued in the context of the *nc*LAD-SR mutant. Saturation mutagenesis libraries of Asp213 were created and screened for activity, and mutants with a >25% activity increase above the parent templates were further analyzed. The majority of clones with increased activity contained an asparagine at residue 213, and the triple mutant Asp211→Ser/Ile212→Arg/Asp213→Asn (*nc*LAD-SRN) was further characterized. Kinetic analysis of this mutant was consistent with a 3-fold loss in affinity for NADP⁺. However, as this triple variant was less stable than the *nc*LAD-SR parent enzyme and prone to aggregation during purification, the kinetic parameters for this variant likely represent an underestimation of the true catalytic efficiency,

consistent with the increased activity measured during the initial screening.

In order to further improve on the results of our rational engineering experiments, we subjected the *nc*LAD-SR double mutant to an error-prone PCR library screen. This approach identified a third mutation (Ser348→Thr) that improved upon the solubility and activity of the *nc*LAD-SR template. This triple mutant *nc*LAD-3x shows a clear switch in cofactor preference, with a catalytic efficiency that is nearly 20-fold higher with NADP⁺ than with NAD⁺. While the effects of Asp211→Ser and Ile212→Arg can be understood in the context of the structural data, the results of the Ser348→Thr mutation are more difficult to rationalize. Within the cocrystal structure, Ser348 is situated nearly 15 Å away from the active site and does not make any direct contact with residues implicated in catalysis. Additionally, this residue is 13 Å away from the C2' hydroxyl of the NAD⁺, which would bear the phosphate group in the alternate cofactor NADP⁺.

The increased catalytic efficiency of the Ser348→Thr variant is likely a consequence of modest alterations in the secondary structure in the context of the Asp211→Ser/Ile212→Arg double mutation. Residue Ser348 is located at the end of helix α 11, which is located almost perpendicularly to the loop that harbors Asp211 and Ile212. In both human sorbitol dehydrogenase and whitefly ketose reductase, the equivalent residue is either an arginine or a lysine that helps stabilize the orientation of helix α 11. Replacement of Ile212 with a large arginine side chain would result in a protrusion of this side chain along the cofactor binding cleft, where it may disrupt the orientation of helix α 11. The Ser348→Thr mutation likely provides a compensatory stabilization of this helix in the context of the cofactor specificity altering mutations. The improvement of *nc*LAD-3x over the parent *nc*LAD-SR template is due mostly to an increase in k_{cat} , consistent with our proposal that the Ser348→Thr mutation likely affects secondary structure stability to facilitate improved turnover.

Although the *nc*LAD-3x triple mutant exhibits a clear switch in cofactor preference, the affinity of this variant for substrate L-arabinitol is significantly diminished ($K_m=290\ \text{mM}$ for L-arabinitol) relative to wild type ($K_m=12\ \text{mM}$ for L-arabinitol). Consequently, the catalytic efficiency is decreased by nearly 20-fold relative to the wild type, in spite of the fact that the L-arabinitol concentrations were maintained at 25 mM during the screening assays, in the range close to the K_m of the wild-type enzyme. An examination of the *nc*LAD-NAD⁺ cocrystal structure in the context of the mutations suggests that the protrusion of Arg212 along the cofactor binding cleft likely precludes binding of the substrate L-arabinitol at the active site.

In conclusion, this work represents the first reported structural data of an LAD enzyme, which was used to provide a molecular understanding of the cofactor specificity of the novel engineered NADP⁺-dependent LAD variants. Through this work, the effects of the consolidation of the cofactor usage of the initial pathway enzymes in the L-arabinose catabolic pathway are currently being investigated.

Materials and Methods

Materials

Escherichia coli BL21(DE3) and pET-28a were purchased from Novagen (Madison, WI). Kanamycin, isopropyl- β -D-thiogalactopyranoside (IPTG), NAD, NADP, NADH, and NADPH were obtained from Sigma-Aldrich (St. Louis, MO). Other salts and reagents were purchased from either Fisher (Pittsburg, PA) or Sigma-Aldrich. Co²⁺ Talon™ immobilized metal affinity resin was purchased from Clontech BD Biosciences (San Jose, CA). Amicon® Ultra-15 filter devices were purchased from Millipore. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). Kits for plasmid, gel, and column purification of DNA fragments were obtained from Qiagen (Valencia, CA). DNA oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA).

Protein expression and purification

The construction of an expression vector for the heterologous production of *ncLAD* in *E. coli* has been previously described.¹⁰ *E. coli* expression strain BL21(DE3) was transformed with this expression vector, and single colonies of transformed *E. coli* were used to inoculate 5 ml of LB medium supplemented with kanamycin (50 μ g/ml). Five hours following inoculation, the small-scale culture was added to 1 L of LB medium containing kanamycin (50 μ g/ml) and 0.5 mM ZnCl₂ and grown at 37 °C. When the OD₆₀₀ of the culture had reached 0.3, protein expression was induced with 0.1 mM IPTG, and the cells were further grown for 18 h at 30 °C. Bacterial cells were pelleted by centrifugation (4000 rpm for 30 min) and resuspended in 100 mM KCl, 20 mM Tris-HCl (pH 8.3), 10% glycerol, and a cocktail of protease inhibitors. Resuspended cells were disrupted by multiple passes through an Avestin C5 Emulsiflex French press cell, and insoluble aggregates and cellular debris were removed by centrifugation (15,000 rpm for 1 h).

Recombinant proteins were purified from the clarified supernatant by virtue of the amino terminal polyhistidine tag using a Talon resin (Clontech BD Biosciences) column charged with cobalt chloride. Following elution from the cobalt affinity resin, the cleavable polyhistidine tag was removed using thrombin (1 U/mg protein; GE Healthcare). The protein was further purified by anion-exchange chromatography (5 ml of HiTrap Q; GE Healthcare) and size-exclusion chromatography (Superdex 75 16/60; GE Healthcare) prior to crystallization.

Crystallization and X-ray data collection

Initial crystallization conditions were established by sparse-matrix sampling methods using commercial screens. Refinement of promising conditions yielded large crystals suitable for diffraction analysis. Crystals of the *ncLAD*-NAD⁺ complex were grown using hanging-drop vapor diffusion. A 2- μ l drop containing the preincubated sample [25 mg/ml *ncLAD* in 100 mM KCl, 10 mM Hepes (pH 7.5), and 10 mM NAD⁺] was added to 2 μ l of precipitant (1.8–2 M ammonium sulfate) and equilibrated over a well containing the precipitant solution at 8 °C. Crystals grew within a week and reached a maximum size of 0.2 mm \times 0.4 mm \times 0.4 mm. Crystals were briefly soaked in the precipitant solution supplemented with 25% glycerol and vitrified by being plunged directly into liquid nitrogen prior to data collection.

Phasing and structure determination

Crystals of *ncLAD*-NAD⁺ occupied space group *P*6₄22 with unit cell parameters $a=135.04$ Å, $b=135.04$ Å, and $c=225.08$ Å, and contained two molecules of the complex in the crystallographic asymmetric unit. Diffraction data were collected to a limiting resolution of 2.6 Å at an insertion device line (LS-CAT-Sector 21ID-D; Advanced Photon Source, Argonne, IL), integrated, and scaled using the HKL2000 package.³⁰ An 11-fold redundant data set was collected to a limiting resolution of 2.6 Å (overall $R_{\text{merge}}=0.094$; $I/\sigma(I)=2.9$ in the highest resolution shell). Crystallographic phases were determined by the molecular replacement method³¹ using the refined coordinates of human sorbitol dehydrogenase (PDB ID 1PL8; 48% identity over 306 residues),^{12,13} omitting all solvent molecules and bound ligands.

Following rigid-body refinement of the initial molecular replacement solution, the atomic model was subjected to automatic rebuilding using ARP/wARP,³² resulting in placement of roughly 80% of the main-chain atoms and 60% of the side-chain atoms. Clear density for metal ions and bound ligand could be observed prior to any crystallographic refinement. The remainder of the model was fitted using Coot³³ and further improved by rounds of refinement with REFMAC5.^{34,35} Cross-validation used 5% of the data in the calculation of R_{free} .³⁶ Metal ions, ligands, and solvent molecules were built into the model only after the R_{free} had dropped to below 0.35. The stereochemistry of the models was routinely monitored throughout the course of refinement using PROCHECK,³⁷ and the final model was validated using MolProbity.^{38,39} The final model consists of 2 molecules of the LAD-NAD⁺ complex, 4 zinc ions, and 146 solvent molecules, and has been refined to an R -factor/ R_{free} of 0.21/0.26. Crystal parameters, data collection parameters, and refinement statistics for each of the structures are summarized in Table 3. The refined coordinates have been deposited in the PDB with accession number 3M6I.

Mutagenesis and screening assay

Random mutagenesis was carried out using error-prone PCR.⁴⁰ To identify LAD mutants that showed improved NADP⁺ activity while keeping NAD⁺ activity limited, we

Table 3. Data collection, phasing, and refinement statistics

	LAD-NAD ⁺ complex
<i>Data collection</i>	
Space group	P6 ₄ 22
Unit cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	134.5, 134.5, 224.3
Resolution (Å)	50–2.6 (2.69–2.6) ^a
<i>R</i> _{sym} (%)	9.4 (63.2)
<i>I</i> / σ (<i>I</i>)	13.3 (2.9)
Completeness (%)	96.3 (92.0)
Redundancy	10.6 (8.6)
<i>Refinement</i>	
Resolution (Å)	25.0–2.6
Number of reflections	33,999
<i>R</i> _{work} / <i>R</i> _{free} ^b (%)	20.8/26.6
Number of atoms	
Protein	2864
Solvent	146
NAD	92
Average <i>B</i> -value	
Protein	60.6
Solvent	54.1
NAD	54.7
rmsd	
Bond lengths (Å)	0.013
Bond angles (°)	1.56

^a The highest-resolution shell is shown in parentheses.

^b R -factor = $\sum(|F_{\text{obs}}| - k|F_{\text{calc}}|) / \sum|F_{\text{obs}}|$, and *R*_{free} is the *R* value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.

developed a screening procedure based on a 96-well plate. Library colonies were selected with sterile toothpicks and placed in single wells to inoculate 96-well plates containing 100 μ l of LB medium supplemented with 50 μ g/ml kanamycin. The plates were incubated at 37 °C for 5 h with shaking (250 rpm). Cell cultures were then induced with IPTG (final concentration of 0.3 mM IPTG), followed by incubation at 30 °C for 12–16 h. After protein expression, the plates were centrifuged at 4000 rpm for 15 min, and the medium was decanted. Cell pellets were lysed by resuspension in 100 μ l of lysis buffer consisting of 100 mM phosphate buffer (pH 7.0) with 1 mg/ml lysozyme. The 96-well plates were sealed with adhesive labels and vortexed until the cell pellets had uniformly resuspended; this was then followed by two freeze–thaw steps. Ninety-six-well plates containing cell lysates were then centrifuged at 4000 rpm at 4 °C for 15 min, and 20 μ l of cell lysate was transferred to a fresh 96-well plate. Addition of 100 μ l of activity assay mix, comprising 100 mM KH₂PO₄ buffer (pH 7.0), 25 mM L-arabinitol, and 0.125 mM NADP⁺, was immediately followed by monitoring of the change in absorbance at 340 nm for 5 min in a SpectraMax 340PC microplate reader (Molecular Devices) to determine the initial activities of each library mutant. Positive mutants were verified in a scaled-up tube assay, which also served to determine residual NAD⁺ activity.

Enzyme kinetics

The kinetic rate constants for the purified N-His₆-tagged wild-type and mutant LAD enzymes were determined as

described elsewhere.¹⁰ Briefly, the initial rates of reaction were determined by monitoring the increase in absorbance of NAD(P)H at 340 nm. Initial rates were measured in triplicate at 25 °C using a Cary UV 100 Bio UV–Vis spectrophotometer (Varian). Reactions were initiated by addition of ~1–20 μ g of purified LAD enzyme with N-terminal His₆ tag. Enzyme concentrations were determined using the Bradford Protein Assay (Bio-Rad), with bovine serum albumin (Sigma-Aldrich) as standard. For determination of kinetics, L-arabinitol or NAD(P)⁺ concentrations were varied from below the *K*_m value to at least five times higher than the *K*_m value, with saturating concentrations of the other substrate/cofactor present. The data were fitted to the Michaelis–Menten equation, using nonlinear curve fitting with least squares regression analysis in OriginPro 8 (OriginLab Corporation), to determine the *k*_{cat} and *K*_m values.

Accession numbers

Coordinates and structure factors have been deposited in the PDB with accession number 3M6I.

Acknowledgements

This work was supported by the Biotechnology Research and Development Consortium (Project 1-4-158).

References

- Lee, J. (1997). Biological conversion of lignocellulosic biomass to ethanol. *J. Biotechnol.* **56**, 1–24.
- van Maris, A. J., Abbott, D. A., Bellissimi, E., van den Brink, J., Kuyper, M., Luttik, M. A. *et al.* (2006). Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. *Antonie Van Leeuwenhoek*, **90**, 391–418.
- Margeot, A., Hahn-Hagerdal, B., Edlund, M., Slade, R. & Monot, F. (2009). New improvements for lignocellulosic ethanol. *Curr. Opin. Biotechnol.* **20**, 372–380.
- Barnett, J. A. (1976). The utilization of sugars by yeasts. *Adv. Carbohydr. Chem. Biochem.* **32**, 125–234.
- Matsushika, A., Inoue, H., Kodaki, T. & Sawayama, S. (2009). Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Appl. Microbiol. Biotechnol.* **84**, 37–53.
- Jennings, D. H. (1984). Polyol metabolism in fungi. *Adv. Microb. Physiol.* **25**, 149–193.
- Dien, B. S., Cotta, M. A. & Jeffries, T. W. (2003). Bacteria engineered for fuel ethanol production: current status. *Appl. Microbiol. Biotechnol.* **63**, 258–266.
- Richard, P., Londesborough, J., Putkonen, M., Kalkkinen, N. & Penttila, M. (2001). Cloning and expression of a fungal L-arabinitol 4-dehydrogenase gene. *J. Biol. Chem.* **276**, 40631–40637.
- Chiang, C. & Knight, S. G. (1960). A new pathway of pentose metabolism. *Biochem. Biophys. Res. Commun.* **3**, 554–559.

10. Sullivan, R. & Zhao, H. (2007). Cloning, characterization, and mutational analysis of a highly active and stable L-arabinitol 4-dehydrogenase from *Neurospora crassa*. *Appl. Microbiol. Biotechnol.* **77**, 845–852.
11. Eklund, H. & Ramaswamy, S. (2008). Medium- and short-chain dehydrogenase/reductase gene and protein families: three-dimensional structures of MDR alcohol dehydrogenases. *Cell. Mol. Life Sci.* **65**, 3907–3917.
12. Johansson, K., El-Ahmad, M., Kaiser, C., Jorvall, H., Eklund, H., Hoog, J. & Ramaswamy, S. (2001). Crystal structure of sorbitol dehydrogenase. *Chem. Biol. Interact.* **130–132**, 351–358.
13. Pauly, T. A., Ekstrom, J. L., Beebe, D. A., Chrunchy, B., Cunningham, D., Griffor, M. *et al.* (2003). X-ray crystallographic and kinetic studies of human sorbitol dehydrogenase. *Structure*, **11**, 1071–1085.
14. Banfield, M. J., Salvucci, M. E., Baker, E. N. & Smith, C. A. (2001). Crystal structure of the NADP(H)-dependent ketose reductase from *Bemisia argentifolii* at 2.3 Å resolution. *J. Mol. Biol.* **306**, 239–250.
15. Wolfe, G. R., Smith, C. A., Hendrix, D. L. & Salvucci, M. E. (1999). Molecular basis for thermoprotection in *Bemisia*: structural differences between whitefly ketose reductase and other medium-chain dehydrogenases/reductases. *Insect Biochem. Mol. Biol.* **29**, 113–120.
16. Ishikawa, K., Higashi, N., Nakamura, T., Matsuura, T. & Nakagawa, A. (2007). The first crystal structure of L-threonine dehydrogenase. *J. Mol. Biol.* **366**, 857–867.
17. Watanabe, S., Kodaki, T. & Makino, K. (2005). Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *J. Biol. Chem.* **280**, 10340–10349.
18. Scrutton, N. S., Berry, A. & Perham, R. N. (1990). Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature*, **343**, 38–43.
19. Mittl, P. R. E., Berry, A., Scrutton, N. S., Perham, R. N. & Schulz, G. E. (1993). Structural differences between wild-type NADP-dependent glutathione-reductase from *Escherichia coli* and a redesigned NAD-dependent mutant. *J. Mol. Biol.* **231**, 191–195.
20. Metzger, M. H. & Hollenberg, C. P. (1995). Amino acid substitutions in the yeast *Pichia stipitis* xylitol dehydrogenase coenzyme-binding domain affect the coenzyme specificity. *Eur. J. Biochem.* **228**, 50–54.
21. Ehsani, M., Fernandez, M. R., Biosca, J. A. & Dequin, S. (2009). Reversal of coenzyme specificity of 2,3-butanediol dehydrogenase from *Saccharomyces cerevisiae* and *in vivo* functional analysis. *Biotechnol. Bioeng.* **104**, 381–389.
22. Zeng, Q. K., Du, H. L., Wang, J. F., Wei, D. Q., Wang, X. N., Li, Y. X. & Lin, Y. (2009). Reversal of coenzyme specificity and improvement of catalytic efficiency of *Pichia stipitis* xylose reductase by rational site-directed mutagenesis. *Biotechnol. Lett.* **31**, 1025–1029.
23. Andreadeli, A., Platis, D., Tishkov, V., Popov, V. & Labrou, N. E. (2008). Structure-guided alteration of coenzyme specificity of formate dehydrogenase by saturation mutagenesis to enable efficient utilization of NADP⁺. *FEBS J.* **275**, 3859–3869.
24. Flores, H. & Ellington, A. D. (2005). A modified consensus approach to mutagenesis inverts the cofactor specificity of *Bacillus stearothermophilus* lactate dehydrogenase. *Protein Eng. Des. Sel.* **18**, 369–377.
25. Woodyer, R., van der Donk, W. A. & Zhao, H. (2003). Relaxing the nicotinamide cofactor specificity of phosphite dehydrogenase by rational design. *Biochemistry*, **42**, 11604–11614.
26. Wiegert, T., Sahm, H. & Sprenger, G. A. (1997). The substitution of a single amino acid residue (Ser-116 → Asp) alters NADP-containing glucose-fructose oxidoreductase of *Zymomonas mobilis* into a glucose dehydrogenase with dual coenzyme specificity. *J. Biol. Chem.* **272**, 13126–13133.
27. Carugo, O. & Argos, P. (1997). NADP-dependent enzymes: I. Conserved stereochemistry of cofactor binding. *Proteins*, **28**, 10–28.
28. Carugo, O. & Argos, P. (1997). NADP-dependent enzymes: II. Evolution of the mono- and dinucleotide binding domains. *Proteins*, **28**, 29–40.
29. Ehrensberger, A. H., Elling, R. A. & Wilson, D. K. (2006). Structure-guided engineering of xylitol dehydrogenase cosubstrate specificity. *Structure*, **14**, 567–575.
30. Otwinowski, Z., Borek, D., Majewski, W. & Minor, W. (2003). Multiparametric scaling of diffraction intensities. *Acta Crystallogr. Sect. A*, **59**, 228–234.
31. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
32. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179.
33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. Sect. D*, **66**, 486–501.
34. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.
35. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. Sect. D*, **55**, 247–255.
36. Brunger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
37. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & 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.
38. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619.
39. Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J. *et al.* (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D*, **66**, 12–21.
40. Zhao, H. & Arnold, F. H. (1999). Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng.* **12**, 47–53.