SCIENTIFIC REPORTS

OPEN

Structural insight into D-xylose utilization by xylose reductase from *Scheffersomyces stipitis*

Received: 7 March 2018
Accepted: 9 November 2018
Published online: 28 November 2018

Hyeoncheol Francis Son[1,2], Sun-Mi Lee[3] & Kyung-Jin Kim[1,2]

Lignocellulosic biomass, of which D-xylose accounts for approximately 35% of the total sugar, has attracted attention as a future energy source for biofuel. To elucidate molecular mechanism of D-xylose utilization, we determined the crystal structure of D-xylose reductase from *Schefferzomyces stipitis* (*Ss*XR) at a 1.95 Å resolution. We also determined the *Ss*XR structure in complex with the NADPH cofactor and revealed that the protein undergoes an open/closed conformation change upon NADPH binding. The substrate binding pocket of *Ss*XR is somewhat hydrophobic, which seems to result in low binding affinity to the substrate. Phylogenetic tree analysis showed that AKR enzymes annotated with bacterial/archaeal XRs belonged to uncharacterized AKR families and might have no XR function, and yeast/fungi derived enzymes, which belong to the same group with *Ss*XR, can be candidates for XR to increase xylose consumption.

Industrialization and population growth lead to increased energy demands, and interest in alternative energy has increased because of fossil fuel depletion and environmental problems. Lignocellulosic biomass has attracted attention as a future energy source for biofuel and polymer production, because it is abundant, sustainable and does not compete with edible resources[1–3]. Particularly, the recent application of lignocellulosic biomass has expanded not only to conventional bioethanol but also to biodiesel production, and thus numerous studies have been conducted to solve the energy problem using ligonocellulosic biomass[4]. However, biofuel producing organisms, such as *Saccharomyces cerevisiae* and *Yarrowia lipolytica*, lack the pentose metabolic pathway, which limits the use of lignocellulosic biomass as a resource[4,5]. Therefore, many studies have been conducted to overcome the limitation of pentose utilization by introducing the pentose metabolic pathway from heterologous organisms[5–11].

D-Xylose, which accounts for approximately 35% of the total sugar in lignocellulosic biomass, is considered a good industrial resource[12–15]. D-Xylose can be converted to D-xylulose by two different biological pathways: the oxidoreductase pathway and isomerase pathway (Fig. 1a). In the isomerase pathway, xylose isomerase (XI) directly converts D-xylose to D-xylulose, and has a high theoretical yield because XI enzyme does not require any cofactors. Therefore, various heterologous XI enzymes have been introduced into *S. cerevisiae* and *Y. lipolytica* to assimilate D-xylose and recently, the crystal structures of XI from *Piromyces* sp. (*Ps*XI) were determined[4,16–22]. In the oxidoreductase pathway, two enzymes, xylose reductase (XR) and xylitol dehydrogenase (XDH), are used to convert D-xylose to D-xylulose using xylitol as an intermediate (Fig. 1a), and XR/XDH enzymes derived from *Scherrsomyces stipitis* (*Ss*XR, *Ss*XDH, respectively) have been introduced into *S. cerevisiae* and *Y. lipolytica* for D-xylose intake[4,23,24]. XR/XDH enzymes can offer higher metabolic fluxes than the XI enzymes, however D-xylose intake by the oxidoreductase pathway causes cofactor imbalance, because XR and XDH use NADPH and NAD^+ as cofactors, respectively (Fig. 1a). Although several protein engineering trials to change the cofactor specificity were performed[25–28], there has been not fundamental solution for cofactor imbalance problem, due to the lack of the crystal structure of *Ss*XR and *Ss*XDH.

Aldo-keto reductases (AKRs), which include D-xylose reductase, are found in almost all organisms, and has broad substrate specificity for sugars, lipid aldehydes, ketosteroids, ketoprostaglandins, and chemical carcinogens as substrates[29–34]. Based on amino acid sequence similarities, more than 190 AKR members are divided into 16 families, with XR enzymes belonging to the AKR family 2B[35]. Until now, only one XR structure derived

[1]School of Life Sciences, KNU Creative BioResearch Group, Kyungpook National University, Daegu, 41566, Republic of Korea. [2]KNU Institute for Microorganisms, Kyungpook National University, Daegu, 41566, Republic of Korea. [3]Clean Energy Research Center, Korea Institute of Science and Technology (KIST), Seoul, 02792, Republic of Korea. Correspondence and requests for materials should be addressed to S.-M.L. (email: smlee@kist.re.kr) or K.-J.K. (email: kkim@knu.ac.kr)

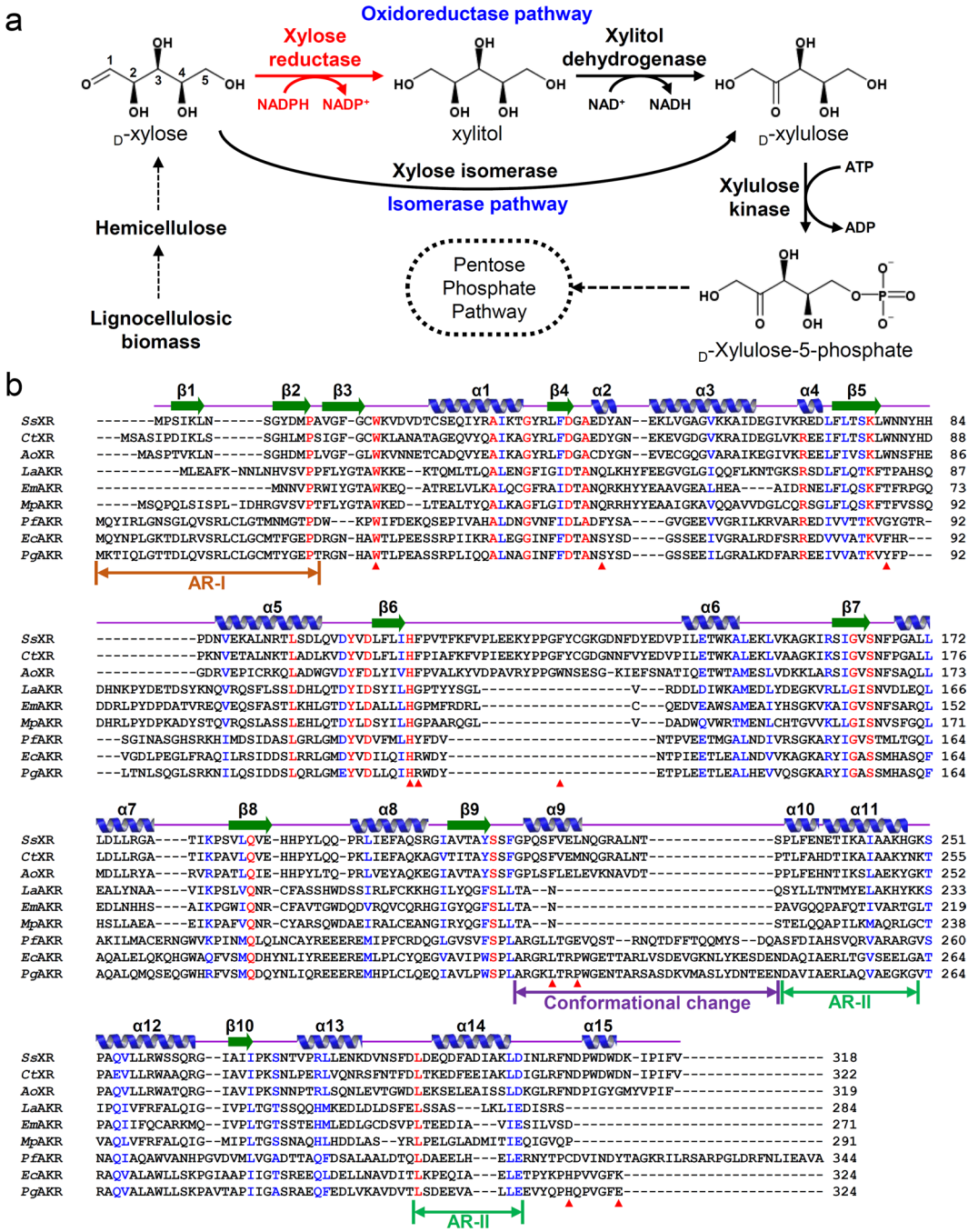

Figure 1. D-Xylose metabolic pathway and amino acid sequence alignment of XRs. (**a**) D-xylose metabolic pathway. (**b**) Amino acid sequence alignment of XRs, UNC AKR I, and UNC AKR II. The secondary structure elements are drawn based on the structure of *Ss*XR in the closed conformation. The residues involved in the formation of D-xylose substrate binding pocket of *Ss*XR are indicated by red colored triangles. *Ss*XR, *Ct*XR, and *Ao*XR are abbreviation of XRs from *Scheffersomyces stipitis*, *Candida tenuis*, and *Aspergillus oryzae*, respectively. *Pf*AKR, *Ec*AKR, *Pa*AKR, *La*AKR, *Em*AKR, *Mp*AKR are abbreviation of AKR from *Legionella anisa*, *Endozoicomonas montiporae*, *Methylovulum psychrotolerans*, *Pseudomonas fluorescens*, *Escherichia coli*, and *Pantoea agglomerans*, respectively.

from *Candida tenuis* (*Ct*XR) was reported[36], however, the crystal structure of *Ss*XR used in D-xylose consumption by introduction into *S. cerevisiae* and *Y. lipolytica* has not been determined yet. Here, we report the crystal structures of D-xylose reductase from *S. stipitis* (*Ss*XR) in its apo form and in complex with NADPH cofactor. We revealed that the protein undergoes an open/closed conformation change upon the binding of NADPH and has somewhat hydrophobic D-xylose binding pocket. Phylogenetic tree analysis showed that bacteria/archaea derived XR-annotated AKRs might have no XR activity, and yeast/fungi derived XRs can be candidate to increase xylose consumption.

	*Ss*XR_Apo	*Ss*XR_NADPH
Data collection		
Space group	$P2_12_12_1$	$P4_22_12$
Cell dimensions		
a, *b*, *c* (Å)	69.2,87.2,122.6	97.7,97.7,160.1
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00-1.95 (1.98–1.95)	50.00–2.00 (2.03–2.00)
R_{sym} or R_{merge} (%)	9.7 (20.8)	11.1 (27.3)
R_{meas} (%)	10.6(23.0)	11.6 (29.8)
I/σ (I)	55.2 (15.1)	26.3 (3.3)
Completeness (%)	98.8 (99.8)	98.6 (96.6)
Redundancy	6.7 (5.4)	8.7 (5.4)
Refinement		
Resolution (Å)	50.00–1.95	50.00–2.00
No. reflections	51177	49336
R_{work}/R_{free}	18.9 (23.6)	17.7 (22.4)
No. atoms	5491	5526
Protein	5084	5068
Ligand/ion	12	96
Water	395	362
B-factors (Å^2)	36.514	29.337
Protein	36.152	28.988
Ligand/ion	37.886	38.841
Water	42.366	34.135
R.m.s. deviations		
Bond lengths (Å)	0.0184	0.0189
Bond angles (°)	1.7601	1.983
Ramachandran statistics		
Favoured (%)	98	96
Allowed (%)	2	4
Outliers (%)	0	0

Table 1. Data collection and refinement statistics.

Results and Discussion

Overall structure of *Ss*XR. To elucidate the molecular mechanism of *Ss*XR, we determined its crystal structure at 1.95 Å resolution. The refined structure was in good agreement with the X-ray crystallographic statistics for bond angles, bond lengths, and other geometric parameters (Table 1). The overall structure of *Ss*XR is similar to that of the AKR superfamily enzymes. The monomeric structure of *Ss*XR is composed of 15 α-helices (α1-α15) and 10 β-strands (β1-β10) (Fig. 1b). The monomeric structure of *Ss*XR consists of a core domain and two auxiliary regions (ARs), AR-I and AR-II. The core domain consists of 13 α-helices (α1-α10, α12-α13, and α15) and eight β-strands (β3-β10) and forms a TIM-barrel motif (Figs 1b and 2a). As the conventional TIM-barrel motif, in *Ss*XR eight parallel β-strands (β3-β10) are arranged in a cylindrical shape with eight surrounding α-helices (α1, α3, α5-α8, and α12-α13). Four α-helices (α2, α9-α10, and α15) are located at the back of the TIM-barrel motif and contribute to binding of the NADPH cofactor (Fig. 2a). AR-I is composed of two β-strands (β1-β2) and is located on the opposite side of the TIM-barrel. AR-II consists of two α-helices (α11 and α14) and is positioned next to the α12 helix (Fig. 2a). As reported in other enzymes belonging to AKR families, four catalytic residues, Asp43, Tyr48, Lys77, and His110, are also conserved in the *Ss*XR, and catalytic mechanism can be proposed[37] (Fig. 2b).

It has been reported that AKR enzymes function as a monomer, dimer, and tetramer[37] and *Ss*XR is known to exist as a dimer in the presence of 50 mM NaCl[38]. In our current structure, there are two *Ss*XR polypeptide chains in an asymmetric unit forming a dimer. PISA and EPPIC calculations also showed that the protein exists as a dimer, although the interaction was quite low. We then performed size-exclusion chromatography experiments under various NaCl concentrations. Unexpectedly, *Ss*XR generally separated into monomers in 50 and 100 mM NaCl and completely separated into monomers in150 mM NaCl, although it existed as dimers in the absence of NaCl (Figs 2c, S1). Based on these results, we suggest that *Ss*XR exists as a monomer under the physiological NaCl concentration and tends to form a dimer in the presence of low NaCl concentrations. Furthermore, in order to compare the enzyme activity according to the oligomec state, kinetic values at 0 mM and 150 mM were measured. Interestingly, all values at 0 mM NaCl condition were better than 150 mM NaCl, indicating that oligomer formation affects enzyme activity.

Open/closed conformation for cofactor binding of *Ss*XR. AKR superfamily enzymes utilize NADH and/or NADPH as a cofactor. To examine the cofactor preference of *Ss*XR, we performed kinetic analysis using

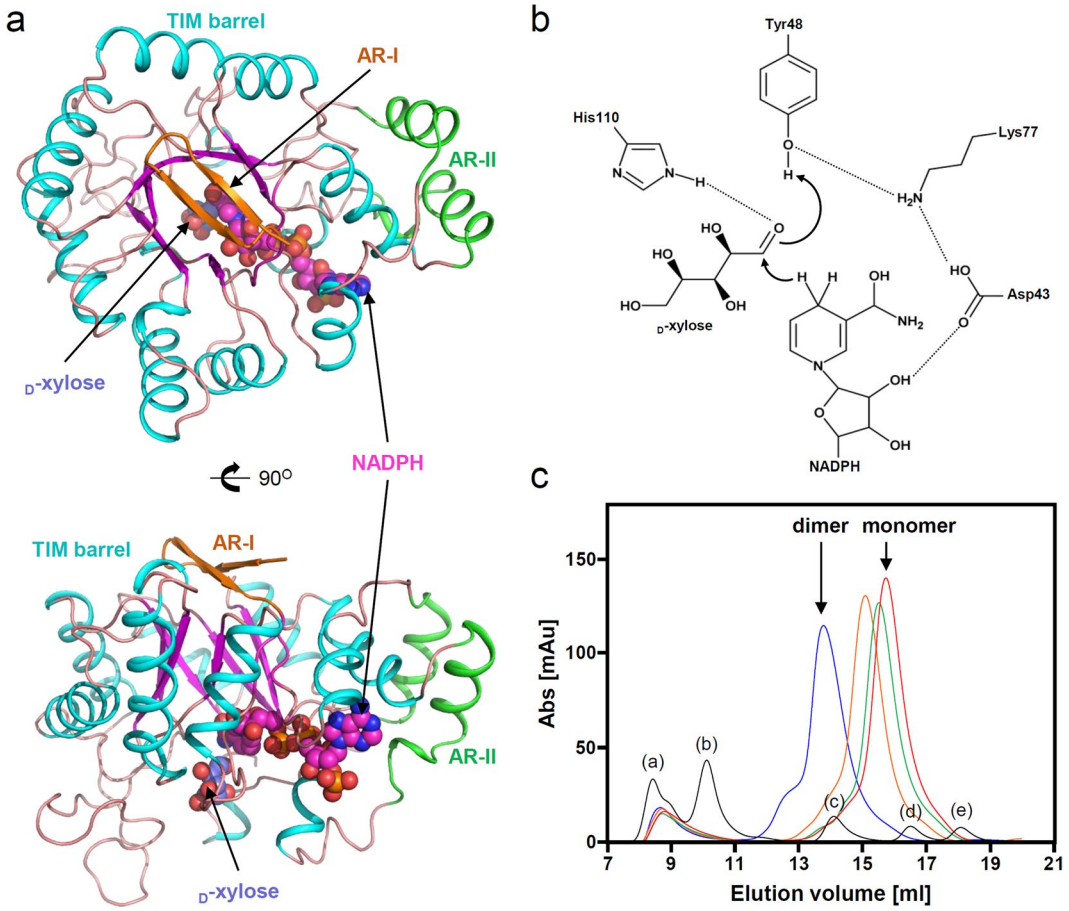

Figure 2. Overall structure of *Ss*XR. (**a**) The monomeric structure of *Ss*XR. The monomeric structure of *Ss*XR is presented as a cartoon diagram. α-helices, β-strands, and loops in core domain are distinguished cyan, magenta, and orange colors, respectively, and auxiliary region I, and II (AR-I and AR-II) are distinguished with orange, and green colors, respectively, and labeled. The bound NADPH cofactors and modeled D-xylose substrates are presented as sphere models with magenta and light blue colors, respectively. The bottom figure is rotated by 90 degree horizontally from the top figure. (**b**) Putative catalytic mechanism of *Ss*XR. Dashed likes show hydrogen bonds. (**c**) Size-exclusion chromatography of *Ss*XR. The *Ss*XR samples in condition of 0, 50, 100, and 150 mM NaCl are distinguished with blue, orange, green, and red colors, respectively. The *Ss*XR with 0 and 150 mM NaCl are eluted as a dimeric and monomeric form, respectively. The standard graph was drawn with black color. (**a**) Indicates void volume and (**b**–**d**), € indicate the standard samples of ferritin (440 kDa), conalbumin (75 kDa), carbonic anhydratse (29 kDa), and ribonuclease A (13.7 kDa), respectively.

		K_m [mM]	k_{cat} [s^{-1}]
0 mM NaCl	NADPH	0.00928 ± 0.00108	7.65 ± 0.23
	NADH	0.0187 ± 0.0026	2.68 ± 0.11
	$xylose^{NADPH}$	32.37 ± 3.02	8.37 ± 0.26
	$xylose^{NADH}$	39.61 ± 3.56	4.34 ± 0.14
150 mM NaCl	NADPH	0.0277 ± 0.0020	7.63 ± 0.19
	NADH	0.136 ± 0.017	2.39 ± 0.15
	$xylose^{NADPH}$	39.40 ± 2.65	6.69 ± 0.16
	$xylose^{NADH}$	59.72 ± 9.76	3.02 ± 0.20

Table 2. The kinetic analysis of *Ss*XR.

NADH and NADPH as cofactors. At the physiological condition, the K_m values for NADPH and NADH were 0.0277 and 0.136 mM, respectively, and the k_{cat} values for NADPH and NADH were 7.63 and 2.39s^{-1}, respectively (Table 2). These results indicate that *Ss*XR prefers NADPH as a cofactor and has a 15-fold higher k_{cat}/K_m value for NADPH than for NADH. To elucidate the cofactor binding mode of *Ss*XR, we determined the crystal structure of *Ss*XR in complex with NADPH at 2.0 Å resolution (Table 1, Fig. 3a). The NADPH cofactor bound at the back of the TIM barrel motif (Fig. 2a). The residues Phe216, Gln219, Glu223, Phe236, Ala253, Lys270, Asn272, Arg276,

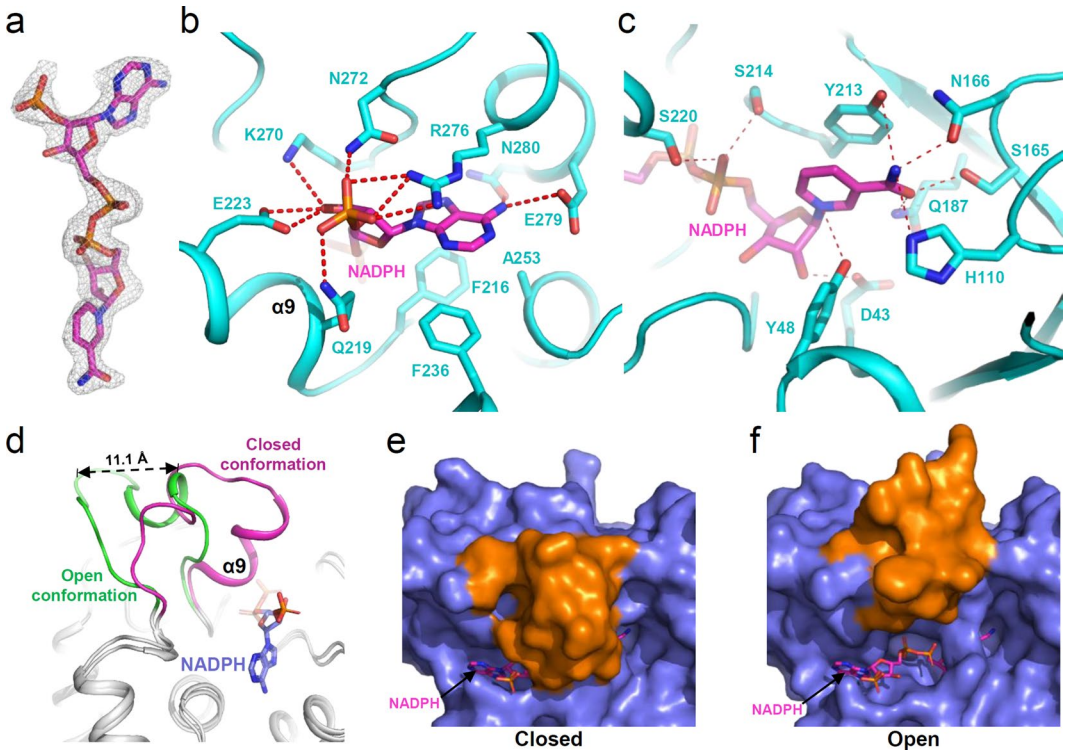

Figure 3. NADPH cofactor binding mode of *Ss*XR. (**a**) Electron density map of the bound NADPH cofactor. The Fo-Fc electron density map of NADPH cofactor is shown with a gray-colored mesh, and contour 2.5 σ. The NADPH cofactor is shown as a stick model with magenta color. (b,c) NADPH cofactor binding mode of *Ss*XR. The binding mode of nucleotide part (**b**), and pyrophosphate, ribose ring, and nicotinamide ring (**c**) of *Ss*XR. The structure of *Ss*XR is shown as a cartoon diagram with cyan color. The residues involved in the NADPH binding are shown as stick models and labeled appropriately. The bound NADPH cofactor is shown as a stick model with a magenta color. Red color dotted lines indicate hydrogen bonds contributing to NADPH binding. (**d**) Structural comparison of the open and closed conformation of *Ss*XR. The TIM-barrel motif of the open and closed conformation of *Ss*XR were superposed, and the relative positions of the α9 were compared. The moving part of the open and closed conformation of *Ss*XR were with green and magenta colors, respectively, and labeled. The bound NADPH cofactor was shown as sphere model with light-blue color. (**e**,**f**) The open and closed conformation of *Ss*XR. The closed (**e**) and open (**f**) conformations of *Ss*XR were shown as surface model. The moving part and rigid body were distinguished with orange and light-blue colors, respectively. The bound NADPH cofactors were shown as stick model with magenta color.

Glu279, and Asn280 contribute to formation of the nucleotide binding pocket (Fig. 3b). The adenine ring is stabilized by Glu279 and Asn280 and the 2′-phosphate group of NADPH is stabilized by Gln219, Asn272, and Arg276 through hydrogen bonds. The Glu223 and Lys270 residues form hydrogen bonds with the ribose moiety (Fig. 3b). Two serine residues, Ser214 and Ser220, are involved in stabilizing pyrophosphate moiety and the nicotinamide ring of NADPH is stabilized by the residues Asp43, Tyr48, His110, Ser165, Asn166, Gln187, and Tyr213 through hydrogen bond networks (Fig. 3c).

Interestingly, when we superimposed the *Ss*XR structure in the apo-form with that in complex with NADPH, we observed a large structural difference near the NADPH binding site between these two structures (Fig. 3d). In the NADPH-bound form, the α9 helix and α9-α10 connecting loop (Gly217-Ser232) were in the closed conformation without crystal contact, and the Gln219 and Glu223 residues form hydrogen bonds with the phosphate atom of the phosphoribose moiety (Fig. 3b,d,e). In contrast, the region moved away from the NADPH binding site by 11.1 Å and formed an open conformation (Fig. 3d,f). Based on these observations, we propose that *Ss*XR undergoes an open/closed conformational change upon binding of the NADPH cofactor.

$_D$-Xylose binding mode of *Ss*XR. To elucidate the binding mode of the $_D$-xylose substrate, we performed molecular docking simulations of *Ss*XR with the $_D$-xylose substrate. Molecular docking simulations revealed that the $_D$-xylose substrate fits well into the predicted substrate binding pocket (Fig. 4a). The $_D$-xylose binding pocket is constituted by 10 residues: Trp20, Asp47, Trp79, His110, Phe111, Phe128, Phe221, Leu224, Asn306, and Trp311 (Fig. 4a,b). The Asp47 residue contributes to the stabilization of two hydroxyl groups (OH2 and OH3), and the aldehyde group of $_D$-xylose is stabilized by Asn306 through hydrogen bonding (Fig. 4b). The residues involved in formation of the $_D$-xylose binding pocket were confirmed by site-directed mutagenesis experiments (Fig. 4c). Most mutants, including W20A, D47A W79A, H110A, F111A, F128A, F221A, N306A, and W311A, exhibited almost complete loss of enzyme activity, indicating that these residues are crucial for $_D$-xylose binding. However, the L224A mutant showed 35% XR activity compared to the wild type, indicating that the Leu224 residue does

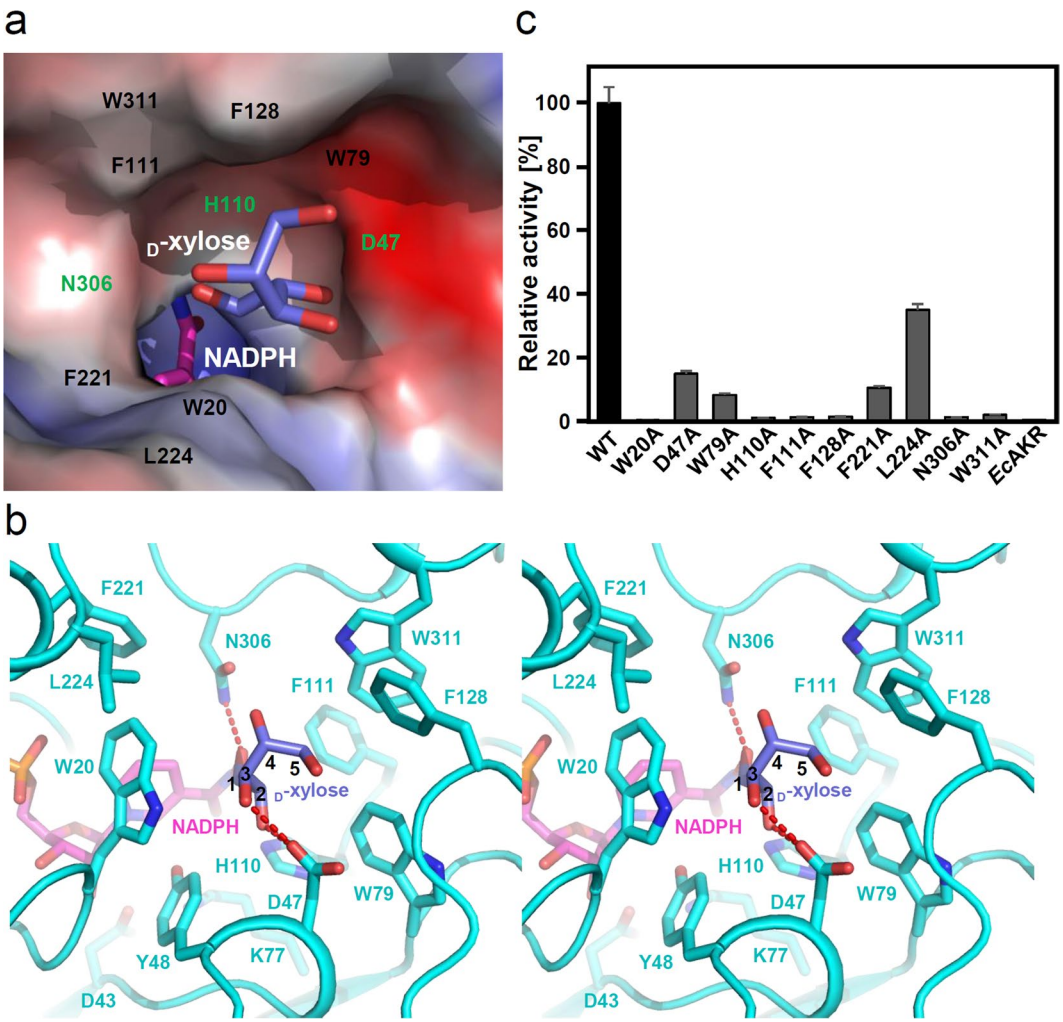

Figure 4. Substrate binding mode of *Ss*XR. (**a**) Electrostatic potential surface presentation of the D-xylose substrate binding mode of *Ss*XR. The *Ss*XR structure is shown as an electrostatic potential surface presentation using PYMOL software. The NADPH cofactor and D-xylose substrate presented by stick models with magenta and light-blue colors, respectively, and labeled. The sites of hydrophilic and hydrophobic residues involved in the substrate binding pocket formation were labelled with green and black colors, respectively. (**b**) Stereo-view of substrate binding mode of *Ss*XR. The *Ss*XR structure is shown as a cartoon diagram with cyan color. The residues involved in the substrate binding and catalysis are shown as stick models and labeled appropriately. The bound NADPH cofactor and modeled D-xylose substrate were shown as stick model, with magenta and light-blue colors, respectively, and labeled. Hydrogen bonds involved in the substrate binding are shown with red-colored dotted lines. 1 to 5 indicate the carbon number of the xylose substrate. (**c**) Site-directed mutagenesis of *Ss*XR. Residues involved in the D-xylose substrate binding are replaced by alanine residues. The relative activities of recombinant mutant proteins were measured and compared with that of wild-type *Ss*XR.

not influence substrate binding as much as other residues (Fig. 4b,c). Interestingly, the D-xylose binding pocket is somewhat hydrophobic and seven of 10 residues involved in formation of the D-xylose binding pocket are hydrophobic (Fig. 4a,b). Considering that the D-xylose substrate is highly hydrophilic with four hydroxyl groups and one aldehyde group, we can predict that the substrate binding pocket of *Ss*XR requires more hydrophilic residues. Thus, the hydrophobicity of the substrate binding site of *Ss*XR indicates that the binding affinity of the enzyme for D-xylose is low. We then performed kinetic analysis of the enzyme and the Km value for D-xylose was extremely high with a value of 39.4 mM (Table 2). Because the Km value of D-xylose was significantly higher than that of the NADPH cofactor (0.0277 mM) (Table 2), the bottleneck of the enzyme activity in *Ss*XR appears to be the binding affinity for D-xylose.

Phylogenetic tree analysis and classification of XRs. Because utilization of D-xylose by bio-fuel producing organisms, such as *S. cerevisiae* and *Y. lipolytica*, requires heterologous introduction of XR enzymes, selection of highly efficient XR is a key issue in high bio-fuel production[4]. In order to explore the XR candidates that can be introduced to microorganisms for xylose consumption, enzymes annotated as XR in NCBI and UNIPROT server were analyzed (Table. S1). There are 161 AKR enzymes (67 yeast/fungal, 83 bacterial, and 11 archaeal

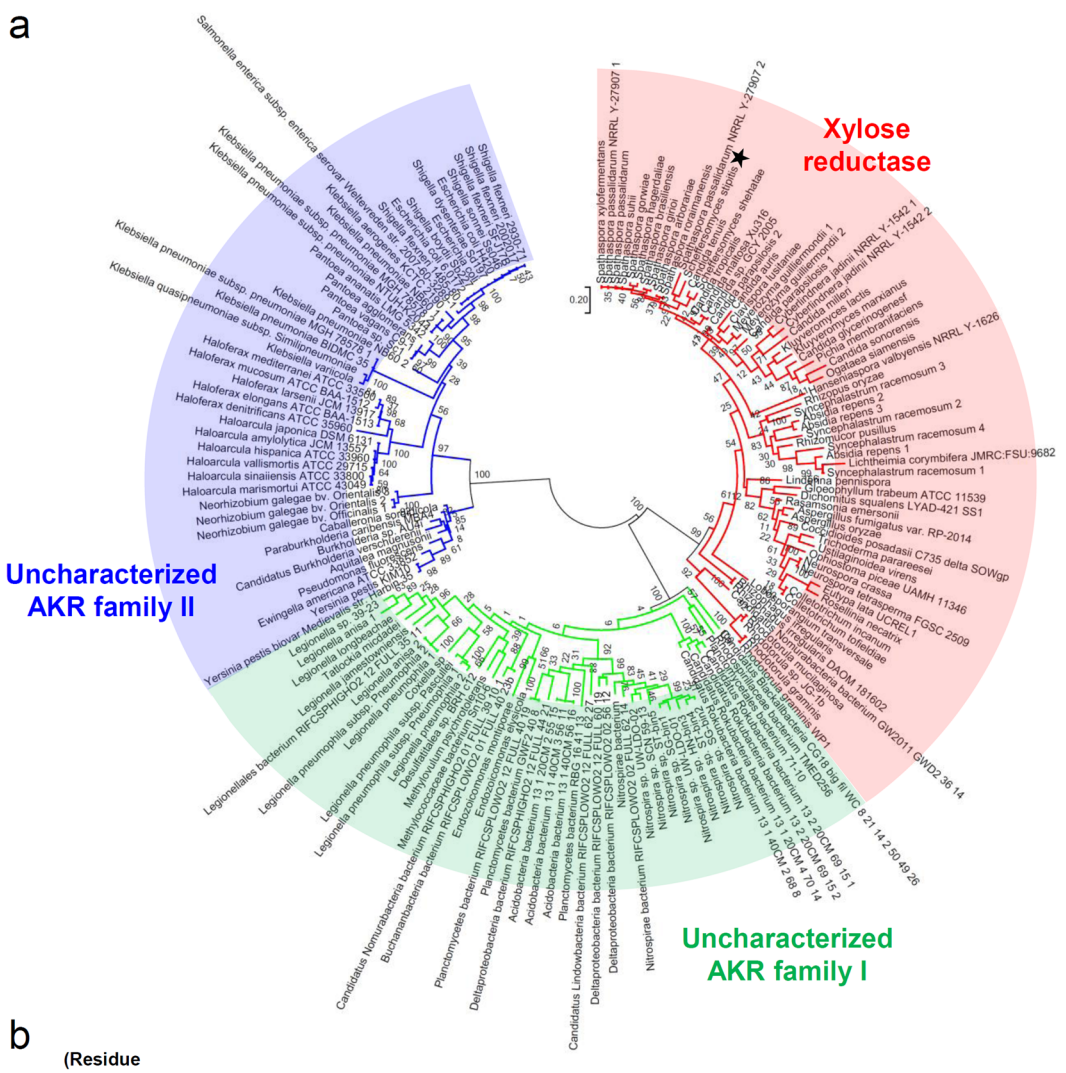

Figure 5. Phylogenetic analysis of XR proteins. (**a**) Unrooted maximum Likelihood tree of annotated XRs in NCBI. XRs, uncharacterized AKR family I and II are labeled as red, green, and blue colors, respectively, and labeled. *Ss*XR indicates by black colored star. The accession codes are listed at Supporting Table S1. (**b**) Amino acid sequence alignment of the residues involved in the substrate binding. Each amino acid is represented by single letter code, and 'Va' is an abbreviation for 'Various'.

enzymes) annotated with XR, and these enzymes can be divided into three groups by phylogenetic tree analysis (Fig. 5a). Interestingly, only yeast and fungi derived enzymes belonged to the same group with *Ss*XR, and all other enzymes from bacteria and archaea belonged to the different group from *Ss*XR. Furthermore, amino acid sequence alignment also showed that 161 AKR enzymes were classified in the same way as the phylogenetic analysis, and only yeast/fungi derived enzymes have amino acid sequence similar to *Ss*XR (Fig. S2). In particular, when the residues involved in the substrate binding of *Ss*XR were compared each other, there are many gaps in bacterial/archaeal AKR enzymes (Fig. 5b). These observations indicate that AKR enzymes annotated with bacterial/archaeal XRs belonged to uncharacterized AKR families (UNC AKR I and II) and might have no XR function. In order to confirm that the enzymes belonging to UNK AKR families cannot use xylose as a substrate, we measured an XR activity using purified AKR from *E. coli*, an enzyme belonging to UNK ASK II (Accession codes of NCBI and UNIPROT are WP_001199831 and C3TM25, respectively), and the enzyme showed no enzyme

activity against xylose (Fig. 4c). Consequently, we propose that the yeast/fungi derived enzymes, which belong to the same group with *Ss*XR, can be candidates for XR. In addition, when we compared to the amino acid residues involved in the formation of xylose binding pockets, only four of ten residues are conserved in the yeast/fungi derived XRs, and various amino acids are positioned in the xylose binding pocket (Figs 5b, S2). Therefore, structural and biochemical studies on other yeast/fungi derived XRs are required to select more efficient XRs and to increase biofuel productivity using lignocellulosic biomass.

We report the crystal structure of *Ss*XR in the apo form and in complex with the NADPH cofactor and provide structural insight into the open/closed conformational change upon cofactor binding. Molecular docking simulations of *Ss*XR for the $_{D}$-xylose substrate and kinetic analysis of the protein revealed that why *Ss*XR shows low binding affinity for the substrate. Moreover, based on phylogenetic tree analysis and detailed amino acid sequence alignment, we propose that various yeast/fungi derived XRs can be candidate for more efficient XRs. These results may be useful for developing methods with much higher $_{D}$-xylose utilization for lignocellulosic biomass in biofuel production.

Material and Methods

Cloning, expression and purification. The *Ss*XR coding gene was amplified through PCR from synthetic gene by Bioneer (Republic of Korea), and primers (forward: 5′-GCGC**CATATG**CCTTCTATTAAGTTGAACTC TGGTTAC-3′, reverse: 5′-GCGC**CTCGAG**TTAGACGAAGATAGGAATCTTGTCC-3′). The PCR product was digested by restriction enzymes (NdeI and XhoI), and sub-cloned into pProEX-HTa vector (Thermo Fischer Scientific), which contained a 6x -His tag with rTEV protease cleavage site at N-terminus. The pProEX-HTa:*Ss*XR was transformed into a *E. coli* BL21(DE3)-T1^{R} strain, and transformed *E. coli* was grown to an OD_{600} of 0.65 in LB medium containing 100 mg L^{-1} ampicillin at 310 K and *Ss*XR protein expression was induced by 0.5 mM isopropyl β-D-1-thiogalatopyranoside (IPTG). After 18 h at 293 K, the cell was harvested at 277 K. The cell pellet was resuspended in 40 mM Tris-HCl, pH 8.0 (buffer A) and disrupted by ultrasonication. The cell debris was removed by centrifugation at 13,000 g for 30 min, and the lysate was applied onto a Ni-NTA agarose column (Qiagen). After washing with 40 mM Tris-HCl, pH 8.0 and 25 mM Imidazole (buffer B), the bound proteins were eluted with 40 mM Tris-HCl, pH 8.0 and 300 mM Imidazole (buffer C). Finally, the trace amount of contaminants was removed by gel filtration experiment using HiPrep 26/60 Sephacryl S-300 HR column (GE Healthcare Life Sciences) equilibrate with buffer A. The eluted protein had about 75 kDa molecular weight, indicating a dimeric structure. The protein was concentrated to 75 mg mL^{-1}, and kept at 193 K for further experiments.

Crystallization of *Ss*XR. Crystallization of the *Ss*XR protein was initially performed with commercially available sparse-matrix screens, including PEG ion I and II, and Index, (Hampton Research), and Wizard Classic I and II (Rigaku Reagents), using the sitting-drop vapor diffusion method at 295 K. Each experiment consisted of mixing 1.0 μL protein solution (75 mg mL^{-1}, buffer A) with 1.0 μL reservoir solution and then equilibrating against 50 uL reservoir solution. The apo-form of *Ss*XR crystals were observed from various crystallization conditions. After several steps for crystal improvement, crystals of the best quality appeared in 22% (w/v) polyethylene glycol 3350 and 0.1 M ammonium citrate tribasic, pH 7.0. For the crystal of the *Ss*XR in complex with NADPH cofactor, the *Ss*XR proteins was prepared in the same manner as that of apo form. The *Ss*XR protein was mixed with 20 mM NADPH cofactor, and incubated for 1 hr at 277 K. *Ss*XR crystals in complex with NADPH were crystallized in the condition of 20% (w/v) polyethylene glycol 3350 and 8% (v/v) Tacsimate, pH 7.0.

Data collection and structure determination. The crystals of *Ss*XR in its apo form and in complex with NADPH were transferred to cryo-protectant solution containing 30% (v/v) glycerol with the crystallization buffer condition. The crystals were mounted in a 100 K nitrogen stream. Data of apo and NADPH complex were collected to a resolution of 1.95 and 2.0 Å, respectively, at 7 A beamline of the Pohang Accelerator Laboratory (PAL, Pohang, Republic of Korea), using a Quantum 270 CCD detector (ADSC, USA)[39]. All data were indexed, integrated, and scaled together using the HKL2000 software package[40]. The apo-form crystals of *Ss*XR belonged $P2_12_12_1$ with unit cell parameters a = 69.248 Å, b = 87.151 Å, c = 122.62 Å, α = β = γ = 90°. Assuming two *Ss*XR molecules in asymmetric unit, the crystal volume per unit of protein mass was 2.58 $Å^3$ Da^{-1}, which means the solvent content was approximately 52.29%[41]. The crystals of *Ss*XR in complex with NADPH belonged to the space group $P4_22_12$, with unit cell parameters a = b = 97.654 Å, c = 160.12 Å, α = β = γ = 90°. Assuming two *Ss*XR molecules in asymmetric unit, the crystal volume per unit of protein mass was 2.73 $Å^3$ Da^{-1}, which means the solvent content was approximately 54.92%. The apo form structure of *Ss*XR was determined by molecular replacement with the CCP4 version of MOLREP[42] using the structure of xylose reductase from *Candida tenuis* (*Ct*XR, PDB code 1Z9A) as a search model[43]. Model building was performed manually using WinCoot software [44], and refinement was performed with refmac5 [45]. The structure of *Ss*XR in complex with NADPH was determined by molecular replacement using the crystal structure of the apo-form of *Ss*XR (PDB code 5Z6U). The data statistics are summarized in Table 1. The *Ss*XR structures of apo-form and in complex with NADPH were deposited in the protein data bank with PDB codes of 5Z6U and 5Z6T, respectively.

Size-exclusion chromatographic analysis. To investigate the oligomeric state of *Ss*XR, analytical size-exclusion chromatography was performed using a Superdex increase 200 10/300 GL column (GE Healthcare Life Sciences) equilibrated with 40 mM Tris-HCl, pH 8.0 and various NaCl concentrations, such as 0, 50, 100, and 150 mM. Protein sample of 0.5 mL with concentration of 1 mg mL^{-1} was analyzed. In order to calculate the molecular weight of eluted *Ss*XR sample, the calibration curve was drawn using standard sample of ferritin (440 kDa), conalbumin (75 kDa), carbonic anhydrate (29 kDa), and ribonuclease A (13.7 kDa) (GE Healthcare Life Sciences).

Kinetic analysis. For the kinetic analysis, the *Ss*XR protein was purified in the same manner as the protein for crystallization. The activities of *Ss*XR were determined by measuring the decrease of absorbance at 340 nm (extinction coefficient of $6.22 \times 10^3\,M^{-1}\,cm^{-1}$). For the kinetic analysis of *Ss*XR on NADPH and NADH cofactor, enzyme activity was measured with a reaction mixture of 0.5 ml total volume at 303 K. The reaction mixture contained 100 mM Tris-HCl, pH 8.0, 100 mM $_D$-xylose, and various concentration of NADPH/NADH cofactor from 1 to 200 μM. The reactions were initiated by the addition of enzyme to a final concentration of 0.5 and 2 μM for the analysis of NADPH and NADH, respectively. For the kinetic analysis of *Ss*XR on $_D$-xylose substrate, enzyme reactions were performed with a reaction mixture of 0.5 ml total volume at 303 K. The reaction mixture contained 100 mM Tris-HCl, pH 8.0, 0.2 mM NADPH, and various concentrations of $_D$-xylose substrate from 1 to 200 mM. The reaction was initiated by the addition of enzyme to a final concentration of 2 μM. The *Ss*XR activity measurement was performed in duplicate reaction.

Molecular docking simulation. Molecular docking simulations of $_D$-xylose to *Ss*XR structure was performed using AutoDock Vina software[46]. *Ss*XR structure of PDB code 5Z6T, in complex with NADPH, was used and the $_D$-xylose ligand was prepared using the MarvinScketch software. The *pdbqt* files were generated by AutoDock Vina manual. Side-chain of Asp47, Tyr48, and Asn306 were treated as flexible residues, and the grid size for $_D$-xylose was x = 18, y = 36, z = 40, and grid center was designated at x = 17.169, y = 21.352, z = 33.799. The final conformations produced in this simulation were confirmed using PyMOL software[47]. The calculated free energy of binding (ΔG_{bind}) of the final pose was −4.5 kcal/mol.

Site-directed mutagenesis and activity assay. Site-specific mutations were created with the QuikChange kit (Stratagene), and sequencing was performed to confirm correct incorporation of the mutations. The mutant proteins were purified in the same manner as the wild type. The activities of *Ss*XR were determined by measuring the decrease of absorbance at 340 nm (extinction coefficient of $6.22 \times 10^3\,M^{-1}\,cm^{-1}$). Enzyme reactions for the relative activity of *Ss*XR were performed with a reaction sample of 0.5 ml total volume at 303 K. The reaction sample contained 100 mM Tris-HCl, pH 8.0, 100 mM $_D$-xylose, and 200 μM NADPH. The reactions were initiated by the addition of enzyme to a final concentration of 0.5 μM. The *Ss*XR activity assay was performed in duplicate reaction.

Phylogenetic tree analysis of reported XRs. Annotated XR enzymes are searched by protein search in National Center for Biotechnology Information (NCBI) and UNIPROT server. Multiple sequence alignment was performed using Clustal Omega[48]. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model[49]. The tree with the highest log likelihood (−19893.7669) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 161 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 133 positions in the final dataset. Evolutionary analyses were conducted in MEGA7[50].

References

1. Jonsson, L. J., Alriksson, B. & Nilvebrant, N. O. Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol Biofuels* **6**, 16, https://doi.org/10.1186/1754-6834-6-16 (2013).
2. Sims, R. E., Mabee, W., Saddler, J. N. & Taylor, M. An overview of second generation biofuel technologies. *Bioresour Technol* **101**, 1570–1580, https://doi.org/10.1016/j.biortech.2009.11.046 (2010).
3. Metzger, J. O. & Huttermann, A. Sustainable global energy supply based on lignocellulosic biomass from afforestation of degraded areas. *Naturwissenschaften* **96**, 279–288, https://doi.org/10.1007/s00114-008-0479-4 (2009).
4. Ko, J. K. & Lee, S. M. Advances in cellulosic conversion to fuels: engineering yeasts for cellulosic bioethanol and biodiesel production. *Curr Opin Biotechnol* **50**, 72–80, https://doi.org/10.1016/j.copbio.2017.11.007 (2017).
5. Lee, S. M., Jellison, T. & Alper, H. S. Systematic and evolutionary engineering of a xylose isomerase-based pathway in Saccharomyces cerevisiae for efficient conversion yields. *Biotechnol Biofuels* **7**, 122, https://doi.org/10.1186/s13068-014-0122-x (2014).
6. Van Vleet, J. H. & Jeffries, T. W. Yeast metabolic engineering for hemicellulosic ethanol production. *Curr Opin Biotechnol* **20**, 300–306, https://doi.org/10.1016/j.copbio.2009.06.001 (2009).
7. Hahn-Hagerdal, B., Karhumaa, K., Jeppsson, M. & Gorwa-Grauslund, M. F. Metabolic engineering for pentose utilization in Saccharomyces cerevisiae. *Adv Biochem Eng Biotechnol* **108**, 147–177, https://doi.org/10.1007/10_2007_062 (2007).
8. Sanchez, R. G. *et al.* Improved xylose and arabinose utilization by an industrial recombinant Saccharomyces cerevisiae strain using evolutionary engineering. *Biotechnology for Biofuels* **3**, https://doi.org/10.1186/1754-6834-3-13 (2010).
9. Bera, A. K., Ho, N. W. Y., Khan, A. & Sedlak, M. A genetic overhaul of Saccharomyces cerevisiae 424A(LNH-ST) to improve xylose fermentation. *J Ind Microbiol Biot* **38**, 617–626, https://doi.org/10.1007/s10295-010-0806-6 (2011).
10. Kim, S. R., Park, Y. C., Jin, Y. S. & Seo, J. H. Strain engineering of Saccharomyces cerevisiae for enhanced xylose metabolism. *Biotechnol Adv* **31**, 851–861, https://doi.org/10.1016/j.biotechadv.2013.03.004 (2013).
11. Jin, Y. S., Alper, H., Yang, Y. T. & Stephanopoulos, G. Improvement of xylose uptake and ethanol production in recombinant Saccharomyces cerevisiae through an inverse metabolic engineering approach. *Appl Environ Microb* **71**, 8249–8256, https://doi.org/10.1128/Aem.71.12.8249-8256.2005 (2005).
12. Nogue, V. S. & Karhumaa, K. Xylose fermentation as a challenge for commercialization of lignocellulosic fuels and chemicals. *Biotechnol Lett* **37**, 761–772, https://doi.org/10.1007/s10529-014-1756-2 (2015).
13. Kaparaju, P., Serrano, M., Thomsen, A. B., Kongjan, P. & Angelidaki, I. Bioethanol, biohydrogen and biogas production from wheat straw in a biorefinery concept. *Bioresource Technol* **100**, 2562–2568, https://doi.org/10.1016/j.biortech.2008.11.011 (2009).
14. Mosier, N. *et al.* Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour Technol* **96**, 673–686, https://doi.org/10.1016/j.biortech.2004.06.025 (2005).
15. Haldar, D., Sen, D. & Gayen, K. A review on the production of fermentable sugars from lignocellulosic biomass through conventional and enzymatic route-a comparison. *Int J Green Energy* **13**, 1232–1253, https://doi.org/10.1080/15435075.2016.1181075 (2016).
16. Karhumaa, K., Garcia Sanchez, R., Hahn-Hagerdal, B. & Gorwa-Grauslund, M. F. Comparison of the xylose reductase-xylitol dehydrogenase and the xylose isomerase pathways for xylose fermentation by recombinant Saccharomyces cerevisiae. *Microb Cell Fact* **6**, 5, https://doi.org/10.1186/1475-2859-6-5 (2007).
17. Madhavan, A. *et al.* Xylose isomerase from polycentric fungus Orpinomyces: gene sequencing, cloning, and expression in Saccharomyces cerevisiae for bioconversion of xylose to ethanol. *Appl Microbiol Biot* **82**, 1067–1078, https://doi.org/10.1007/s00253-008-1794-6 (2009).

18. Matsushika, A., Inoue, H., Kodaki, T. & Sawayama, S. Ethanol production from xylose in engineered Saccharomyces cerevisiae strains: current state and perspectives. *Appl Microbiol Biot* **84**, 37–53, https://doi.org/10.1007/s00253-009-2101-x (2009).
19. Tanino, T. *et al*. Construction of a xylose-metabolizing yeast by genome integration of xylose isomerase gene and investigation of the effect of xylitol on fermentation. *Appl Microbiol Biot* **88**, 1215–1221, https://doi.org/10.1007/s00253-010-2870-2 (2010).
20. Ha, S. J., Kim, S. R., Choi, J. H., Park, M. S. & Jin, Y. S. Xylitol does not inhibit xylose fermentation by engineered Saccharomyces cerevisiae expressing xylA as severely as it inhibits xylose isomerase reaction *in vitro*. *Appl Microbiol Biot* **92**, 77–84, https://doi.org/10.1007/s00253-011-3345-9 (2011).
21. Lee, M. *et al*. Metal Dependence of the Xylose Isomerase from *Piromyces* sp. E2 Explored by Activity Profiling and Protein Crystallography. *Biochemistry* **56**, 5991–6005, https://doi.org/10.1021/acs.biochem.7b00777 (2017).
22. Son, H. F., Lee, S. M. & Kim, K. J. Crystal structure and biochemical characterization of xylose isomerase from *Piromyces* sp. E2. *J Microbiol Biotechnol*, https://doi.org/10.4014/jmb.1711.11026 (2018).
23. Chiang, C. & Knight, S. G. Metabolism of d-xylose by moulds. *Nature* **188**, 79–81 (1960).
24. Karhumaa, K., Fromanger, R., Hahn-Hagerdal, B. & Gorwa-Grauslund, M. F. High activity of xylose reductase and xylitol dehydrogenase improves xylose fermentation by recombinant Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* **73**, 1039–1046, https://doi.org/10.1007/s00253-006-0575-3 (2007).
25. Bengtsson, O., Hahn-Hagerdal, B. & Gorwa-Grauslund, M. F. Xylose reductase from Pichia stipitis with altered coenzyme preference improves ethanolic xylose fermentation by recombinant Saccharomyces cerevisiae. *Biotechnol Biofuels* **2**, 9, https://doi.org/10.1186/1754-6834-2-9 (2009).
26. Lee, S. H., Kodaki, T., Park, Y. C. & Seo, J. H. Effects of NADH-preferring xylose reductase expression on ethanol production from xylose in xylose-metabolizing recombinant Saccharomyces cerevisiae. *J Biotechnol* **158**, 184–191, https://doi.org/10.1016/j.jbiotec.2011.06.005 (2012).
27. Runquist, D., Hahn-Hagerdal, B. & Bettiga, M. Increased ethanol productivity in xylose-utilizing Saccharomyces cerevisiae via a randomly mutagenized xylose reductase. *Appl Environ Microbiol* **76**, 7796–7802, https://doi.org/10.1128/AEM.01505-10 (2010).
28. Watanabe, S. *et al*. Ethanol production from xylose by recombinant Saccharomyces cerevisiae expressing protein engineered NADP+ -dependent xylitol dehydrogenase. *J Biotechnol* **130**, 316–319, https://doi.org/10.1016/j.jbiotec.2007.04.019 (2007).
29. Bohren, K. M., Bullock, B., Wermuth, B. & Gabbay, K. H. The aldo-keto reductase superfamily. cDNAs and deduced amino acid sequences of human aldehyde and aldose reductases. *J Biol Chem* **264**, 9547–9551 (1989).
30. Burczynski, M. E., Sridhar, G. R., Palackal, N. T. & Penning, T. M. The reactive oxygen species–and Michael acceptor-inducible human aldo-keto reductase AKR1C1 reduces the alpha,beta-unsaturated aldehyde 4-hydroxy-2-nonenal to 1,4-dihydroxy-2-nonene. *J Biol Chem* **276**, 2890–2897, https://doi.org/10.1074/jbc.M006655200 (2001).
31. Srivastava, S., Chandra, A., Bhatnagar, A., Srivastava, S. K. & Ansari, N. H. Lipid peroxidation product, 4-hydroxynonenal and its conjugate with GSH are excellent substrates of bovine lens aldose reductase. *Biochem Biophys Res Commun* **217**, 741–746, https://doi.org/10.1006/bbrc.1995.2835 (1995).
32. Penning, T. M. *et al*. Human 3alpha-hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *Biochem J* **351**, 67–77 (2000).
33. Komoto, J., Yamada, T., Watanabe, K. & Takusagawa, F. Crystal structure of human prostaglandin F synthase (AKR1C3). *Biochemistry* **43**, 2188–2198, https://doi.org/10.1021/bi036046x (2004).
34. Suzuki-Yamamoto, T. *et al*. cDNA cloning, expression and characterization of human prostaglandin F synthase. *FEBS Lett* **462**, 335–340 (1999).
35. Hyndman, D., Bauman, D. R., Heredia, V. V. & Penning, T. M. The aldo-keto reductase superfamily homepage. *Chem Biol Interact* **143–144**, 621–631 (2003).
36. Kavanagh, K. L., Klimacek, M., Nidetzky, B. & Wilson, D. K. The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from Candida tenuis. *Biochemistry* **41**, 8785–8795 (2002).
37. Penning, T. M. The aldo-keto reductases (AKRs): Overview. *Chem Biol Interact* **234**, 236–246, https://doi.org/10.1016/j.cbi.2014.09.024 (2015).
38. Rizzi, P. E., Bui-Thanh, N.-A. & Xylose, H. D. fermentation by yeasts. *Appl Microbiol Biot* **29**, 148–154 (1988).
39. Suk-Youl Park, S.-C. H. & Kim, Y.-G. The Protein Crystallography Beamlines at the Pohang Light Source II. *Biodesign* **5**, 30–34 (2017).
40. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326 (1997).
41. Matthews, B. W. Solvent content of protein crystals. *J Mol Biol* **33**, 491–497 (1968).
42. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22–25, https://doi.org/10.1107/S0907444909042589 (2010).
43. Kratzer, R., Leitgeb, S., Wilson, D. K. & Nidetzky, B. Probing the substrate binding site of Candida tenuis xylose reductase (AKR2B5) with site-directed mutagenesis. *Biochem J* **393**, 51–58, https://doi.org/10.1042/BJ20050831 (2006).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132, https://doi.org/10.1107/S0907444904019158 (2004).
45. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255, https://doi.org/10.1107/S0907444996012255 (1997).
46. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455–461, https://doi.org/10.1002/jcc.21334 (2010).
47. Alexander, N., Woetzel, N. & Meiler, J. bcl::Cluster: A method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System. *IEEE Int Conf Comput Adv Bio Med Sci* **2011**, 13–18, https://doi.org/10.1109/ICCABS.2011.5729867 (2011).
48. Sievers, F. *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, https://doi.org/10.1038/msb.2011.75 (2011).
49. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275–282 (1992).
50. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870–1874, https://doi.org/10.1093/molbev/msw054 (2016).

Acknowledgements

This work was supported by the New & Renewable Energy Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (20153030091360), and also supported by the Technology Development Program to Solve Climate Changes of the National Research Foundation (NRF) funded by the Ministry of Science and ICT (NRF-2017M1A2A2087631). H-F Son was supported by the NRF-2015-Global PhD Fellowship Program of the Korean Government (2015H1A2A1034233).

Author Contributions

S.-M.L., K.-J.K. designed the project. H.F.S. performed the experiments. H.F.S. and K.-J.K. wrote the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-35703-x.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.