The FEBS Journal

FEBS PRESS science publishing by scientists

# Crystal structure of yeast xylose reductase in complex with a novel NADP-DTT adduct provides insights into substrate recognition and catalysis

Bhaskar Paidimuddala[1], Samar B. Mohapatra[2], Sathyanarayana N. Gummadi[1] and Narayanan Manoj[2]

1 Applied and Industrial Microbiology Laboratory, Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India

2 Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India

**Keywords**

xylose reductase; AKR superfamily; *Debaryomyces nepalensis*; crystal structure complex; NADP-DTT adduct

**Correspondence**

S. N. Gummadi, Applied and Industrial Microbiology Laboratory, Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India

E-mail: gummadi@iitm.ac.in

or

N. Manoj, Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India

E-mail: nmanoj@iitm.ac.in

(Received 11 June 2018, revised 13 September 2018, accepted 26 September 2018)

doi:10.1111/febs.14667

Aldose reductases (ARs) belonging to the aldo-keto reductase (AKR) superfamily catalyze the conversion of carbonyl substrates into their respective alcohols. Here we report the crystal structures of the yeast *Debaryomyces nepalensis* xylose reductase (*Dn*XR, AKR2B10) in the apo form and as a ternary complex with a novel NADP-DTT adduct. Xylose reductase, a key enzyme in the conversion of xylose to xylitol, has several industrial applications. The enzyme displayed the highest catalytic efficiency for L-threose ($138 \pm 7$ mM$^{-1}$·s$^{-1}$) followed by D-erythrose ($30 \pm 3$ mM$^{-1}$·s$^{-1}$). The crystal structure of the complex reveals a covalent linkage between the C4N atom of the nicotinamide ring of the cosubstrate and the S1 sulfur atom of DTT and provides the first structural evidence for a protein mediated NADP–low-molecular-mass thiol adduct. We hypothesize that the formation of the adduct is facilitated by an *in-crystallo* Michael addition of the DTT thiolate to the specific conformation of bound NADPH in the active site of *Dn*XR. The interactions between DTT, a four-carbon sugar alcohol analog, and the enzyme are representative of a near-cognate product ternary complex and provide significant insights into the structural basis of aldose binding and specificity and the catalytic mechanism of ARs.

**Database**

Structural data are available in the PDB under the accession numbers 5ZCI and 5ZCM.

## Introduction

Aldose reductases (ARs) (EC 1.1.1.21) are members of the aldo-keto reductase (AKR) superfamily that contains homologs present across all phyla. AKRs catalyze the conversion of a wide variety of carbonyl substrates to their corresponding alcohols with concomitant oxidation of NAD(P)H [1,2]. The superfamily has been classified into 16 families in the AKR superfamily database [1]. In this scheme, the ARs are present in the AKR 1, 2, 3, and 4 families. Human ARs (hARs) belonging to the AKR1 family are implicated in diabetic complications, fatty liver, chronic kidney diseases, and the oxidative stress-induced activation of NF-kB and AP1 signals [1,3–6]. Fungal ARs belong to the AKR2 and AKR3 families and are inducible enzymes with important roles in growth and metabolism [1,7]. The AKR2 family xylose reductase

**Abbreviations**

AKR, Aldo-keto reductase; AR, aldose reductase; *Ct*XR, *Candida tenuis* xylose reductase; *Dn*XR, *Debaryomyces nepalensis* xylose reductase; hAR, human aldose reductase; XR, xylose reductase.

(XR) (EC 1.1.1.307) is the first enzyme in the xylose utilization pathway and can be exploited for the biosynthetic production of polyols such as xylitol, arabitol, and erythritol for applications in food and pharmaceutical industries [7–9]. The ARs of the AKR3 family are plant erythrose reductases involved in the regulation of plant abiotic stress defense and share similar substrate preferences of the homologous AKR2 members [1,10–14].

The superfamily members share high levels of sequence conservation with the human, plant, and fungal ARs displaying nearly 40% sequence identity. The mostly conserved residues involved in catalysis and cosubstrate binding suggest a common mode of catalytic mechanism [1]. Numerous crystal structures of ARs from the AKR1 family have been determined owing to their therapeutic importance for inhibitor development. These include the well characterized *Homo sapiens* (hARs; PDB: 1ADS, 3V36, 2IKI, 1US0) [2,15–17], *Sus scrofa* (*Ss*AR; PDB: 1DLA, 1AH0) [18,19], *Rattus norvegicus* (*Rn*AR; PDB: 3O3R, 3QKZ) [20,21], and *Giardia lamblia* (*Gl*AR; PDB: 3KRB) homologs [22]. The AKR fold is a conserved $(\beta/\alpha)_8$ barrel with the active site located in a large, deep elliptical pocket at the C-terminal end of the barrel and a bound NADPH in an extended conformation [2,23]. Another structural feature common to all AKRs is the large conformational change of a 'safety belt' element that acts to sequester the cosubstrate from solvent [23]. The enzyme follows an ordered bi-bi catalytic mechanism in which NADPH binds first followed by the carbonyl substrate and the conformational changes, primarily of the 'safety belt' loops, regulate the uptake and release of the cosubstrate [24,25]. The structural basis of substrate recognition and specificity in this superfamily primarily is derived from studies of inhibitor bound complexes of AKR1 members [19,26].

The fungal ARs (AKR2 and AKR3 families) have been poorly characterized with respect to their structure–function properties despite the potential for use in biotechnological applications. To date, the XR (AKR2B5) from *Candida tenuis* [*C. tenuis* XR (*Ct*XR); PDB: 1JEZ, 1K8C] represents the sole instance of a structurally and biochemically well-characterized fungal AR [27]. It displays dual cosubstrate specificity to utilize both NADPH and NADH to convert xylose into xylitol [27–32]. The safety belt loop is disordered in the apo form of *Ct*XR and also in the wild-type apo forms of most reported AR structures [18,27,33,34]. The specific transfer of the 4-pro-*R* hydride to the carbonyl carbon atom of the substrate is known to be highly conserved in this superfamily. However, despite extensive structural studies (69 inhibitor bound AKR1 structures in the PDB), it is intriguing that only two structures of a substrate bound ternary complex of an AKR superfamily member have been reported, namely, hAR bound to D-glyceraldehyde (PDB: 3V36) and to glucose-6-phosphate (PDB: 2ACQ) [15,35].

Here we present the crystal structures of the *Debaryomyces nepalensis* XR (*Dn*XR) in the apo form and as a ternary complex with cosubstrate (NADPH) and a product analog, DTT. *D. nepalensis*, a nonpathogenic *saccharomycetes* yeast can utilize both hexose and pentose sugars to produce polyols [36]. *Dn*XR, a key metabolic enzyme in the D-xylose utilization pathway belongs to the AKR2 family (AKR2B10) and has previously been shown to have strict dependence on NADPH and displays activity for a range of aldehydes [37]. In this study, an ordered conformation of the safety belt loop was observed in the absence of the cosubstrate and provides a bonafide depiction of the apo form of an AR. The structure also showed a novel covalent adduct between the nicotinamide ring and DTT (NADP-DTT), a feature hitherto unobserved in a protein mediated complex. The first structure of an AKR as a ternary complex with a four-carbon sugar alcohol analog provides new insights into substrate binding and specificity and the catalytic mechanism of AKRs. The results of this study may help design and develop engineered yeast ARs to fine tune both substrate and cosubstrate selectivity for biotechnological applications.

## Results

### Enzyme activity, kinetics, and inhibition

Recombinant *Dn*XR was overexpressed in *Escherichia coli* Rosetta and purified to homogeneity by Ni-NTA affinity chromatography as described previously [37]. The protein is composed of 320 amino acid residues and exists as a dimer in solution as indicated by size-exclusion chromatography (SEC) studies. *Dn*XR was earlier shown to exhibit broad substrate specificity, with the highest catalytic efficiency for C5 sugars like arabinose, xylose, and ribose and a strict preference for cosubstrate NADPH [37]. In this study, *Dn*XR was tested for activity and kinetic parameters using two C4 sugars, namely, D-erythrose and its diastereoisomer, L-threose. Interestingly, *Dn*XR showed higher activity and catalytic efficiency (20- to 92-fold) for the four-carbon sugars in comparison to the five-carbon sugars. The enzyme displayed the highest catalytic efficiency for L-threose ($138 \pm 7$ $mM^{-1} \cdot s^{-1}$) followed by D-erythrose ($30 \pm 3$ $mM^{-1} \cdot s^{-1}$)

and is similar to the pattern of activity displayed by *Ct*XR [38]. Furthermore, the effect of the product erythritol and its dithio analog DL-DTT, on *Dn*XR activity were examined. Both alcohols inhibited the catalytic activity in a concentration-dependent manner. In order to identify the type of inhibition mechanism, the kinetics of *Dn*XR were performed using different concentrations of erythritol/DTT using erythrose as a substrate. Enzyme kinetics were performed using erythritol in the range 500–1100 mM and showed an increase in $K_M$ from 6.5 ± 0.4 to 8.6 ± 0.5 mM, while the $V_{max}$ decreased from 176 ± 3 to 124 ± 3 U·mg$^{-1}$ (Table 1). The data suggest that erythritol inhibits *Dn*XR via mixed inhibition mechanism (Fig. 1A,B). Interestingly, DTT displayed a similar behavior in affecting the kinetic parameters (Table 1) (Fig. 1C,D). Although the kinetic studies suggest a mixed inhibition mechanism for both erythritol and DTT, the latter is a far stronger inhibitor with an $K_I$ value of 1.8 mM in comparison to erythritol ($K_I$ = 1577 mM) (Table 1).

## Overall structures of apo *Dn*XR and the ternary complex with cosubstrate and DTT

The crystal structure of apo *Dn*XR was determined at a resolution of 2.0 Å in the space group $P2_12_12_1$ and the structure solved by molecular replacement using one subunit of apo form of *Ct*XR. The tertiary structure contains the characteristic AKR superfamily $(\beta/\alpha)_8$ barrel fold (Fig. 2A) [1,23]. The asymmetric unit contains two subunits related by a non-crystallographic twofold axis with the dimeric arrangement identical to that in *Ct*XR (Fig. 2B). The root-mean-square deviation (rmsd) between the subunits of *Dn*XR and *Ct*XR is 0.9 Å, over 298 $C_\alpha$ atoms. The crystal structure of the complex of *Dn*XR bound to the cosubstrate and D-DTT was determined at a resolution of 1.7 Å in the space group $C222_1$ (Fig. 3). The asymmetric unit in the complex contains one subunit. However, the crystal packing arrangement reveals an identical arrangement of the expected dimer, with the subunits related by a crystallographic twofold axis in this case. The subunit model in the complex contains ordered electron density for an additional nine residues from the N-terminal $His_6$-tag.

The tertiary structure with an eight stranded barrel (β1–β8) is flanked by eight helices (α1–α8) (Fig. 3A). A β-hairpin caps the N-terminal end of the barrel while the C-terminal end contains the conserved active site groove. The extended C-terminal region containing two helices reaches over the active site with the terminal loop that is stabilized by interactions with residues from loop 4. The overall quaternary, tertiary, and secondary structures are largely identical between the two forms except for the cosubstrate-binding regions (see below). The independent subunits in the apo form are virtually identical with a rmsd of 0.46 Å, over 315 $C_\alpha$ atoms, whereas the rmsd between the subunit in the apo form and that of the complex is 2.1 Å. The rmsd between the apo enzyme dimer and the symmetry-related dimer in the complex is 0.75 Å. A summary of the data collection, refinement and validation statistics is given in Table 2.

## Binding of NADP-DTT covalent adduct

A well-conserved feature of AKRs is the presence of the cosubstrate-binding groove that spans the region from the center of the barrel and extends to the solvent exposed edge lined by helix α8 (Fig. 3A,B) [2,27]. Electron density (higher than 3σ) for the cosubstrate and the bound ligand D-DTT can be clearly seen in

**Table 1.** Kinetic parameters of *Dn*XR in the presence of meso-erythritol (product) and DL-DTT (product analog).

| | | Michaelis–Menten plot | | Lineweaver–Burk plot | | | |
|---|---|---|---|---|---|---|---|
| | Conc (mM) | $K_M$ (mM) | $V_{max}$ (U·mg$^-$) | $K_M$ (mM) | $V_{max}$ (U·mg$^{-1}$) | $K_I$ (mM) | Inhibition mechanism |
| Control[a] | – | 5.8 ± 0.37 | 236 ± 4.6 | 5.5 ± 0.31 | 221 ± 2.0 | – | – |
| Meso-Erythritol | 500 | 6.5 ± 0.42 | 176 ± 3.5 | 6.5 ± 0.42 | 174 ± 2.9 | 1577 ± 127 | Mixed |
| | 800 | 7.8 ± 0.73 | 149 ± 4.5 | 7.3 ± 0.45 | 138 ± 2.0 | | |
| | 1100 | 8.6 ± 0.56 | 124 ± 2.7 | 8.2 ± 0.12 | 121 ± 4.0 | | |
| DL-DTT | 0.4 | 6.8 ± 0.52 | 194 ± 4.7 | 6.7 ± 0.63 | 198 ± 4.0 | 1.8 ± 0.03 | Mixed |
| | 0.7 | 7.5 ± 0.82 | 159 ± 5.6 | 7.7 ± 0.18 | 154 ± 3.1 | | |
| | 1.0 | 8.6 ± 0.50 | 127 ± 7.5 | 8.1 ± 0.70 | 119 ± 7.6 | | |

The kinetic parameters of erythrose reduction by *Dn*XR were determined by using Michaelis–Menten and Lineweaver–Burk plots by linear and nonlinear regression fitting ($R^2 > 0.95$). Both product and product analog exhibited a similar mechanism of inhibition. The represented values are mean of three analytical replicates with ± SD. $P < 0.001$. $P$ values are determined by two-way ANOVA method using the GRAPHPAD PRISM software. $K_M$, Michaelis–Menten constant; $V_{max}$, maximal velocity; $K_I$, inhibitory constant.

[a] Kinetic parameters as $K_M$ and $V_{max}$.

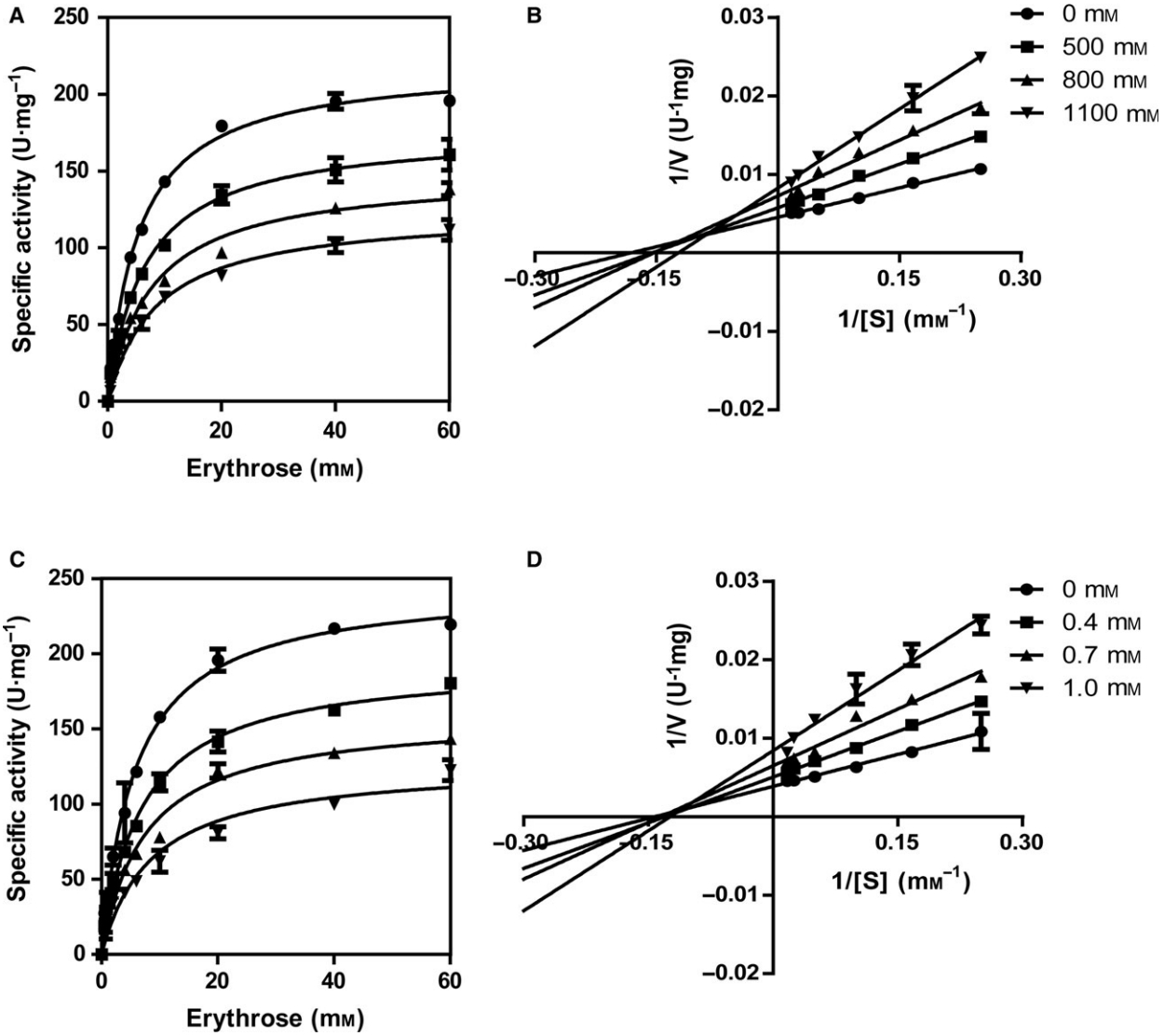



**Fig. 1.** *Dn*XR kinetic parameters for reduction of erythrose in the presence of meso-erythritol (product) and DL-DTT (product analog). The kinetic parameters were measured in the presence of three different concentrations of meso-erythritol (A and B) and DL-DTT (C and D) at standard assay conditions. The obtained data were analyzed by Michaelis–Menten plot (A and C) and Lineweaver–Burk plot (B and D). Both meso-erythritol and DL-DTT showed mixed inhibition of *Dn*XR by affecting $K_M$ and $V_{max}$ of the enzyme. Results are means of triplicates and error bars denote ± SD. The determination coefficients of regression ($R^2$) were found to be > 0.95; $P < 0.001$.

the unbiased omit $F_o - F_c$ map and in simulated annealing omit maps of this region for the ternary complex. Examination of the maps showed an unexpected feature suggestive of a carbon-sulfur covalent bond between the C4N atom of the nicotinamide ring and the sulfur atom (S1) of DTT in the linear form (distance of ~1.8 Å) (Fig. 4A). Hence, a nicotinamide-DTT adduct was inferred and modeled and refinement carried out by introducing a covalent link between the C4N and S1 atoms. There was no evidence for residual difference electron density after the adduct model was refined with full occupancy. In the ternary complex, the average B-factors for the all non-hydrogen protein atoms (3011 atoms) is 15.7 Å$^2$, while that of all atoms in the NADP-DTT adduct is 16.7 Å$^2$, indicating a well ordered conformation of the bound adduct.

The nicotinamide ring is in the boat-like conformation (anti-antiperiplanar) and the attached ribose ring in the C2′-exo conformation. The deviation of the positions of the N1 and C4 atoms of the nicotinamide ring from planarity can be described by the angle ($\alpha_N$) between the planes N1N-C2N-C6N and C2N-C3N-C6N and the angle ($\alpha_C$) between the planes C2N-C3N-C6N and C3N-C4N-C5N, respectively [39]. The values of

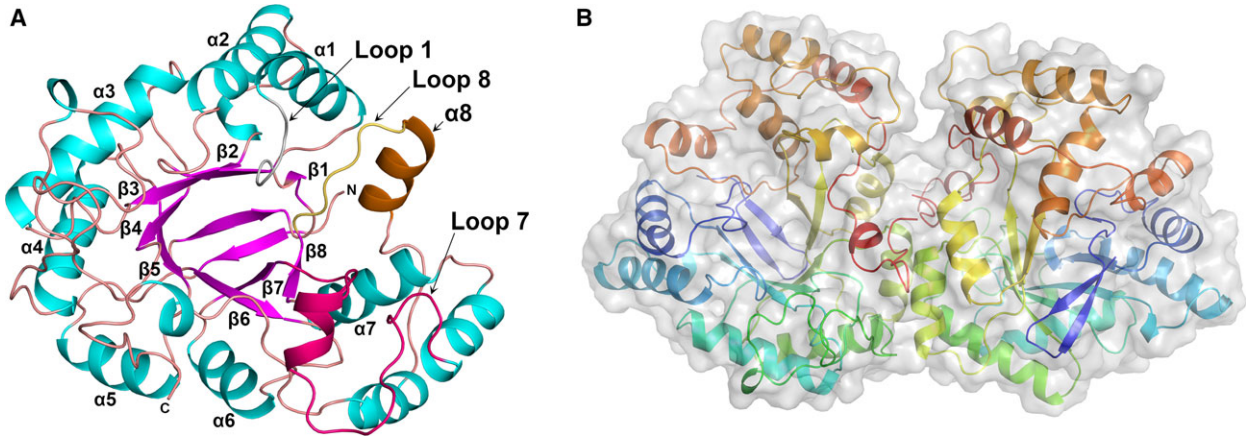



**Fig. 2.** Overall structure of apo *Dn*XR. (A) Ribbon representation of a subunit showing the TIM-barrel fold. Helices are colored cyan, except helix α8 in orange, strands are magenta. Loops are brown, except loop 1 (gray), loop 7 region with associated helices (pink), and loop 8 (yellow). In this figure and in Fig. 3, secondary structure elements corresponding to the standard TIM-barrel nomenclature are labeled (β1–β8) and (α1–α8) and the loops are numbered according to the elements that they connect. (B) Ribbon diagram of the *Dn*XR dimer (apo form) with the corresponding protein surface (gray) in the background. Rainbow coloring from blue to red follows the N- to C-terminal positions of the residues in each subunit.

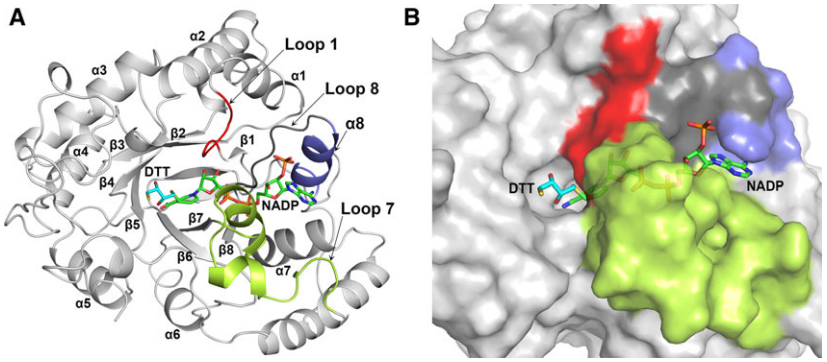



**Fig. 3.** Structure of the *Dn*XR–NADP-DTT adduct complex. (A) Ribbon diagram of a subunit of the *Dn*XR complex shows bound cosubstrate NADP and DTT. Loop 1 (red), loop 8 (black), helix α8 (blue), and loop 7 (lemon) that are involved in cosubstrate binding are highlighted. (B) Surface representation showing bound DTT at the catalytic site and bound NADP at the cosubstrate-binding site. NADP and DTT are displayed as green and cyan stick representation, respectively. Two structural elements, loop 1 (red) and loop 7 region (lemon) that undergo conformational changes upon binding the cosubstrate are highlighted. These two regions together constitute the 'safety belt' interaction to fasten the cosubstrate. Other regions, namely, loop 8 (black) and helix α8 (blue) also undergo concomitant conformational changes. In this figure and Figs 4, 5, 6, 7, 8, and 11, the oxygen, nitrogen, phosphorous, and sulfur atoms are shown in red, blue, orange, and yellow, respectively.

$\alpha_N = 160°$ and $\alpha_C = 156°$ indicate significant departures of the nicotinamide ring from planarity (Fig. 4B–D). The resultant configuration of the C4N atom is $sp^3$ like, resembling the reduced NADPH form. The carboxamide group is almost in plane with respect to the pyridine ring and maintains a dihedral angle (C2N-C3N-C7N-N7N) of 3°. The S1 atom of the adduct occupies the structurally equivalent position of the $H_R$ atom on the A-side (4-pro-*R*) face of the cosubstrate (Fig. 4D).

The cosubstrate adopts an extended anti-conformation with only one intramolecular hydrogen bond present between atoms O3D of ribose and O2N of pyrophosphate. A large number of specific nonbonded interactions are involved in the binding. Among, these, the side chain of Tyr212 stacks with the nicotinamide ring and hydrogen bond interactions between the side chains of Ser164, Asn165, and Gln186 with the carboxamide group orient the A-side of the nicotinamide group toward the substrate binding cavity. The

**Table 2.** Summary of crystal parameters, data collection, and refinement statistics of *Dn*XR

| Data set | Apo form | NADP-DTT complex |
|---|---|---|
| PDB ID | 5ZCI | 5ZCM |
| Data collection | | |
| Wavelength (Å) | 1.54 | 1.54 |
| Space group | $P2_12_12_1$ | $C222_1$ |
| Unit cell parameters *a, b, c* (Å) | 61.8, 69.4, 159.7 | 99.6, 108.2, 71.8 |
| Molecules in asymmetric unit | 2 | 1 |
| Resolution range (Å)[a] | 52.4–2.0 (2.11–2.0) | 36.6–1.7 (1.73–1.70) |
| $R_{merge}$ on $I$ (%)[a] | 15.7 (74.5) | 14.7 (76.9) |
| Mean ($I/\sigma$ ($I$))[a] | 12.6 (3.1) | 13.0 (3.6) |
| Total reflections | 442 170 | 569 904 |
| Unique reflections | 42 060 | 42 718 |
| Completeness (%)[a] | 89.2 (81.1) | 99.4 (98.3) |
| Multiplicity[a] | 10.5 (10.7) | 13.3 (13.1) |
| $CC_{1/2}$[a,b] | 99.7 (77.9) | 99.4 (70.6) |
| Model and refinement statistics | | |
| Free $R$ reflections (%) | 5.0 | 5.0 |
| $R_{work}/R_{free}$ (%)[c] | 18.8/21.8 | 16.2/18.5 |
| Non-hydrogen protein atoms | 5431 | 3011 |
| Solvent atoms | 399 | 308 |
| Ligand atoms | 0 | 56 |
| Mean B value, all atoms (Å$^2$) | 27.2 | 16.9 |
| RMS deviations from ideal geometry | | |
| Bond angle (°) | 0.484 | 1.12 |
| Bond length (Å) | 0.002 | 0.01 |
| Ramachandran plot analysis (%) | | |
| Favored | 96.6 | 97.9 |
| Allowed | 3.2 | 2.1 |
| Outliers | 0.2 | 0.0 |

[a] Value in parentheses refer to the highest resolution shell.
[b] $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data [81].
[c] $R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$, $R_{free}$ is calculated for a set of 5% of reflections randomly chosen and not used for refinement; $R_{work}$ is calculated for the rest of the reflections.

nicotinamide-ribose and the pyrophosphate are mostly buried by the residues of loop 7 and loop 1, whereas the adenosine and the associated ribose 2′-phosphate are located at the distal solvent exposed edge of the groove formed between helices α7 and α8 (Fig. 3).

Extensive attempts to obtain crystals of cognate substrate/product bound ternary complexes with tetrose, pentose, and hexose sugar/sugar alcohols using co-crystallization and soaking techniques proved unsuccessful. Crystals containing the NADP-DTT adduct were obtained in a co-crystallization experiment condition which contained 10 mM racemic DL-DTT, in addition to the cosubstrate. It is noteworthy that the D-isomer of DTT is preferentially bound at the putative substrate-binding pocket. Examination of the crystal packing interactions of the ternary complex showed that the substrate-binding pocket is occluded from bulk solvent by crystal packing interactions from a neighboring symmetry-related molecule. The S4 atom of DTT makes multiple van der Waals interactions (< 4.0 Å) with two residues (Arg-5 and Pro-6) of the N-terminal $His_6$-tag of a neighboring subunit, locking the ligand in place (Fig. 5A). However, in the apo form crystal packing, the substrate-binding pocket is accessible to the bulk solvent in both subunits. Furthermore, differences in the crystal packing interactions have not induced any conformational changes of residues lining the substrate-binding site. Apparently, a Michael addition of the DTT thiolate to the double bond of the α,β-unsaturated nicotinamide is favored and resulted in a derivative under these conditions (Fig. 5B).

Covalent NADP-thiol adducts have previously been reported in the structures of oxido-reductases. However, all these instances are adducts between the pyridine and a Cys residue from the protein (Fig. 5C–E) [40–42]. To the best of our knowledge, this is the first structural evidence of a protein mediated NADP–low-molecular-mass thiol adduct bound to a protein active site. Based on earlier reports of NADP-thiol adducts being identified in solution, we attempted to detect the formation of the adduct in the presence and absence of the enzyme using spectrophotometric measurements under conditions used for the crystallization [41–43]. However, the expected increase in absorbance around 340 nm, indicating the formation of an adduct was not observed in solution (data not shown). In our case, it is probably a result of insufficient thiol oxidation of DTT compared to that of cysteine, at pH 7.5.

## Comparison of apo form and ternary complex of *Dn*XR

In the reported structures of apo forms of wild-type ARs, the loop 7 is disordered and consequently, description of conformational changes upon cosubstrate binding is incomplete in the ARs [18,27,33]. *Dn*XR undergoes significant conformational changes upon formation of the complex with the adduct. The rearrangements and conformational changes results in burying 1011 Å$^2$ out of 1200 Å$^2$ the total accessible surface area of the adduct. Loop 7 (residues 213–237), together with loop 1 constitutes the 'safety belt' that sequesters the cosubstrate from the bulk solvent (Fig. 3). Cosubstrate binding causes the side chain of Trp19 of loop 1 to move out (by ~2.1 Å) of the

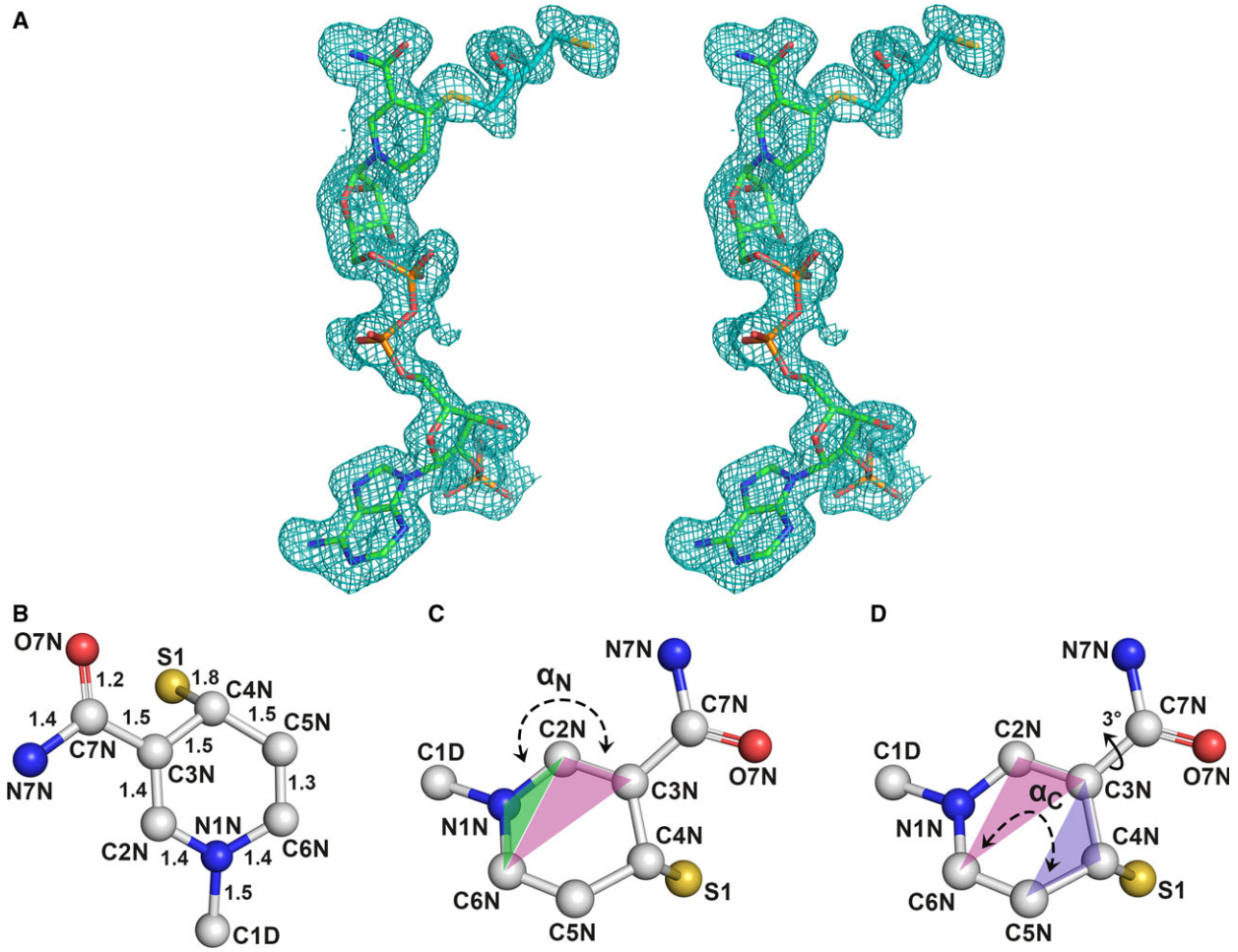



**Fig. 4.** NADP-DTT adduct. (A) Stereo view of the difference electron density ($F_o - F_c$) (marine mesh surface) contoured at 3σ representing bound NADP-DTT adduct. The density corresponds to a simulated annealing omit map (5000 K) calculated from a model lacking the atoms of the adduct. The covalent bond is present between C4N of the pyridine ring and the sulfur (S1) of the DTT with a bond length of 1.8 Å. Carbon atoms in NADP and DTT are represented as green and cyan stick representation, respectively. (B) Structure of the nicotinamide ring of the NADP-DTT adduct. Front view showing bond lengths in Å and atom numbers. (C) Side view showing the angle $\alpha_N$ (160°) is that formed by the planes defined by N1N-C2N-C6N (green) and C2N-C3N-C6N (magenta). (D) Side view showing the angle $\alpha_C$ (156°) is that between the planes defined by C2N-C3N-C6N (magenta) and C3N-C4N-C5N (purple). The dihedral angle (3°) is that defined by C2N-C3N-C7N-N7N. The atom S1 belongs to DTT.

binding pocket to accommodate the nicotinamide ring and the pyrophosphate. The largest change is in the conformation of loop 7 that shows a rmsd value of 3.96 Å for the $C_\alpha$ atoms of the region containing residues 218–231 (Fig. 6A). The Oγ atom of Ser219 in loop 7 shifts as much as 8.4 Å toward the cosubstrate to interact with the O1N atom of pyrophosphate (Fig. 6B). These changes place the nicotinamide ring deep in the center of the barrel close to active site (Fig. 3B).

Another major conformational change associated with cosubstrate binding involves the region containing loop 8 and helix α8. In the complex, the helix undergoes a rotation of ~6°, toward the cosubstrate, clamping it and allowing for the formation of non-bonded interactions between residues 269–279 (rmsd of ~1.0 Å for $C_\alpha$ atoms) and the ribose 2′-phosphate and adenosine groups (Fig. 6A). This is accompanied by several side-chain rearrangements as well, resulting in the key interactions that enable the binding of the ribose 2′-phosphate via two hydrogen bonds with Oγ atom of Ser270 and the main chain amide group from residue Asn271 and two ionic interactions with the side chains of Lys269 and Arg275. The side chain of the highly conserved Arg275 is however disordered in the apo form and appears to be stabilized on binding

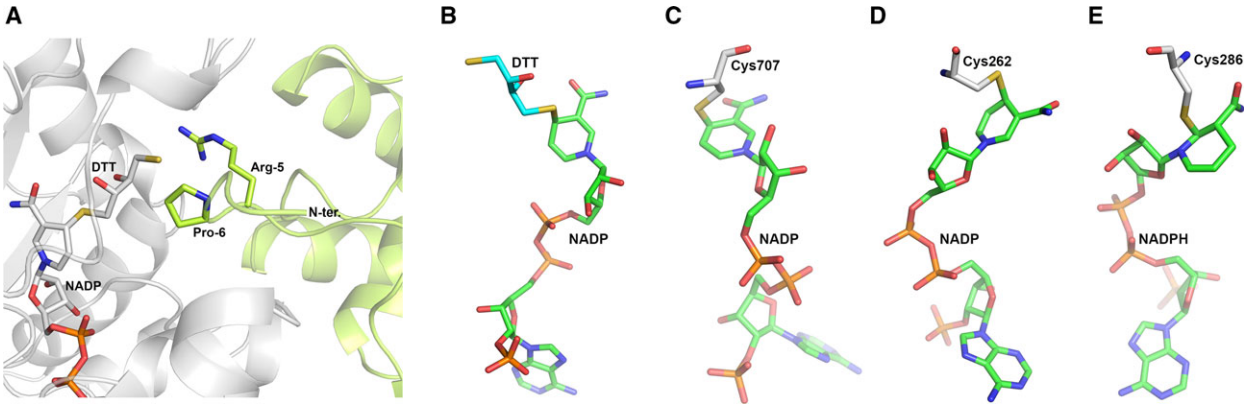



**Fig. 5.** Crystal packing interactions at the substrate-binding pocket of *Dn*XR and comparison of NADP adducts observed in different enzymes. (A) Bound DTT in the subunit (white ribbon) is locked in place via van der Waals interactions (< 4.0 Å) of the S4 atom with two residues, Arg-5 and Pro-6 located in the N-terminal $His_6$-tag region of a symmetry-related subunit (lemon ribbon). The NADP-DTT adduct and the interacting residues of the symmetry-related subunit are displayed as white and lemon-colored sticks, respectively. (B) NADP-DTT in *Dn*XR (this study). (C) NADP-Cys in Rat 10-formyltetrahydrofolate dehydrogenase (PDB: 2O2Q). (D) NADP-Cys in *Synechococcus* Succinic semialdehyde dehydrogenase (PDB: 4ITA). (E) NADPH-Cys in *Pseudomonas aeruginosa* aldehyde dehydrogenase (PDB: 2WOX). NADP(H), Cysteine, and DTT are displayed as green, white, and cyan-colored carbon sticks. The Cys residue in each case is the catalytic residue of the protein.

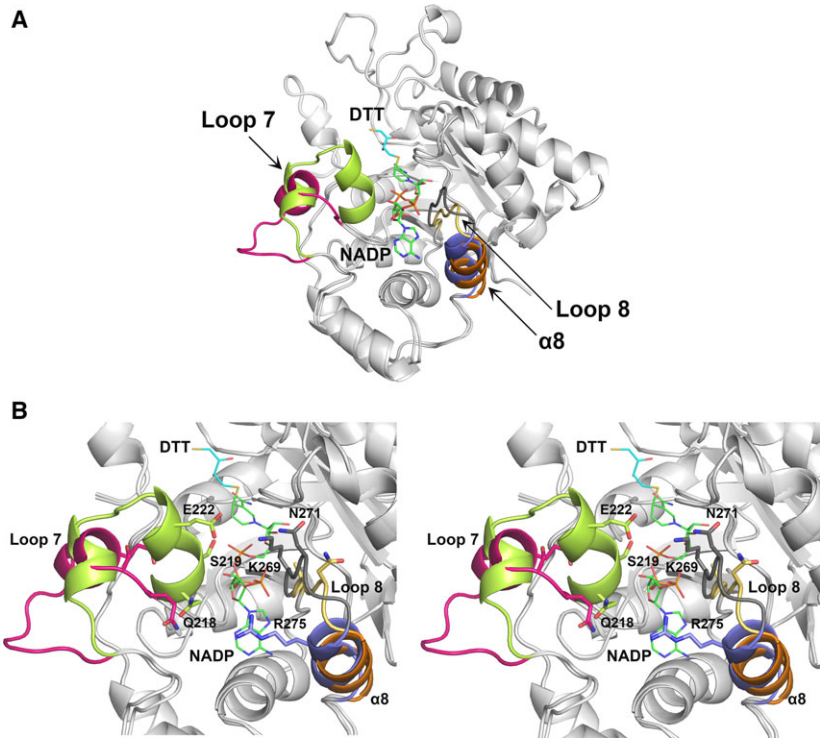



**Fig. 6.** Conformational changes upon cosubstrate binding. (A) Ribbon representation of the superposition of the *Dn*XR apo and complex forms. The structural elements that undergo major conformational changes are highlighted in both forms. (B) Stereo view of the arrangement of specific side chains of residues in the regions that undergo significant conformational changes upon cosubstrate binding. Residues are represented in sticks in corresponding colors of their structural elements. In both illustrations, the structural elements, loop 7 (apo form-pink; complex-lemon), loop 8 (apo form-yellow; complex-black), and helix α8 (apo form-orange; complex-blue) are highlighted. NADP and DTT are represented in green and cyan, respectively.

of the cosubstrate (Fig. 6B). These interactions appear to be primarily responsible for the specificity of the enzyme for NADPH over NADH. Finally, the side chain of Asn279 shifts by ~2.2 Å to establish bidentate hydrogen bond interactions with the N6A and N7A atoms of adenosine.

## Substrate-binding site and catalytic mechanism

The catalytic site architecture of AKRs includes a highly conserved tetrad of residues Asp42, Tyr47, Lys76, and His109 (*Dn*XR numbering) lining the bottom of a deep open cavity. Superposition of *Dn*XR on

the hAR ternary complexes (PDBs: 3V36, 2ACQ) [15,35] reflects the high degree of structural conservation of the tetrad across the families (Fig. 7). The shape of the open substrate-binding pocket in AKRs is defined by loop 2, loop 4, and the C-terminal loop. To the best of our knowledge, the structures of hAR·$NADP^+$ with D-glyceraldehyde and glucose-6-phosphate represent the only examples of substrate bound ternary complexes of an AKR superfamily member [15,35]. Although Zheng *et al.* [15] have not described the enzyme–glyceraldehyde interactions in detail, the structure shows that the acceptor carbonyl oxygen atom is 3.1 Å from the donor C4N atom of $NADP^+$ and 2.9 Å from the OH group of Tyr48, a catalytic residue. However, in the complex with glucose-6-phosphate, the substrate is bound in a 'backward' and unproductive cyclic sugar conformation with the phosphate group occupying the position equivalent to the carbonyl group of an aldehyde substrate [35]. The equivalent distances of the O1P atom are 3.3 and 2.9 Å from the C4N atom and the OH group of Tyr48, respectively (Fig. 7).

In the absence of relevant complexes, numerous studies of AKRs have utilized molecular modeling of substrates into the putative active sites of holo forms to establish the enzyme–substrate interactions and explain key features, including the enzyme mechanism, substrate specificity, and active site plasticity of the AKRs [27,44–49]. In the *Dn*XR complex, DTT is bound in the substrate-binding pocket in an orientation expected for a four-carbon open chain sugar alcohol, reminiscent of a product-bound form (Fig. 7). Apparently, DTT, a dithio analog of erythritol, can occupy the substrate-binding pocket since it is a product inhibitor of the forward reaction of D-erythrose reduction. The configuration of the DTT C1 atom (equivalent of the carbonyl carbon of the substrate) is tetrahedral $sp^3$ like and is present on the C4N-pro-*R* side of the cosubstrate. Superposition of the *Dn*XR complex and the two hAR complexes shows that the ligands occupy the same location in the substrate-binding site (Fig. 7). The positions of the glyceraldehyde carbonyl oxygen atom, the phosphate O1P atom, and the equivalent S1 atom of DTT are within 0.6 Å. The S1 atom of DTT is within hydrogen bonding distance to the conserved catalytic residues, namely, OH group of Tyr47 (3.2 Å) and Nε2 atom of His109 (3.5 Å). In the hAR-glyceraldehyde complex, the corresponding distances from the carbonyl oxygen atom are 2.9 and 3.0 Å, respectively, while the corresponding distances of the O1P atom in the hAR-glucose-6-phosphate complex are 2.9 and 3.4 Å, respectively. Thus, it is

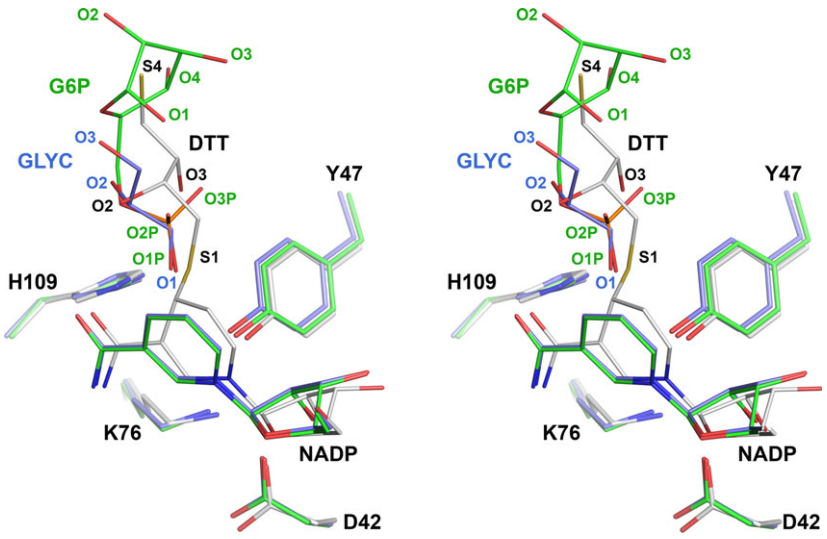



**Fig. 7.** Stereo representation of the superposition of the catalytic tetrads of *Dn*XR complex and hAR complexes with glyceraldehyde and glucose 6-phosphate. The catalytic residues (D42, Y47, K76, and H109) and bound NADP-DTT are labeled according to *Dn*XR numbering and shown in stick representation. White (*Dn*XR; this study), blue (hAR; PDB: 3V36), and green (hAR; PDB: 2ACQ). Non-carbon atoms of the bound glyceraldehyde and glucose 6-phosphate in hAR ternary complexes and DTT in *Dn*XR are labeled. The product analog DTT in *Dn*XR occupies the substrate-binding site in an orientation that permits the hydride transfer as evident by comparison with the position of the cognate substrate glyceraldehyde in an hAR. However, in another the hAR ternary complex, glucose-6-phosphate is bound in a 'backward' and unproductive cyclic sugar conformation with the phosphate group occupying the position equivalent to the carbonyl group of an aldehyde substrate.

reasonable to suppose that the E·NADP-DTT complex obtained here represents the geometry of the enzyme–near-cognate sugar alcohol product complex, allowing for a precise description of the substrate-binding site and the putative interactions in an AKR-four-carbon linear sugar substrate complex for the first time.

The DTT moiety maintains an extended conformation (2R, 3R configuration) with the S4 thiol group exposed to the bulk solvent (Fig. 8A,B). The S1 atom occupies the position equivalent to the carbonyl oxygen of the substrate at the bottom of the pocket. The binding mode of DTT is primarily determined by two specific hydrogen bond interactions anchoring the ligand hydroxyls on either side, namely, between Nδ2 atom of Asn305 and the C2 hydroxyl group (3.0 Å) and between the Oδ2 atom of Asp46 (2.7 Å) and the C3 hydroxyl group. Additionally, DTT makes multiple van der Waals contacts with the side-chain groups of nonpolar residues that line the active site cavity. These involve the C1 atom with Trp19 (3.5 Å), the C4 atom

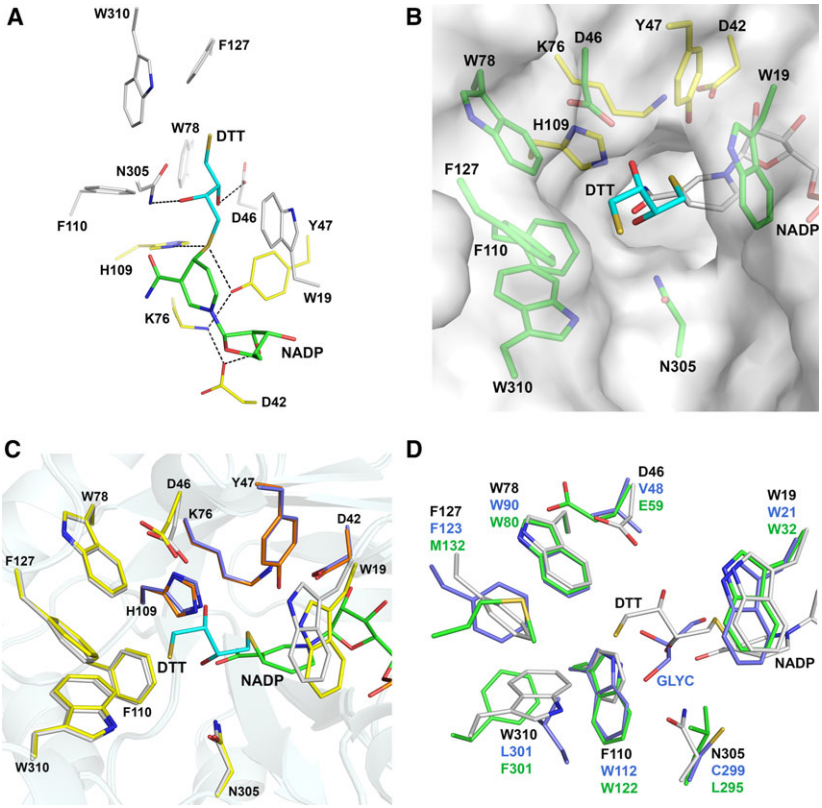



**Fig. 8.** Substrate-binding pocket. (A) Interactions between bound DTT and *Dn*XR. Close-up view of the proposed substrate-binding pocket showing nonbonded interactions (distances in the range 2.3–4.0 Å) of the DTT moiety with residues lining the active site pocket. The catalytic tetrad residues and the substrate-binding residues are shown in yellow and gray stick representation, respectively. Hydrogen bond interactions are displayed as dashed black lines. NADP and DTT are depicted in green and cyan stick representation, respectively. (B) Surface representation of the substrate-binding pocket of the *Dn*XR complex. The outer edge of the pocket is wide and lined by side chains of nonpolar residues (W310, F110, F127, W78, W19). The bottom of the pocket constituting the catalytic site (residues D42, K76, Y47, H109) and accommodating the carbonyl carbon and the oxygen atom, is far narrower. The substrate-binding residues and catalytic tetrad are depicted as green and yellow-colored carbon sticks, respectively. NADP and DTT in the adduct are shown as white and cyan-colored carbon sticks. (C) Close up view of the superposition of the active sites of the DnXR apo and complex forms depicting conformational changes of catalytic and substrate-binding residues of *Dn*XR upon binding of the adduct. Active site residues are shown as blue and orange-colored carbon sticks for the apo and ternary complex forms, respectively. Substrate-binding residues are depicted as yellow- and white-colored carbon sticks for the apo and complex forms, respectively. No significant changes occur except for the change in the side-chain conformations of W19 and D46. (D) Comparison of substrate-binding residues of *Dn*XR (yeast), hAR (human), and *Hv*AR (barley). Close-up view of the substrate-binding residues in stick representation showing superposition of *Dn*XR (white), hAR (blue, PDB: 3V36), and *Hv*AR (green, PDB: 2BGQ). The bound cognate substrate glyceraldehyde in hAR and NADP-DTT adduct in *Dn*XR are shown as blue- and white-colored carbon sticks, respectively. The residues corresponding to positions N305, D46, and F127 in *Dn*XR are found to be divergent across different lineages.

with the Trp78 (3.7 Å), and the S4 thiol group with Phe127 (3.9 Å) and Trp310 (3.8 Å). The substrate-binding site remains mostly unchanged upon ligand binding suggesting the existence of a rigid and preformed cavity for sugar substrates (Fig. 8C). An induced rearrangement occurs in the side chain of Asp46 where the Oδ2 atom moves toward the C3 hydroxyl group by ~0.8 Å to form a hydrogen bond. The slight rearrangement observed in the substrate-binding pocket involving the side chain of Trp19 of loop 1 is induced by cosubstrate binding to the protein (Fig. 8C). This rearrangement results in significant van der Waals contacts between the Trp residue and the C1 atom (equivalent of the carbonyl carbon of the substrate) and has been observed in the cosubstrate bound holo forms of hAR, pig AR, *Ct*XR, and plant AR [2,18,27,33]. Furthermore, conformations of several conserved residues lining this pocket are nearly identical between the yeast, human, and plant AR homologs (Fig. 8D).

The reaction mechanism and the roles of the catalytic tetrad in the AKRs (Asp42, Tyr47, Lys76, His109) have been dissected in detail using crystal structures of numerous apo and holo forms in combination with biochemical and kinetic data from substitution mutants and computational studies involving docking and QM/MM calculations [1,2,23,27,31,44–53]. The observed geometry of nonbonded interactions between the enzyme and the adduct is consistent with the general mechanism for erythrose reduction by *Dn*XR (Fig. 9A). The C4N group of the nicotinamide moiety in the reduced puckered geometry is poised for the stereospecific hydride transfer to the *re* side of the C1 atom of the ligand, the distance being 2.7 Å. The typical donor-acceptor hydride transfer distances in similar enzymes, for example in alcohol dehydrogenases, are in the range of 2.8–3.9 Å [51,52]. The phenolic oxygen of Tyr47 is expected to have a lowered p*K*a as a result of the hydrogen bond to Lys76 amino group (3.1 Å), which in turn maintains a salt bridge with the Oδ2 of Asp42 (2.5 Å). The reprotonation of Tyr47 is expected to be assisted by a proton relay involving His109 and to the bulk solvent, in turn [1,23,50].

## Discussion

Fungal XRs, in general, have been poorly studied with respect to their detailed structure–function properties. In particular, the structural basis of the broad specificity for *in vitro* substrates that include aldoses of varying lengths and a range of non-sugar substituent groups at the non-reacting end of the aldehyde is not clear. Furthermore, the ability of the enzyme to select the physiological substrate xylose in the promiscuous substrate-binding pocket has not been established primarily due to the lack of structural data from an appropriate ternary complex. This report describes the structures of *Dn*XR in the apo form and as a complex of the holo form with DTT, a dithio C4 sugar alcohol analog, and shows how recognition and selectivity for aldoses is achieved in the enzyme.

### Formation of NADP-DTT adduct

Attempts to generate ternary complexes of AKRs with cognate substrates/products appear to have been generally unsuccessful. A likely explanation could be the

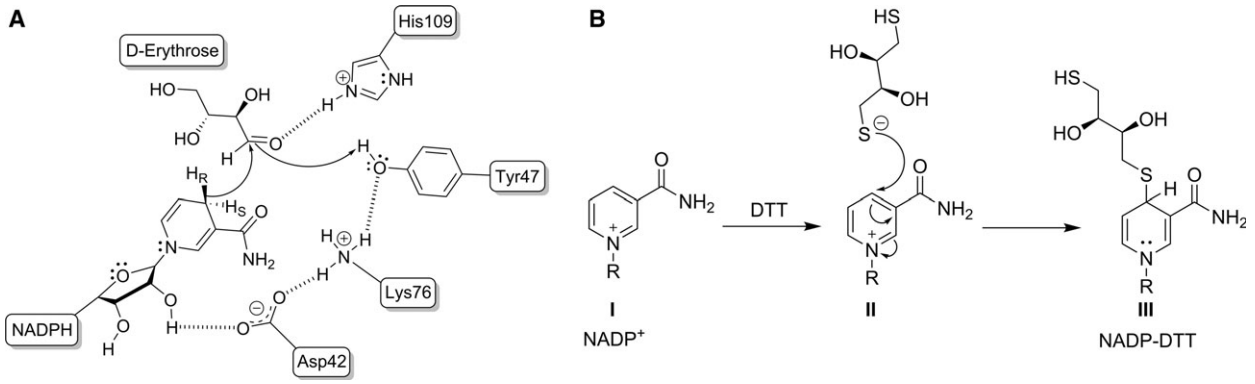



**Fig. 9.** Proposed reaction mechanisms of *Dn*XR catalysis and NADP-DTT adduct formation. (A) Catalytic mechanism of the AR activity by *Dn*XR (see text for details). The arrangement of catalytic residues, their interactions with the cosubstrate and the substrate erythrose, and the orientations of the *re*-/*si*- sides are as observed in the structure of the *Dn*XR complex. Hydrogen bonds interactions are shown as short dashed lines. (B) Mechanism for the *in-crystallo* formation of the NADP-DTT adduct. The oxidized form of NADPH (I) participates in the addition reaction when the thiolate (DTT) performs a nucleophilic attack at the C4N atom of the pyridine ring (II), thus producing a nicotinamide derivative, the NADP-DTT adduct (III).

presence of an electrostatic barrier created by bound $NADP^+$, a substrate inhibitor that may form during the crystal growth phase preventing the entry of neutral substrates into the active site [54]. Alternatively, the open substrate-binding cavity is lined by multiple nonpolar residues and hence the binding energy provided by these secondary hydrophobic interactions under crystallization conditions may not be favorable for the formation of complexes with polar substrates like sugars. The structure presented here is the first structure of a holo AKR in its active conformation with the substrate-binding site occupied by a tetrose sugar alcohol analog. This complex appears to be a result of a serendipitous *in-crystallo* formation of a covalent NADP-DTT adduct. Under alkaline conditions, $NADP^+$ is known to participate in nucleophilic addition reactions with sulfhydryl compounds like sulfides, cysteines and a range of mercaptans, including dithiols to form 1,2- or 1,4-adducts with the nicotinamide moiety [43]. Structural evidence for the NADP-Cys adduct has been reported earlier in *Synechococcus* succinic semialdehyde dehydrogenase (*Sy*SSADH), *Rattus* 10-formyltetrahydrofolate dehydrogenase (*Ra*FDH), and *Pseudomonas aeruginosa* betaine aldehyde dehydrogenase [40–42]. Although these complexes may result from 'charge-transfer' reactions, the ability to purify the adducts suggests an actual addition reaction [40–42,55].

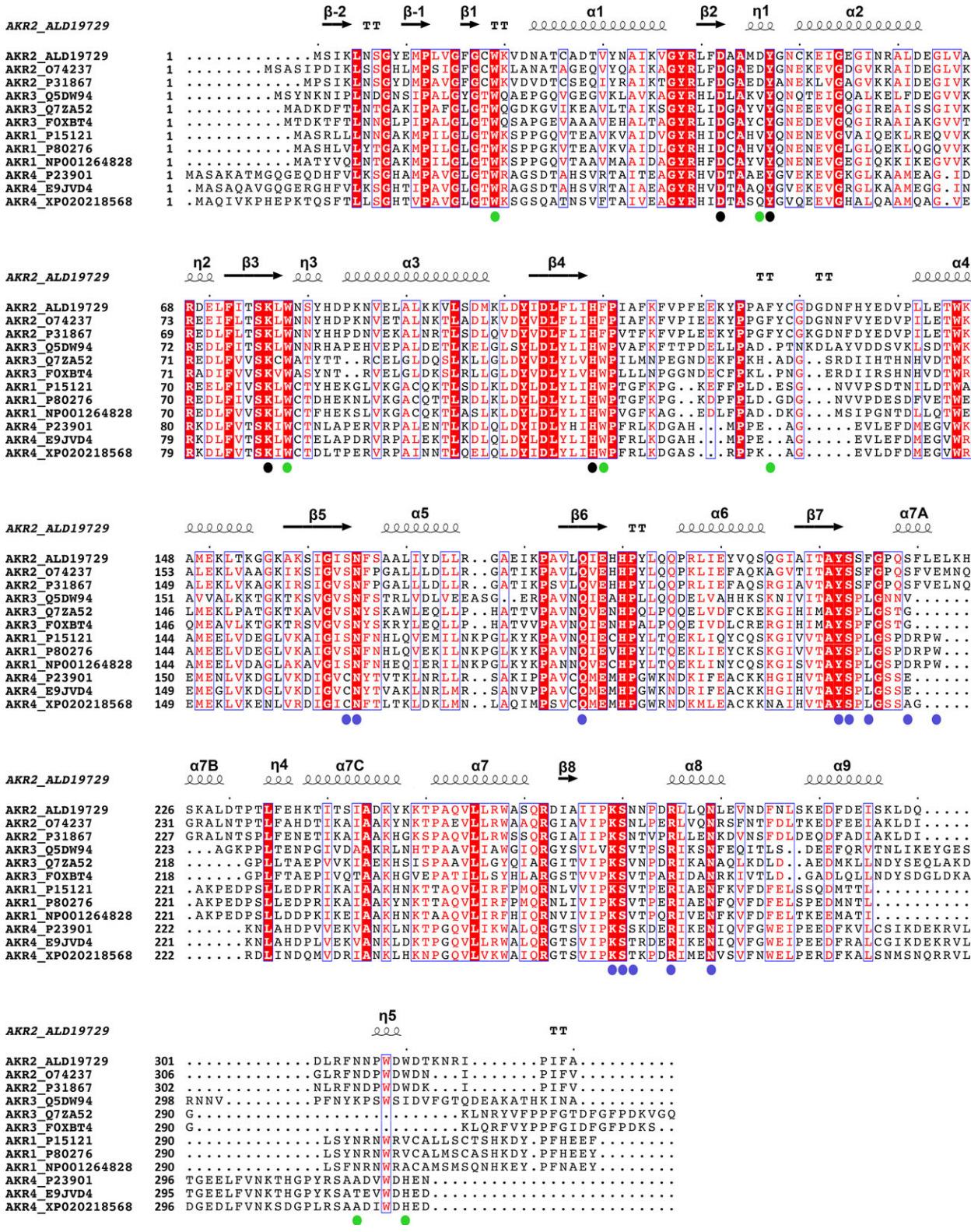
The crystallographic evidence in *Dn*XR for the NADP-DTT covalent linkage provides the first demonstrated evidence for the formation of a protein-mediated NADP–low-molecular-mass thiol adduct. We propose a mechanism where the thiolate group of DTT performs a nucleophilic attack on the electrophilic C4N atom of the oxidized nicotinamide, producing the 1,4-nicotinamide adduct (Fig. 9B). Although NADPH and reduced DTT were initially present in the crystallization cocktail, it is likely that coupled auto-oxidation has occurred under the alkaline pH (HEPES pH 7.5) and aerobic nature of the crystallization conditions, thus favoring the reaction [42,56,57]. We also propose a structural argument where competitive inhibition by DTT, a thiol analog of the reaction product, helps retain an appropriate orientation that favors the interaction of the C4N atom of the nicotinamide ring and the thiolate. Finally, the crystal packing environment appears to be crucial to hamper the cleavage of the adduct by an elimination reaction. While the adduct species may not be an intermediate in the catalytic cycle of AKRs, given the widely prevalent practice of adding DTT as a reducing agent for many biochemical experiments, the side effect of such adduct formation may not always be negligible.

## Cosubstrate binding and specificity in ARs

The mode of cosubstrate binding in the AKR superfamily is highly conserved with mostly identical and structurally conserved residues and water molecules involved in the interactions. Unlike in the apo forms of most other wild-type ARs, loop 7 is present in an ordered form in *Dn*XR. Interestingly, in a $hAR{\cdot}NADP^+$ complex with citrate in the substrate-binding pocket, two discrete conformations corresponding to a fully closed and a 'half-open' state of loop 7 were observed. This transition induced by conformational changes in bound citrate is representative of a situation where reorientation of the product could trigger the release of the cosubstrate via its interactions with loop 7 [58]. Together, changes observed in loop 7 corroborate the two-step kinetic mechanism of carbonyl group reduction in the AKR superfamily where the release of $NADP^+$ from the $E{\cdot}NADP^+$ complex is known to be the rate limiting step [1,23,24]. Furthermore, it is likely that the divergence in the length and residue composition of this loop across the superfamily is an evolutionary mechanism for the modulation of the kinetic properties of its members [1,23] (Fig. 10). For instance, in the plant AKR4 family, specific interactions of a shortened loop 7 that stabilize the closed conformation of the loop were held responsible for the lower cosubstrate affinity and activity of *Zm*AR in comparison to hAR [33,59].

Although the AKRs in general are specific for NADPH, the *Ct*XR represents an example of an AR that can utilize both NADPH and NADH [28,30]. However, *Dn*XR, the closest homolog of

**Fig. 10.** Structure-based sequence alignment of ARs belong to fungal (AKR 2, 3 families), animals (AKR 1 family), and plants (AKR 4 family). The multiple sequence alignment was performed for each family by considering more than 200 nonredundant sequences with identities in range 99–45%. A set of divergent sequences from each family are represented here. The secondary structure elements corresponding to *Dn*XR are indicated on top and conserved residues are highlighted in red. The colored circles at the bottom of alignment indicate the residues involving in catalysis (black), substrate binding (green), and cosubstrate binding (blue) in *Dn*XR. The sequence similarities and secondary structure information from aligned sequences are rendered using the program ESPript 3.0 for representation. The analysis of multiple sequence alignment of the AKR 1–4 families revealed that the catalytic tetrad (Asp42, Lys76, Tyr47, and His109), substrate-binding residues Trp19 and Trp78, and cosubstrate-binding residues Asn165, Gln186, Tyr212, Ser213, Lys269, Ser270, Arg275, and Asn 279 (*Dn*XR numbering) are conserved among all four families.


AKR2_ALD19729
β-2 TT β-1 β1 TT α1 β2 η1 α2
AKR2_ALD19729 1 .....MSIKLNSGYEMPLVGFGCWKVDNATCADTVYNAIKVGYRLFDAAMDYGNCKEIGEGINRALDEGLVA
AKR2_O74237 1 .....MSASIPDIKLSSGHLMPSIGFGCWKLANATAGEQVYQAIKAGYRLFDGAEDYGNEKEVGDGVKRAIDEGLVK
AKR2_P31867 1 .....MPSIKLNSGYDMPAVGFGCWKVDVDTCSEQIYRAIKTGYRLFDGAEDYANEKLVGAGVKKAIDEGIVK
AKR3_Q5DW94 1 .....MSYNKNIPLNDGNSIPALGYGTWQAEPGQVGEGVKLAVKAGYRHLDLAKVYQNQTEIGQALKELFDEGVVK
AKR3_Q7ZA52 1 .....MADKDFTLNTGAKIPAFGLGTWQGDKGVIKEAVLTAIKSGYRLIDGAYVYGNEEEVGQGIREAISSGIVK
AKR3_F0XBT4 1 .....MTDKTFTLNNGLPIPALGLGTWQSAPGEVAAAVEHALTAGYRLIDGAYCYGNEDEVGQGIRAAIAKGVVT
AKR1_P15121 1 .....MASRLLLNNGAKMPILGLGTWKSPPGQVTEAVKVAIDVGYRHIDCAHVYQNENEVGVAIQEKLREQVVK
AKR1_P80276 1 .....MASHLVLYTGAKMPILGLGTWKSPPGKVTEAVKVAIDLGYRHIDCAHVYQNENEVGLGLQEKLQGQVVK
AKR1_NP001264828 1 .....MATYVQLNTGAKMPILGLGTWKSPPGQVTAAVMAAIDAGYRHFDCAYVYQNENEVGEGIQKKIKEGVVK
AKR4_P23901 1 MASAKATMGQGEQDHFVLKSGHAMPAVGLGTWRAGSDTAHSVRTAITEAGYRHVDTAAEYGVEKEVGKGLKAAMEAGLID
AKR4_E9JVD4 1 .MASAQAVGQGERGHFVLKSGHTIPAVGLGTWRAGSDTAHSVRTAIAEAGYRHVDTAAQYGVEKEVGRGLKAAMEGGLIN
AKR4_XP020218568 1 .MAQIVKPHEPKTQSFTLLSGHTVPAVGLGTWKSGSQATNSVFTAIVEAGYRHIDTASQYGVQEEVGHALQAAMQAGLVE
η2 β3 η3 α3 β4 TT TT α4
AKR2_ALD19729 68 RDELFITSKLWNSYHDPKNVELALKKVLSDMKLDYIDLFLIHFPIAFKFVPFEEKYPPAFYCGDGDNFHYEDVPLLETWK
AKR2_O74237 73 REEIFLTSKLWNNYHDPKNVETALNKTLADLKVDYVDLFLIHFPIAFKFVPIEEKYPPGFYCGDGNNFVYEDVPILETWK
AKR2_P31867 69 REDLFLTSKLWNNYHHPDNVEKALNRTLSDLQVDYVDLFLIHFPVTFKFVPLEEKYPPGFYCGKGDNFDYEDVPILETWK
AKR3_Q5DW94 72 REDLFITSKLWNNRHAPEHVEPALDETLKELGLSYLDLYLIHWPVAFKFTTPDELLPADPTNKDLAYVDDSVKLSDTWK
AKR3_Q7ZA52 71 REDLFVVSKCWATYTT..RCELGLDQSLKLLGLDYVDLYLVHWPILMNPEGNDEKFPKH.ADG..SRDIIHTHNHVDTWK
AKR3_F0XBT4 71 RADIFVVSKVWASYNT..RVELGLDKSLRLLGLDYVDLYLVHWPLLLNPGGNDECFPKL.PNG..ERDIIRSHNHVDTWR
AKR1_P15121 70 REELFIVSKLWCTYHEKGLVKGACQKTLSDLKLDYLDLYLIHWPTGFKPG..KEFFPLD.ESG...NVVPSDTNILDTWA
AKR1_P80276 70 REDLFIVSKLWCTDHEKNLVKGACQTTLRDLKLDYLDLYLIHWPTGFKPG..KDPFPLD.GDG...NVVPDESDFVETWE
AKR1_NP001264828 70 REDLFVVSKLWCTFHEKSLVKGACQKTLASLKLDYLDLYLIHWPVGFKAG..EDLFPAD.DKG...MSIPGNTDLLQTWE
AKR4_P23901 80 RKDLFVTSKIWCTNLAPERVRPALENTLKDLQLDYIDLYHIHWPFRLKDGAH..MPPE..AG.....EVLEFDMEGVWK
AKR4_E9JVD4 79 RKDLFVTSKLWCTELAPDRVRPALEKTLKDLQLDYLDLYLIHWPFRLKDGAH..MPPE..AG.....EVLEFDMEGVWR
AKR4_XP020218568 79 RKDLFVTSKIWCTDLTPERVRPAINNTLQELQLDYIDLYLIHWPFRLKDGAS..RPPK..AG.....EVLDFDMEGVWR
β5 α5 β6 TT α6 β7 α7A
AKR2_ALD19729 148 AMEKLTKGGKAKSIGISNFSAALIYDLLR..GAEIKPAVLQIEHHPYLQQPRLIEYVQSQGIAITAYSSFGPQSFLELKH
AKR2_O74237 153 ALEKLVAAGKIKSIGVSNFPGALLLDLLR..GATIKPAVLQVEHHPYLQQPKLIEFAQKAGVTITAYSSFGPQSFVEMNQ
AKR2_P31867 149 ALEKLVKAGKIRSIGVSNFPGALLLDLLR..GATIKPSVLQVEHHPYLQQPRLIEFAQSRGIAVTAYSSFGPQSFVELNQ
AKR3_Q5DW94 151 AVVALKKTGKTKSVGVSNFSTRLVDLVEEASG..ERPAVNQIEAHPLLQQQDELVAHHKSKNIVITAYSPLGNNV.....
AKR3_Q7ZA52 146 LMEKLPATGKTKAVGVSNYSKAWLEQLLP..HATTVPAVNQVENHPQLPQQELVDFCKEKGIHIMAYSPLGSTG.....
AKR3_F0XBT4 146 QMEAVLKTGKTRSVGVSNYSKRYLEQLLP..HATVVPAVNQIENHPALPQQEIVDLCRERGIHIMAYSPFGSTG.....
AKR1_P15121 144 AMEELVDEGLVKAIGISNFNHLQVEMILNKPGLKYKPAVNQIECHPYLTQEKLIQYCQSKGIVVTAYSPLGSPDRPW...
AKR1_P80276 144 AMEELVDEGLVKAIGVSNFNHLQVEKILNKPGLKYKPAVNQIEVHPYLTQEKLIEYCKSKGIVVTAYSPLGSPDRPW...
AKR1_NP001264828 144 AMEELVDAGLAKAVGISNFNHEQIERILNKPGLKYKPANNQVECHPYLTQEKLINYCQSKGISVTAYSPLGSPDRPW...
AKR4_P23901 150 EMENLVKDGLVKDIGVCNYTVTKLNRLLR..SAKIPPAVCQMEMHPGWKNDKIFEACKKHGIHVTAYSPLGSSE.....
AKR4_E9JVD4 149 EMEGLVKDGLVKDIGVCNYTVAKLNRLMR..SANVPPAVCQMEMHPGWKNDRIFEACKKHGIHVTAYSPLGSSE.....
AKR4_XP020218568 149 EMEKLVKENLVRDIGICNFTLTKLDKLMN..LAQIMPSVCQMEMHPGWRNDKMLEACKKNAIHVTAYSPLGSSAG.....
α7B η4 α7C α7 β8 α8 α9
AKR2_ALD19729 226 SKALDTPTLFEHKTITSIADKYKKTPAQVLLRWASQRDIAIIPKSNNPDRLLQNLEVNDFNLSKEDFDEISKLDQ.....
AKR2_O74237 231 GRALNTPTLFAHDTIKAIAAKYNKTPAEVLLRWAAQRGIAVIPKSNLPERLVQNRSFNTFDLTKEDFEEIAKLDI.....
AKR2_P31867 227 GRALNTSPLFENETIKAIAAKHGKSPAQVLLRWSSQRGIAIIPKSNTVPRLLENKDVNSFDLDEQDFADIAKLDI.....
AKR3_Q5DW94 223 ...AGKPPLTENPGIVDAAKRLNHTPAAVLIAWGIQRGYSVLVKSVTPSRIKSNFEQITLS..DEEFQRVTNLIKEYGES
AKR3_Q7ZA52 218 .....GPLLTAEPVVKIAEKHSISPAAVLLGYQIARGITVIPKSVNPDRIKANAQLKDLD..AEDMKLLNDYSEQLAKD
AKR3_F0XBT4 218 .....GPLFTAEPIVQTAAKHGVEPATILLSYHLARGSTVVPKSVTPARIDANRKIVTLD..GADLQLLNDYSDGLDKA
AKR1_P15121 215 .AKPEDPSLLEDPRIKAIAAKHNKTTAQVLIRFPMQRNLVVIPKSVTPERIAENFKVFDFELSSQDMTTL.....
AKR1_P80276 221 .AKPEDPSLLEDPRIKAIAAKYNKTTAQVLIRFPMQRNLIVIPKSVTPERIAENFQVFDFELSPEDMNTL.....
AKR1_NP001264828 221 .AKPEDPSLLDDPKIKEIAAKHNKTAAQVLIRFHIQRNVIVIPKSVTPQRIVENFKVFDFELTKEEMATI.....
AKR4_P23901 222 .....KNLAHDPVVEKVANKLNKTPGQVLIKWALQRGTSVIPKSSKDERIKENIQVFGWEIPEEDFKVLCSIKDEKRVL
AKR4_E9JVD4 221 .....KNLAHDPLVEKVANKLDKTPGQVLLRWALQRGTSVIPKSTRDERIKENIQVFGWEIPEEDFRALCGIKDEKRVL
AKR4_XP020218568 222 .....RDLINDQMVDRIANKLHKNPGQVLVKWAIQRGTSVIPKSTKPDRLMENVSVFNWELPERDFKALSNMSNQRRVL
η5 TT
AKR2_ALD19729 301 .....DLRFNNPWDWDTKNRI.....PIFA.....
AKR2_O74237 306 .....GLRFNDPWDWDN...I.....PIFV.....
AKR2_P31867 302 .....NLRFNDPWDWDK...I.....PIFV.....
AKR3_Q5DW94 298 RNNV.....PFNYKPSWSIDVFGTQDEAKATHKINA.....
AKR3_Q7ZA52 290 G.....KLNRYVFPPFGTDFGFPDKVGQ
AKR3_F0XBT4 290 G.....KLQRFVYPPFGIDFGFPDKS..
AKR1_P15121 290 .....LSYNRNWRVCALLSCTSHKDY.PFHEEF.....
AKR1_P80276 290 .....LSYNRNWRVCALMSCASHKDY.PFHEEY.....
AKR1_NP001264828 290 .....LSFNRNWRACAMSMSQNHKEY.PFNAEY.....
AKR4_P23901 296 TGEELFVNKTHGPYRSAADVWDHEN.....
AKR4_E9JVD4 295 TGEELFVNKTHGPYKSATEVWDHED.....
AKR4_XP020218568 296 DGEDLFVNKSDGPLRSAADIWDHED.....

*Ct*XR, displays strict specificity for NADPH. The structural basis of the dual specificity was previously explained for *Ct*XR based on the structures of $NADP^+$ and $NAD^+$ bound holo complexes [28]. The enzyme was shown to adopt different side-chain conformations of two residues, namely Glu227 and Asn276, (Glu222 and Asn271 in *Dn*XR) that present different hydrogen bonding interaction patterns in the presence and absence of the ribose 2′-phosphate [28]. Comparison of the *Dn*XR-NADP complex and the *Ct*XR-$NADP^+$ complex shows that all residues and their interactions with the entire cosubstrate are completely identical and structurally conserved between the two homologs. The only difference is that of Asn272 (Leu in *Ct*XR) lining the outer edge of the 2′-phosphate-binding cavity that does not make direct interactions with the phosphate (distances 4.2–4.9 Å) (Fig. 11). Together, these observations strongly suggest that the side chain interactions that were earlier proposed to determine the dual specificity in the *Ct*XR may not be the sole determinants of cosubstrate specificity within the AKR2 family. Reengineering of cosubstrate specificity from NADPH to NADH is of particular importance for biocatalytic applications due to the higher stability and lower cost of NADH [30,60,61]. Our analysis here of the two almost identical cosubstrate-binding sites reveals the limitations of generalizing the structural basis of cosubstrate specificity based solely on comparisons of crystal structures. Considering the pleotropic nature of residue changes, a systematic mutagenesis strategy at multiple locations scanning different substitutions is imperative here. For instance, in a similar effort in the 2,5-DKG AKR, only one out of 40 unique substitution mutants yielded a significant improvement in NADH utilization over NADPH [62].

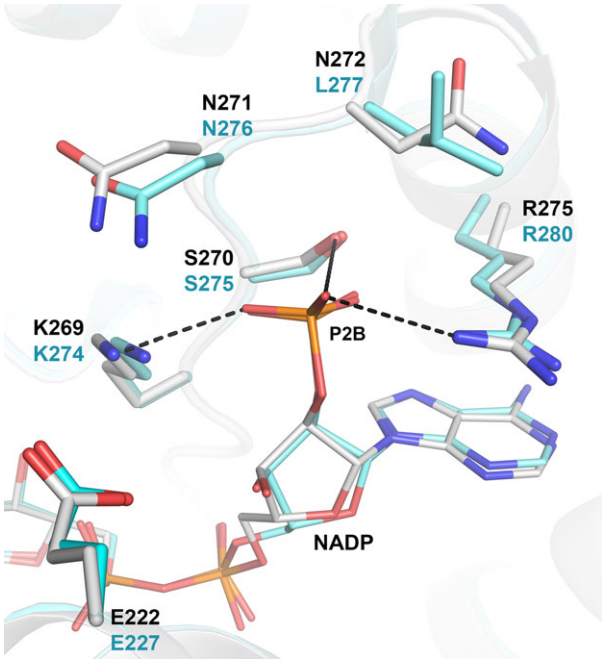



**Fig. 11.** Comparison of residues located in the neighborhood of the ribose 2′-phosphate moieties between *Dn*XR (strict NADPH specificity) and *Ct*XR (dual specificity, NADPH/NADH). The bound cosubstrate and interacting residues in *Dn*XR and *Ct*XR (PDB: 1K8C) are shown as white- and cyan-colored carbon sticks, respectively. All the residues directly interacting with ribose 2′-phosphate of cosubstrate are completely conserved in both XRs despite the differences in their cosubstrate specificities. The only residue variance between these XRs is observed at position L277 in *Ct*XR, which is substituted by N272 in *Dn*XR and are positioned at distances larger than 4.2 Å from the phosphate groups.

## Structural basis of substrate recognition

The ARs accept open chain form of aldoses of varying lengths and different configurations for the carbon atoms at the non-reacting end of the sugars [1,23,27,31]. The outer edge of the AR active site pocket is solvent exposed, wide and lined by side chains of nonpolar residues, whereas the bottom constituting the catalytic site accommodating the carbonyl carbon and the oxygen atoms is far narrower (Fig. 8). It appears that access and activity for a wide range of aldose substrates is facilitated by the open active site.

In the structure presented here, assuming a structural mimic of a bound prochiral substrate, the significant van der Waals interactions that C1 atom makes with the side chain of the Trp19 (six interactions < 4.0 Å) suggests its important steric role in positioning the carbonyl carbon. This residue is completely conserved across the AKR superfamily (Fig. 10). As expected, the catalytic efficiency of a *Ct*XR mutant where the equivalent Trp is substituted by Phe/Tyr shows a decrease of 95% of the value for the wild-type. This is primarily due to a large increase in the $K_M$ (16-fold), while the $k_{cat}$ decreased by 1.5-fold [32]. The interaction between Trp and the substrate is a result of a conformational change occurring in the holo form, subsequent to binding of the cosubstrate. This observation supports the notion that it is indeed the cosubstrate binding that induces the formation of the competent substrate-binding pocket in the AKRs via the Trp residue.

The orientation of DTT places the C2(R) hydroxyl group within hydrogen bonding distance with the Nδ2 of Asn305, strongly indicating that this interaction is the primary determinant of the stereospecificity of the

carbonyl reduction (Fig. 8). Asn305 is completely conserved across diverse members of AKR family 2. The crucial role of Asn305 is also supported by data from kinetic studies of *Ct*XR (Asn309), which revealed that a substitution with Ala caused serious disruption of activity and efficiency [32]. For instance, a 29-fold reduction in catalytic efficiency is observed in the mutant for the reduction reaction of D-galactose. Furthermore, the 30-fold preference in catalytic efficiency of *Ct*XR for D-galactose over 2-deoxy-D-galactose was completely lost in the mutant [32]. Thus, within the AKR family 2, the hydrogen bond between the Asn residue and C2 hydroxyl group also appears to be a determinant of the regioselectivity of the C2 hydroxyl group of aldose substrates, irrespective of the number or the configuration of the carbon atoms at the non-reacting part.

The C-3(R) hydroxyl group of DTT is hydrogen bonded to Oδ2 of Asp46 of *Dn*XR suggesting that this interaction is another key element for binding of polyhydroxylated substrates (Fig. 8). Asp46 is the sole ionizable group in the *Dn*XR directly interacting with the non-reacting part of the substrate and is completely conserved within AKR family 2. Mutation of the equivalent Asp in *Ct*XR (Asp50) lowered the catalytic efficiency for the reduction of D-xylose by ~1.2-fold [32]. Furthermore, the mutant also showed no change in substrate specificity profile compared with the wild-type when tested with other aldoses like D-glucose and D-galactose suggesting that this residue may not be a strong structural determinant for the regioselectivity of hydroxyl groups at carbon positions C3 and further. Taken together, it is evident that the specific sequence features of residues at these two positions may play crucial roles in determining the substrate specificity and selectivity profiles of the AKR family 2.

Among the aldose substrates tested in this study, *Dn*XR has the highest catalytic efficiency for four-carbon sugars compared to five- or six-carbon sugars. Examination of residues that interact with the non-reacting end of the substrate indicate the presence of hydrophobic residues (Trp19, Trp78, Phe110, Phe127, and Trp310 in *Dn*XR) that primarily contribute through van der Waals and hydrophobic interactions [63]. Given the funnel-shaped geometry of the active site and the nonspecific nature of the interactions with the hydroxyl at carbon position C4, it is reasonable to suppose that sugars longer than C4 may possess higher flexibility at the non-reacting end, potentially resulting in fewer/less favorable interactions and thereby affecting the optimal orientation of the substrate in the catalytic site (Fig. 8B).

Interestingly, comparison of the available structures of ARs from AKR families 1, 2, and 4 shows that the main chain positions of residues at the outer edge of the pocket are structurally conserved, whereas some side chains form divergent patterns (Fig. 8D). Trp19 and Trp78 are conserved in all ARs, whereas Phe110 is replaced by Trp in hAR and *Hv*AR. Phe127 in *Dn*XR is conserved in hAR and is substituted by Met132 in *Hv*AR. Trp310 in *Dn*XR is substituted by Leu301 and Phe301 in hAR and *Hv*AR, respectively [15,33] (Fig. 10). We propose that the precise surface geometry presented by variations localized to this region of the substrate-binding pocket provides a structural basis for the extended substrate specificity in the ARs of the superfamily and a similar rationale may be extended for the divergence across the entire superfamily as well. It is likely that protein engineering strategies employing rational or semi-rational redesign of residues identified in our study can be an appropriate starting point to generate variants with desirable biocatalytic properties like enhanced activity, improved enantioselectivity, thermal stability, among others. Previous studies on AKR superfamily members report successful efforts at engineering such variants [64–66]. For instance, in *Ct*XR, a redesign of loop 1 that contains residue Trp19 (in *Dn*XR), yielded a variant with enhanced turnover number and enantioselectivity for bulky ketones of industrial relevance [67].

The structural basis of the broad substrate specificity of the AKR superfamily has been hampered by the lack of the structure of an appropriate complex. The crystal structure of a yeast XR resembling the near-cognate ternary complex presented here provides a structural rationale for the catalytic mechanism, substrate binding, and specificity in the AKR family 2, in particular and in the superfamily, in general.

## Experimental procedures

### Protein expression, purification, and oligomeric state

*Dn*XR (GenBank: KT239024) in the vector pET28a(+) was overexpressed and purified as described previously [37]. The plasmid construct contains a 21-residue-long N-terminal expression and purification tag including a $His_6$ region. A brief description of *Dn*XR purification as follows: the protein was expressed in *E. coli* Rosetta under the IPTG induction and the harvested cell pellet was lysed by sonication. The obtained supernatant containing overexpressed protein was taken for purification by Ni-NTA affinity chromatography. The 100 and 200 mM imidazole elutions of purification were subjected to buffer exchange with 20 mM

Tris/HCl (pH 7.5) and the protein was concentrated using 30 kDa cutoff centrifugal filters. The purity of the obtained protein sample was evaluated on SDS/PAGE (data not shown). The protein concentration was measured by the bicinchoninic acid method and the purified protein was stored at −80 °C until use [68]. The oligomeric state of the purified protein was determined by performing size-exclusion chromatography using BIO-RAD FPLC system at 4 °C. Two milliliter of 20 mM Tris/HCl (pH 7.5) buffer containing 6 mg protein was loaded onto the Superdex 200 GE Healthcare column which was pre-equilibrated with 20 mM Tris/HCl (pH 7.5). The elution was carried out with same buffer at the flow rate of 0.5 $mL \cdot min^{-1}$ while recording the chromatograms by measuring the absorbance at 280 nm.

## Enzyme assays and kinetics

The enzymatic activity of purified protein was measured at optimal pH 7.0 and temperature 45 °C as determined previously [37]. The change in $A_{340}$ was monitored continuously at least for 1 min at every 1 s time point in a 500 μL reaction mixture containing 50 mM sodium phosphate buffer (pH 7.0), 0.3 mM NADPH (cosubstrate), and 6 mM erythrose (substrate) using Lamda 25 UV-VIS spectrophotometer (Perkin Elmer, Wellesley, MA, USA). The linear curve (absorbance vs. time) at least for 1 min was obtained by the addition of a specific amount of enzyme (0.3 μg) into the reaction mixture. One enzyme unit (U) was defined as the amount of enzyme that caused the oxidation of 1 μmol of NADPH per minute [37]. The heat inactivated enzyme and sodium phosphate buffer (pH 7.0) were considered as a references [37]. The C4 sugars D-erythrose and L-threose were assayed for enzyme activity. The kinetic parameters were determined at constant NADPH (0.3 mM as determined previously [37] and different D-erythrose/L-threose concentrations (0.1–60 mM). The obtained data were used to calculate the kinetic constants by fitting for the Michaelis–Menten equation using the software GRAPHPAD PRISM 5 (GraphPad Software, La Jolla, CA, USA).

The effect of product erythritol and its analog DTT on enzyme activity was studied at various concentrations of erythritol (0–2000 mM) and DTT (0–10 mM) at the same assay conditions as mentioned above and the obtained data were used to calculate half maximal inhibitory concentrations ($IC_{50}$) and minimal inhibitory concentrations (MIC) using linear regression function of GRAPHPAD PRISM 5 (data not shown). Using the obtained information on $IC_{50}$ and MIC values, the inhibition kinetics of *Dn*XR was performed at a particular concentration of erythritol/DTT by measuring the initial velocities of enzyme at different erythrose concentrations (0.1–60 mM). The obtained data were used to calculate the kinetic parameters by fitting to the standard noncompetitive/mixed inhibition model for Michaelis–Menten and Lineweaver–Burk plots [69]. The Michaelis–Menten equation for noncompetitive/mixed inhibition model used was

$$v = \frac{V_{max}\,[S]}{\left(1 + \frac{[I]}{K_I}\right)K_M + \left(1 + \frac{[I]}{K_I}\right)[S]}$$

The plot of $K_M$ vs. $[I]$ yielded a linear trend upon linear regression analysis and the values with intercept $K_M$ and slope $K_M/K_I$ and was used for calculation of the inhibitory constant $K_I$.

## Protein crystallization

The protein *Dn*XR was concentrated to 15 $mg \cdot mL^{-1}$ in buffer (20 mM Tris/HCl, pH 7.5). Crystallization experiments were performed in 24-well plates by hanging drop vapor diffusion method at 20 °C using commercial screens (Hampton Research) and each drop was prepared by mixing 1 μL each of protein solution and reservoir solution and equilibrated over 1 mL of reservoir solution. Crystals of the apo form of *Dn*XR (rod shaped) appeared in 2 days in the crystallization condition containing 24% polyethylene glycol 4000, 0.1 M sodium citrate (pH 6.2), and 0.15 M ammonium acetate. Crystals of the NADP-DTT complex were obtained using a protein solution containing 15 $mg \cdot mL^{-1}$ protein, 10 mM NADPH and 10 mM DTT which was preincubated at 4 °C for 4 h. The plate shaped crystals were formed after 3 days with the reservoir solution containing 22% polyethylene glycol 3350, 0.1 M HEPES (pH 7.5), and 0.2 M ammonium sulfate.

## Data collection, structure determination, and refinement

Diffraction data for *Dn*XR crystals were collected at an in-house X-ray source with a MAR345 image plate detector mounted on a Bruker Microstar rotating anode X-ray generator (Cu Kα = 1.5418 Å). Crystals were mounted on cryoloops using the higher concentrations of the reservoir solutions as cryoprotectants and flash-frozen in a nitrogen gas stream at 100 K. All X-ray data were processed and scaled using MOSFLM and SCALA programs as implemented in the CCP4 software package [70–72]. The apo form data were processed to 2.0 Å resolution and the data were indexed in the $P2_12_12_1$ space group. The structure was solved by molecular replacement with the PHASER program using one subunit of apo form of *C. tenuis* XR (PDB: 1JEZ, 71% sequence identity) as the search model [27,73]. The final model of a dimer in the asymmetric unit contains all 320 amino acids encoded in the protein, while the N-terminal $(His)_6$-tag is disordered. The complex data was processed to 1.7 Å resolution and indexed in the space group $C222_1$ and structure was solved using the refined *Dn*XR apo form as a model. The software package PHENIX was used to carry out iterative rounds of restrained refinement and the

program COOT was used for model building [74–76]. The refinement included simulated annealing protocols to remove model bias. R-free calculations were performed to cross validate the refinement with a test set containing 5% of the reflections. The structure and restraints for the adduct were generated using the electronic Ligand Builder and Optimization Workbench (eLBOW) program, as implemented in the PHENIX suite [77]. The coordinates of NADPH and D-DTT available in the PDB Chemical Components dictionary were used to generate the structure and determine the ligand geometry. The resulting model was further optimized using the AM1 semiempirical quantum-chemical method and corresponding restraints were generated as a CIF format file. The output coordinates and the restraints were examined manually using the Restraints Editor Especially Ligand (REEL) tool for validity. The model was then manually fitted into the electron density map and followed by refinement carried out in PHENIX incorporating the generated ligand restraints. The MolProbity program was used for the evaluation of stereochemical quality of the final models [78]. PyMoL was used for the graphical representation of structures [79]. The structure based sequence alignment was carried out using the online program ESPRIPT 3.0 [80].

### Protein Data Bank accession numbers

The atomic coordinates and the corresponding structure factors of the apo form and the ternary complex of *Dn*XR have been deposited in the Protein Data Bank under the accession numbers 5ZCI and 5ZCM, respectively.

## Acknowledgements

Infrastructure support from the Macromolecular X-ray Diffraction Facility, IIT Madras and the DST-FIST Facility, IIT Madras, are acknowledged. BP acknowledges Indian Council of Medical Research for his fellowship.

## Conflict of interest

The authors have no conflict of interest to declare.

## Funding information

This work was supported by a research grant from the Department of Biotechnology, Ministry of Science and Technology, Government of India (Grant BT/PR6157/PID/6/701/2012).

## Author contributions

NM and SNG conceived and coordinated the study and wrote the paper. BP carried out the majority of the experiments, wrote the initial draft of the manuscript, and prepared the figures. SBM and BP carried out data analysis and data visualization. NM carried out the crystallographic analysis and supervised the structural data analysis.

## References

1 Penning TM (2015) The aldo-keto reductases (AKRs): overview. *Chem Biol Interact* **234**, 236–246.
2 Wilson DK, Bohren KM, Gabbay KH & Quiocho FA (1992) An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holo enzyme implicated in diabetic complications. *Science* **257**, 81–84.
3 Kador PF, Wyman M & Oates PJ (2016) Aldose reductase, ocular diabetic complications and the development of topical Kinostat (®). *Prog Retin Eye Res* **54**, 1–29.
4 Lanaspa MA, Ishimoto T, Li N, Cicerchi C, Orlicky DJ, Ruzicky P, Rivard C, Inaba S, Roncal-Jimenez CA, Bales ES *et al.* (2013) Endogenous fructose production and metabolism in the liver contributes to the development of metabolic syndrome. *Nat Commun* **4**, 2434.
5 Andres-Hernando A, Li N, Cicerchi C, Inaba S, Chen W, Roncal-Jimenez C, Le MT, Wempe MF, Milagres T, Ishimoto T *et al.* (2017) Protective role of fructokinase blockade in the pathogenesis of acute kidney injury in mice. *Nat Commun* **8**, 14181.
6 Ramana KV (2011) Aldose reductase: new insights for an old enzyme. *Biomol Concepts* **2**, 103–114.
7 Lee H (1998) The structure and function of yeast xylose (aldose) reductases. *Yeast* **14**, 977–984.
8 Su B, Zhang Z, Wu M, Lin J & Yang L (2016) Construction of plasmid-free *Escherichia coli* for the production of arabitol-free xylitol from corncob hemicellulosic hydrolysate. *Sci Rep* **6**, 26567.
9 Moon HJ, Jeya M, Kim IW & Lee JK (2010) Biotechnological production of erythritol and its applications. *Appl Microbiol Biotechnol* **86**, 1017–1025.
10 Lee JK, Kim SY, Ryu YW, Seo JH & Kim JH (2003) Purification and characterization of a novel erythrose reductase from *Candida magnoliae*. *Appl Environ Microbiol* **69**, 3710–3718.
11 Jovanović B, Mach RL & Mach-Aigner AR (2013) Characterization of erythrose reductases from filamentous fungi. *AMB Express* **3**, 43.
12 Janek T, Dobrowolski A, Biegalska A & Mirończuk AM (2017) Characterization of erythrose reductase from *Yarrowia lipolytica* and its influence on erythritol synthesis. *Microb Cell Fact* **16**, 118.
13 Simpson PJ, Tantitadapitak C, Reed AM, Mather OC, Bunce CM, White SA & Ride JP (2009)

Characterization of two novel aldo-keto reductases from *Arabidopsis*: expression patterns, broad substrate specificity, and an open active-site structure suggest a role in toxicant metabolism following stress. *J Mol Biol* **392**, 465–480.

14 Bartels D, Engelhardt K, Roncarati R, Schneider K, Rotter M & Salamini F (1991) An ABA and GA modulated gene expressed in the barley embryo encodes an aldose reductase related protein. *EMBO J* **10**, 1037–1043.

15 Zheng X, Zhang L, Chen W, Chen Y, Xie W & Hu X (2012) Partial inhibition of aldose reductase by nitazoxanide and its molecular basis. *ChemMedChem* **7**, 1921–1923.

16 Steuber H, Heine A & Klebe G (2007) Structural and thermodynamic study on aldose reductase: nitro-substituted inhibitors with strong enthalpic binding contribution. *J Mol Biol* **368**, 618–638.

17 Howard EI, Sanishivili R, Cachau RE, Mitschler A, Chevrier B, Barth P, Lamour V, Van Zandt M, Sibley E, Bon C *et al.* (2004) Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 A. *Proteins* **55**, 792–804.

18 Rondeau JM, Tête-Favier F, Podjarny A, Reymann JM, Barth P, Biellmann JF & Moras D (1992) Novel NADPH-binding domain revealed by the crystal structure of aldose reductase. *Nature* **355**, 469–472.

19 Urzhumtsev A, Tête-Favier F, Mitschler A, Barbanton J, Barth P, Urzhumtseva L, Biellmann JF, Podjarny A & Moras D (1997) A "specificity" pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil. *Structure* **5**, 601–612.

20 Sundaram K, Dhagat U, Endo S, Chung R, Matsunaga T, Hara A & El-Kabbani O (2011) Structure of rat aldose reductase-like protein AKR1B14 holoenzyme: probing the role of His269 in coenzyme binding by site-directed mutagenesis. *Bioorganic Med Chem Lett* **21**, 801–804.

21 Sundaram K, Endo S, Matsunaga T, Tanaka N, Hara A & El-Kabbani O (2012) Structure of the His269Arg mutant of the rat aldose reductase-like protein AKR1B14 complexed with NADPH. *Acta Crystallogr F* **68**, 400–403.

22 Ferrell M, Abendroth J, Zhang Y, Sankaran B, Edwards TE, Staker BL, Van Voorhis WC, Stewart LJ & Myler PJ (2011) Structure of aldose reductase from *Giardia lamblia*. *Acta Crystallogr F* **67**, 1113–1117.

23 Sanli G, Dudley JI & Blaber M (2003) Structural biology of the aldo-keto reductase family of enzymes: catalysis and cofactor binding. *Cell Biochem Biophys* **38**, 79–101.

24 Kubiseski TJ, Hyndman DJ, Morjana NA & Flynn TG (1992) Studies on pig muscle aldose reductase: kinetic mechanism and evidence for a slow conformational change upon coenzyme binding. *J Biol Chem* **267**, 6510–6517.

25 Kubiseski TJ & Flynn TG (1995) Studies on human aldose reductase: probing the role of arginine 268 by site-directed mutagenesis. *J Biol Chem* **270**, 16911–16917.

26 Rechlin C, Scheer F, Terwesten F, Wulsdorf T, Pol E, Fridh V, Toth P, Diederich WE, Heine A & Klebe G (2017) Price for opening the transient specificity pocket in human aldose reductase upon ligand binding: structural, thermodynamic, kinetic, and computational analysis. *ACS Chem Biol* **12**, 1397–1415.

27 Kavanagh KL, Klimacek M, Nidetzky B & Wilson DK (2002) The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from *Candida tenuis*. *Biochemistry* **41**, 8785–8795.

28 Kavanagh KL, Klimacek M, Nidetzky B & Wilson DK (2003) Structure of xylose reductase bound to $NAD^+$ and the basis for single and dual co-substrate specificity in family 2 aldo-keto reductases. *Biochem J* **373**, 319–326.

29 Leitgeb S, Petschacher B, Wilson DK & Nidetzky B (2005) Fine tuning of coenzyme specificity in family 2 aldo-keto reductases revealed by crystal structures of the Lys274Arg mutant of *Candida tenuis* xylose reductase (AKR2B5) bound to $NAD^+$ and $NADP^+$. *FEBS Lett* **579**, 763–767.

30 Petschacher B, Leitgeb S, Kavanagh KL, Wilson DK & Nidetzky B (2005) The coenzyme specificity of *Candida tenuis* xylose reductase (AKR2B5) explored by site-directed mutagenesis and X-ray crystallography. *Biochem J* **385**, 75–83.

31 Kratzer R, Wilson D & Nidetzky B (2006) Catalytic mechanism and substrate selectivity of aldo-keto reductases: insights from structure-function studies of *Candida tenuis* xylose reductase. *IUBMB Life* **58**, 499–507.

32 Kratzer R, Leitgeb S, Wilson DK & Nidetzky B (2006) Probing the substrate binding site of *Candida tenuis* xylose reductase (AKR2B5) with site-directed mutagenesis. *Biochem J* **393**, 51–58.

33 Olsen JG, Pedersen L, Christensen CL, Olsen O & Henriksen A (2008) Barley aldose reductase: structure, cofactor binding, and substrate recognition in the aldo/keto reductase 4C family. *Proteins Struct Funct Bioinf* **71**, 1572–1581.

34 Khurana S, Powers DB, Anderson S & Blaber M (1998) Crystal structure of 2,5-diketo-D-gluconic acid reductase A complexed with NADPH at 2.1 Å resolution. *Proc Natl Acad Sci U S A* **95**, 6768–6773.

35 Harrison DH, Bohren KM, Ringe D, Petsko GA & Gabbay KH (1994) An anion binding site in human aldose reductase: mechanistic implications for the binding of citrate, cacodylate, and glucose 6-phosphate. *Biochemistry* **33**, 2011–2020.

36 Kumar S & Gummadi SN (2011) Metabolism of glucose and xylose as single and mixed feed in *Debaryomyces nepalensis* NCYC 3413: production of industrially important metabolites. *Appl Microbiol Biotechnol* **89**, 1405–1415.
37 Paidimuddala B, Krishna AG & Gummadi SN (2017) A halotolerant aldose reductase from *Debaryomyces nepalensis*: gene isolation, overexpression and biochemical characterization. *RSC Adv* **7**, 20384–20393.
38 Neuhauser W, Haltrich D, Kulbe KD & Nidetzky B (1997) NAD(P)H-dependent aldose reductase from the xylose-assimilating yeast *Candida tenuis*. Isolation, characterization and biochemical properties of the enzyme. *Biochem J* **326**, 683–692.
39 Almarsson O & Bruice TC (1993) Evaluation of the factors influencing reactivity and stereospecificity in NAD(P)H dependent dehydrogenase enzymes. *J Am Chem Soc* **115**, 2125–2138.
40 Tsybovsky Y, Donato H, Krupenko NI, Davies C & Krupenko SA (2007) Crystal structures of the carboxyl terminal domain of rat 10-formyltetrahydrofolate dehydrogenase: implications for the catalytic mechanism of aldehyde dehydrogenases. *Biochemistry* **46**, 2917–2929.
41 Díaz-Sánchez ÁG, González-Segura L, Rudiño-Piñera E, Lira-Rocha A, Torres-Larios A & Muñoz-Clares RA (2011) Novel NADPH-cysteine covalent adduct found in the active site of an aldehyde dehydrogenase. *Biochem J* **439**, 443–455.
42 Park J & Rhee S (2013) Structural basis for a cofactor-dependent oxidation protection and catalysis of cyanobacterial succinic semialdehyde dehydrogenase. *J Biol Chem* **288**, 15760–15770.
43 Van Eys J & Kaplan NO (1957) The addition of sulfhydryl compounds to diphosphopyridine nucleotide and its analogues. *J Biol Chem* **228**, 305–314.
44 Lee YS, Hodoscek M, Brooks BR & Kador PF (1998) Catalytic mechanism of aldose reductase studied by the combined potentials of quantum mechanics and molecular mechanics. *Biophys Chem* **70**, 203–216.
45 Várnai P, Richards WG & Lyne PD (1999) Modelling the catalytic reaction in human aldose reductase. *Proteins Struct Funct Genet* **37**, 218–227.
46 Várnai P & Warshel A (2000) Computer simulation studies of the catalytic mechanism of human aldose reductase. *J Am Chem Soc* **122**, 3849–3860.
47 Sotriffer CA, Krämer O & Klebe G (2004) Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *Proteins Struct Funct Bioinf* **56**, 52–66.
48 Kolb P, Ferreira RS, Irwin JJ & Shoichet BK (2009) Docking and chemoinformatic screens for new ligands and targets. *Curr Opin Biotechnol* **20**, 429–436.
49 Dréanic MP, Edge CM & Tuttle T (2017) New insights into the catalytic mechanism of aldose reductase: a QM/MM study. *ACS Omega* **2**, 5737–5747.
50 Bohren KM, Grimshaw CE, Lai CJ, Harrison DH, Ringe D, Petsko GA & Gabbay KH (1994) Tyrosine-48 is the proton donor and histidine-110 directs substrate stereochemical selectivity in the reduction reaction of human aldose reductase: enzyme kinetics and crystal structure of the Y48H mutant enzyme. *Biochemistry* **33**, 2021–2032.
51 Cui Q, Elstner M & Karplus M (2002) A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *J Phys Chem B* **106**, 2721–2740.
52 Dzierlenga MW, Antoniou D & Schwartz SD (2015) Another look at the mechanisms of hydride transfer enzymes with quantum and classical transition path sampling. *J Phys Chem Lett* **6**, 1177–1181.
53 Blakeley MP, Ruiz F, Cachau R, Hazemann I, Meilleur F, Mitschler A, Ginell S, Afonine P, Ventura ON, cousido-Siah A *et al.* (2008) Quantum model of catalysis based on a mobile proton revealed by subatomic x-ray and neutron diffraction studies of h-aldose reductase. *Proc Natl Acad Sci U S A* **105**, 1844–1848.
54 Ye Q, Hyndman D, Green NC, Li L, Jia Z & Flynn TG (2001) The crystal structure of an aldehyde reductase Y50F mutant-NADP complex and its implications for substrate binding. *Chem Biol Interact* **130**, 651–658.
55 Dittmer DC & Kolyer JM (1963) Addition compounds of thiols and 1-substituted nicotinamides. *J Org Chem* **28**, 1720–1722.
56 Misra HP (1974) Generation of superoxide free radicals during the autoxidation of thiols. *J Biol Chem* **249**, 2151–2155.
57 Baccanari DP (1978) Coupled oxidation of NADPH with thiols at neutral pH. *Arch Biochem Biophys* **191**, 351–357.
58 Biadene M, Hazemann I, Cousido A, Ginell S, Joachimiak A, Sheldrick GM, Podjarny A & Schneider TR (2007) The atomic resolution structure of human aldose reductase reveals that rearrangement of a bound ligand allows the opening of the safety-belt loop. *Acta Crystallogr D* **63**, 665–672.
59 Giuseppe PO, de Santos MLD, Sousa SM, de Koch KE, Yunes JA, Aparicio R & Murakami MT (2016) A comparative structural analysis reveals distinctive features of co-factor binding and substrate specificity in plant aldo-keto reductases. *Biochem Biophys Res Commun* **474**, 696–701.
60 Zhang M, Jiang ST, Zheng Z, Li XJ, Luo SZ & Wu XF (2015) Cloning, expression, and characterization of a novel xylose reductase from *Rhizopus oryzae*. *J Basic Microbiol* **55**, 907–921.

61 Khoury GA, Fazelinia H, Chin JW, Pantazes RJ, Cirino PC & Maranas CD (2009) Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Sci* **18**, 2125–2138.
62 Banta S, Swanson BA, Wu S, Jarnagin A & Anderson S (2002) Alteration of the specificity of the cofactor-binding pocket of *Corynebacterium* 2,5-diketo-D-gluconic acid reductase A. *Protein Eng* **15**, 131–140.
63 Hoog SS, Pawlowski JE, Alzari PM, Penning TM & Lewis M (1994) Three-dimensional structure of rat liver 3α-hydroxysteroid/dihydrodiol dehydrogenase: a member of the aldo-keto reductase superfamily. *Proc Natl Acad Sci U S A* **91**, 2517–2521.
64 Asako H, Shimizu M & Itoh N (2008) Engineering of NADPH-dependent aldo-keto reductase from *Penicillium citrinum* by directed evolution to improve thermostability and enantioselectivity. *Appl Microbiol Biotechnol* **80**, 805–812.
65 Kratzer R & Nidetzky B (2007) Identification of *Candida tenuis* xylose reductase as highly selective biocatalyst for the synthesis of aromatic alpha-hydroxy esters and improvement of its efficiency by protein engineering. *Chem Commun* **14**, 1047–1049.
66 Wang Z, Zhou S, Zhang S, Zhang S, Zhu F, Jin X, Chen Z & Xu X (2017) Semi-rational engineering of a thermostable aldo-keto reductase from *Thermotoga maritima* for synthesis of enantiopure ethyl-2-hydroxy-4-phenylbutyrate (EHPB). *Sci Rep* **7**, 4007.
67 Krump C, Vogl M, Brecker L, Nidetzky B & Kratzer R (2014) Acceleration of an aldo-keto reductase by minimal loop engineering. *Protein Eng Des Sel* **27**, 245–248.
68 Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ & Klenk DC (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* **150**, 76–85.
69 Cook PF & Cleland WW (2007) Enzyme Kinetics and Mechanism. Garland Science, New York, NY.
70 Kabsch W (1988) Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J Appl Crystallogr* **21**, 916–924.
71 Battye TGG, Kontogiannis L, Johnson O, Powell HR & Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D* **67**, 271–281.
72 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A *et al.* (2011) Overview of the *CCP* 4 suite and current developments. *Acta Crystallogr D* **67**, 235–242.
73 McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with *Phaser*. *Acta Crystallogr D* **63**, 32–41.
74 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213–221.
75 Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* **68**, 352–367.
76 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126–2132.
77 Moriarty NW, Grosse-Kunstleve RW & Adams PD (2009) Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D* **65**, 1074–1080.
78 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* **66**, 12–21.
79 DeLano WL (2006) The PyMOL molecular graphics system http://www.pymol.org.
80 Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, 320–324.
81 Diederichs K & Karplus PA (2013) Better models by discarding data? *Acta Crystallogr D* **69**, 1215–1222.