Crystal Structure and Biophysical Analysis of Furfural Detoxifying Aldehyde Reductase from *Clostridium beijerinkii*

Alan F. Scott[a], *Joel Cresser-Brown*[a], *Thomas L. Williams*[a], *Pierre J. Rizkallah*[b], *Yi Jin*[a], *Louis Y.-P. Luk*[a], *Rudolf K. Allemann*[a]#

[a]School of Chemistry, Cardiff University, Cardiff, United Kingdom

[b]Institute of Infection & Immunology, School of Medicine, Cardiff University, Cardiff, U.K.

#Address correspondence to Rudolf K. Allemann: allemannrk@cardiff.ac.uk.

Running Head: Structure-function of a furfural reductase

Keywords:
lignocellulose, biofuel, detoxification, furfural tolerance, aldehyde reductase, isotope effects, heavy enzyme, dynamic coupling

Abstract

Many aldehydes such as furfural are present in high quantities in lignocellulose lysates and are fermentation inhibitors that make biofuel production from this abundant carbon source extremely challenging. Cbei_3974 has recently been identified as an aldo-keto reductase responsible for partial furfural resistance in *Clostridium beijerinkii*. Rational engineering of this enzyme could

enhance the furfural tolerance of this organism thereby improving biofuel yields. We report an extensive characterization of Cbei_3974 and a single crystal X-ray structure of Cbei_3974 in complex with NADPH at a resolution of 1.75 Å. Docking studies identified residues involved in substrate binding and an activity screen revealed the substrate tolerance of the enzyme. Hydride transfer, which is partially rate limiting under physiological conditions, occurs from the *pro-R* hydrogen of NADPH. Enzyme isotope labeling revealed a temperature-independent enzyme isotope effect of unity, indicating that the enzyme does not use dynamic coupling for catalysis and suggests that the active site of the enzyme is optimally configured for catalysis with the substrate tested.

Importance

Herein is reported the crystal structure and biophysical properties of an aldehyde reductase that can detoxify furfural, a common inhibitor of biofuel fermentation found in lignocellulose lysates. The data contained herein will serve as a guide for protein engineers to develop improved enzyme variants that would infer furfural resistance to the microorganisms used in biofuel production and thus lead to enhanced biofuel yields from this sustainable resource.

Introduction

An ideal source of carbohydrates for biofuel fermentation is lignocellulose, an abundant waste product which is available at low cost and does not affect food security (1). Fermentable sugars are most commonly released from lignocellulose using an acid pre-treatment (2). One of the major drawbacks of this method is the release of aldehydes, organic acids and phenols, which severely inhibit growth and limit the final yield of biofuel (3). While it has been observed that

Clostridium sp. has an increased tolerance against the aldehyde inhibitors, furfural and hydroxymethylfurfural, when compared to other organisms, the high levels of inhibitors found in lignocellulose lysates are nevertheless hugely problematic (4–6).

A recent investigation has identified two genes from *Clostridium beijerinkii* that encode enzymes that reduce aldehydes to less toxic alcohols (7, 8). These enzymes showed activity against furfural, hydroxymethyl furfural and benzaldehyde, which are all common fermentation inhibitors (7). Furthermore, the genes encoding these enzymes are up-regulated during furfural stress, suggesting the physiological relevance of these enzymes to protect *C. beijerinkii* (8). One of these enzymes, Cbei_3904, belongs to the short chain dehydrogenase (SDR) family and the other, Cbei_3974, to the aldo-keto reductase (AKR) family (7). It is highly desirable to engineer greater catalytic efficiency into these enzymes to more rapidly eliminate toxic aldehydes thereby enhancing resistance to aldehyde inhibition.

A prerequisite to rational engineering of an enzyme is a thorough understanding of its mechanism. In this report, the furfural transforming AKR, Cbei_3974, is characterized to provide valuable information for protein engineers. The substrate specificity, steady state kinetic parameters and crystal structure of Cbei_3974 have been determined. In addition, the rate limiting step of the reaction was identified and the coupling of dynamic motions to the active site explored.

Results

Substrate Profile.

63 It has previously been suggested that Cbei_3974 may be useful to alleviate the toxicity of
64 furfural during the fermentation of acid treated lignocellulose lysates (7). NADPH dependent
65 activity towards furfural was previously reported for this enzyme but no characterization of the
66 reaction product was shown (7). To confirm that the enzyme indeed generates the less toxic
67 alcohol from the aldehyde, the reaction product from an enzyme-substrate-NADPH incubation
68 was analyzed by gas chormotography-mass spectrometry (GC-MS) in parallel with controls
69 containing no NADPH or no enzyme. After five hours incubation, a new compound can be
70 detected on the GC trace (Figure 1). The retention time and fragmentation pattern was identical
71 to a commercial standard of furfuryl alcohol. This compound was not detected in either of the
72 controls, showing that its formation was enzyme catalyzed and NADPH dependent.
73
74 Cbei_3974 has previously been shown to also exhibit activity with hydroxymethyl furfural,
75 benzaldehyde and butyraldehyde (7). To more fully explore the substrate scope of the enzyme, a
76 selection of aldehydes, ketones and alcohols were chosen. These putative substrates were
77 incubated at 2 mM with the enzyme and an excess of NADPH, NADH or $NADP^+$. The change in
78 cofactor concentration was measured continuously by UV-spectroscopy to give the reaction rate
79 (Table 1). Surprisingly, *L*-glyceraldehyde-3-phosphate only gave negligible activity despite the
80 enzyme sharing 57.7 % identity with *Escherichia coli* YghZ, which converts *L*-glyceraldehyde-
81 3-phosphate to *L*-glycerol-3-phosphate as part of a novel triose phosphate isomerase (TIM)
82 bypass that allows the formation of dihydroxyacetone phosphate under gluconeogenic
83 conditions, when TIM is genetically inactivated (9). Purified YghZ was shown to reduce *L*-
84 glyceraldehyde-3-phosphate to *L*-glycerol-3-phosphate, which can be converted to
85 dihydroxyacetone by *L*-glycerol-3-phosphate dehydrogenase, thus complementing TIM

deficiency. Rather unexpectedly, YghZ is stereospecific for the *L*-enantiomer of the substrate, whereas the TIM substrate is *D*-glyceraldehyde-3-phosphate. It was therefore proposed that a spontaneous reaction may interconvert the two enantiomers (9). On this basis YghZ and sequence similar enzymes, including Cbei_3974, are annotated in the KEGG database (https://www.kegg.jp/) as *L*-glyceraldehyde-3-phosphate reductases (10, 11). Our results demonstrate this annotation is incorrect for Cebi_3974.

Only minimal activity was found for the previously identified substrates, furfural and butyraldehyde (7), while the enzyme had no measurable activity with benzaldehyde at the concentrations tested (2 mM). In contrast, 4-pyridinecarboxaldehyde, which only differs from benzaldehyde by the presence of a nitrogen atom in the aromatic ring, was the most kinetically efficient among all of the substrates examined. Turnover was 17 times faster than for *L*-glyceraldehyde-3-phosphate and 11 times faster than for furfural. Similarly, propionaldehyde did not show activity at 2 mM while the more polar methylglyoxal, a dialdehyde of the same chain length, gave strong activity 6.8 times faster than furfural. No activity could be detected with ketones or alcohols at the concentrations tested.-The enzyme was specific for NADPH with no activity detectable with 0.4 mM NADH. To determine whether the differences between substrates was caused by differences in k_{cat} or K_M, steady state kinetics were measured for the 5 fastest substrates, not including 4-nitrobenzaldehyde, which was not soluble enough to achieve saturation. All the aldehyde substrates resulted in Michaelis constants in the millimolar range but NADPH had higher affinity as indicated by a lower K_M of 32 μM (Table 2). The best substrate was 4-pyridine carboxaldehyde with a k_{cat} of 10 s^{-1} and K_M of 3.87 mM. This was closely followed by methylglyoxal with a similar k_{cat} of 8.52 s^{-1} but a higher K_M of 12.9 mM. Furfural has previously been shown to have an extraordinarily high k_{cat} of $1.4x10^5$ s^{-1} at 40 °C (7). In our

experiments the k_{cat} was measured at 19 °C and was 2.72 s^{-1}. This was lower than expected, even considering the lower temperature, but is more realistic. The K_M value of 34.9 mM measured here is in agreement with the literature value (7).

Stereochemistry of NADPH transfer

The hydride transfer in aldehyde reductases occurs either from the *pro-R* or *pro-S* hydrogen on C4 of the nicotinamide ring of NADPH. Typically, short-chain reductases (SDR) transfer the *pro-S* hydrogen, while aldo-keto reductases (AKR) transfer the *pro-R* hydrogen (12). To determine the stereospecificity of Cbei_3974, the enzyme was incubated with NADPH or (4*R*)-[4-^{2}H]-NADPD with an excess of substrate 4-pyridinecarboxaldehyde. The reaction products were analyzed to determine whether the deuterium had been incorporated into the alcohol product or remained on the nicotinamide cofactor. The alcohol reaction product was extracted with chloroform and analyzed by GC-MS. The incubation with (4*R*)-[4-^{2}H]-NADPD gave a product which was 1 atomic mass unit higher than the incubation with NADPH, consistent with the incorporation of deuterium. In a duplicate reaction, the nucleotide product, $NADP^+$, was purified by ion exchange chromatography, freeze-dried and dissolved in D_2O. NMR analysis of $NADP^+$ from the reaction compared with commercial $NADP^+$ showed identical spectra, confirming that the deuterium at the *pro-R* position had been transferred from (4*R*)-[4-^{2}H]-NADPD (Figure 2). Cbei_3974 therefore transfers the *pro-R* hydrogen from NADPH in accordance with other members of the AKR superfamily.

Substrate Kinetic Isotope Effect (KIE).

To ascertain if hydride transfer is the rate-limiting step, the substrate KIE for NADPH vs (4*R*)-[4-^{2}H]-NADPD was determined for a range of substrates. A KIE on k_{cat} of 2.13 - 2.58 was observed across the 5 substrates tested (Figure 2) suggesting that the catalytic step is at least partially rate limiting for all substrates tested.

Heavy Enzyme Kinetic Isotope Effect.

The effect of protein dynamics on catalysis was investigated by heavy enzyme production, where the non-exchangeable carbon and nitrogen atoms were replaced with their heavy counterparts (^{13}C, ^{15}N) to slow protein motions without affecting the electrostatics. A reactivity difference between the "heavy" (labeled) and "light" (natural abundance) enzymes indicates that protein motions impact on the catalysis (13). As the substrate KIE measurements indicated that hydride transfer is partially rate limiting, steady-state measurements were used to determine any effect that slower protein motions in heavy enzyme may have on the catalytic step.

Heavy enzyme (^{15}N, ^{13}C) was produced in M9 media with labeled feed-stocks and purified to homogeneity (Figure S3). The incorporation of the heavy isotopes was confirmed by mass spectrometry on the purified enzyme, which revealed a 5.5% mass increase (Figure S4). To determine if the protein was correctly folded, the CD-spectrum and melting temperature were recorded and compared for both the "heavy" and "light" enzymes (Figure S5). Both enzymes gave identical spectra and largely identical melting temperatures of 62.4 °C ± 0.1 and 63.8 °C ± 0.2, respectively, indicating that isotopic labeling does not significantly alter protein folding.

Steady-state kinetics were used to determine k_{cat} for "heavy" and "light" enzyme with a range of substrates at 19 °C (3). All substrates gave an enzyme KIE of near unity, implying that there were no mass dependent effects and that dynamic coupling was minimal (Figure 3). Although

some authors have proposed that enzymes use "promoting motions" to drive catalysis (14–17), this result is consistent with a growing body of literature that shows that dynamic effects only become significant outside physiological conditions and only when poorly tolerated substrates are utilized that necessitate rearrangement of the active site (18–20).The enzyme does not therefore use dynamic motions as a part of its catalytic mechanism.

A recent study on the thermophilic alcohol dehydrogenase BsADH showed that significant heavy enzyme KIEs only manifest below its physiological temperature (40 °C) and only with poor substrates (19). The temperature dependency of heavy enzyme KIEs has been suggested to be an indicator of whether an enzyme is optimized for utilization of a particular substrate (19). The temperature dependence of the KIE for the Cbei_3974 catalyzed reduction of 3-pyridine carboxaldehyde was constant over the temperature range from 11 to 44 °C (Figure 4) suggesting that the active site architecture of the enzyme is optimized for this substrate.

Single crystal X-ray Structure.

The protein was co-crystallized with NADPH and the structure solved by molecular replacement using PDB entry 5T79, which is the crystal structure for STM2406, an AKR from *Salmonella typhimurium* of unknown physiological function but with a similar substrate profile to Cbei_3974 (21). The two proteins have 60.91% sequence identity and a root mean square deviation 0.89. The structure was refined at 1.75 Å to R_{factor} 16.5% (R_{free} 19.3%). The structure consists of alternating α-helices and β-strands forming an 8 stranded TIM barrel with some extra helices (Figure 5).

This motif is conserved across the AKR superfamily (23). Both this structure and STM2406 have an unusual N-terminus consisting of a long loop and a β-hairpin. Most AKR structures, including the structure of *Coptotermes gestroi* AKR1 (another AKR known to reduce furfural),

177 have a shorter N-terminal tail consisting only of the β-hairpin or, in the case of the AKR7 family,
178 have no N-terminal tail (21, 24, 25). The function of this extra sequence is unclear. Conversely,
179 the C-terminus is truncated and is lacking a loop that is present in many AKRs, leaving the active
180 site exposed to solvent (25). AKRs, which omit this loop, have low catalytic efficiency
181 consistent with the measured millimolar Michaelis constants (21, 26, 27). Deletion of the C-
182 terminal loop from human aldose reductase AKR1B1 (27), *Bacillus subtilis* YhdN, YvgN and
183 *Psedomonas aeruginosa* PA1127 (21) resulted in dramatic loss of catalytic efficiency.
184 NADPH sits in a mostly open cleft with a hydrophobic center and polar residues at the ends
185 where the adenine base and nicotinamide ring bind. The adenine base is held in place by
186 hydrogen bonds to Glu-307 and Asn-308. The nucleotide 2'-phosphate that distinguishes
187 NADPH from NADH, is hydrogen bonded to Gln-304 and Ser-300. The diphosphate makes
188 hydrogen bond contact with the backbone oxygen of Leu-225.
189 There is an area of missing electron density between residues Ile-238 to Leu-256. In
190 *Coptotermes gestroi* AKR1 and human aldose reductase, this region forms a mobile loop that
191 would strap the cofactor in place across the diphosphate bridge (24, 28). The lack of density in
192 Cbei_3974 indicates that the region is disordered and suggests that the loop is not trapping the
193 cofactor.
194 The canonical mechanism of AKRs involves hydride transfer from NADPH to the carbonyl
195 acceptor (12). This is followed by protonation from an active site tyrosine as part of a proton
196 relay from histidine and bulk water (12). Neighboring aspartate and lysine residues lower the pK_a
197 of tyrosine to enable it to function as an acid (12). In the close homologue STM2406, the
198 catalytic tetrad consists of Tyr-66, Asp-61, Lys-97 and His-138 (21). All these residues are
199 conserved in Cbei_3974 (identical numbering). It was not possible to obtain crystal structures of

protein-product complexes due to the low affinity of the ligands. Therefore, docking was used to predict the possible binding of substrates. Autodock Vina (29) was used to dock furfural and the best substrate, 4-pyridine carboxaldehyde, into the active site. The best pose was selected on the basis of proximity to NADPH and the proposed catalytic residues. These poses are illustrated in Figure 6. Both substrates are orientated towards the *pro-R* hydrogen of NADPH, consistent with the experimentally determined stereochemistry. The carbonyl oxygen of 4-pyridine carboxaldehyde is within hydrogen bonding distance of the exocyclic amide of NADPH and makes hydrophobic contacts with residues Asn-65, Trp-33 and Tyr-100. These residues are conserved in STM2406 and have been shown to be important for binding in that enzyme (21). Furfural docked into the active site in a similar location but with a different orientation, possibly due to its smaller size. The active site has a lot of polar residues, which may explain why the more hydrophobic aldehydes such as benzaldehyde are less favored. Asn-65 and Tyr-100 contribute to the polar surface of the active site, and therefore alteration of these residues to more hydrophobic ones may help improve activity for hydrophobic substrates. In STM2406, which has a very similar active site, the variant Asn-65-Met (both enzymes have the same residue numbering) gave 341% increase in activity towards 3-pyridinecarboxaldehyde, compared with wild type and a 2-fold decrease in K_M (21). Alterations of Tyr-100 to aspartate, leucine, isoleucine or valine mostly resulted in insoluble proteins, while Tyr-100-Ala showed decreased activity but this may have been due to a loss of steric bulk by replacing a phenyl group with a hydrogen atom (21).

Discussion

Cbei_3974, an enzyme that could putatively address the problem of aldehyde toxicity in biofuel fermentation from lignocellulose (7), has been extensively characterized and its crystal structure solved. Though annotated as an *L*-glyceraldehyde-3-phosphate reductase, it shows only low activity towards this substrate. Instead, it preferentially catalyzes reactions with 4-pyridine carboxaldehyde and methylglyoxal, of which only the latter is likely to be naturally occurring inside the cell. Methylglyoxal is a toxic product formed from dihydroxy-acetone-phosphate by methylglyoxal synthase to release phosphate (30). Detoxification of methylglyoxal by *Clostridium* results in the formation of 1,2-propanediol in a pathway that requires aldo-keto reductase activity (31). The physiological relevance of methylglyoxal reductase activity of Cbei_3974 is however questionable, given the low affinity of the substrate with a K_M in the region of 12 mM. The low affinity may be part of a mechanism to conserve NADPH, as depletion of NADPH can be just as lethal as aldehyde accumulation (32) but this is unlikely given the LD_{50} of methylglyoxal is likely to be significantly less than the K_M, rendering the enzyme useless for detoxification. It is also possible that the enzyme requires an interaction partner or a post-translational modification for activity or that the experimental conditions were not optimal. An examination of the *C. beijerinkii* genome on the KEGG database (10, 11) shows that the *cbei_3974* gene is not part of a biosynthetic gene cluster but it is adjacent to a putative MerR transcription regulator. These typically respond to environmental stimuli to up-regulate stress response proteins (33). In conclusion, the natural substrate of Cbei_3974 remains unclear but it is likely to be involved in a stress response.

Studies with (4*R*)-[4-^{2}H] NADPD demonstrated that hydride transfer occurs from the *pro-R* hydrogen of NADPH and is partially rate limiting. Isotopically labeled heavy enzyme (^{13}C, ^{15}N)

gave identical k_{cat} constants to those obtained with the natural abundance enzyme, indicating that the slower protein motions in the heavy enzyme did not impact the catalytic step. This shows that the enzyme does not use "promoting motions" to drive the chemical transformation.

The enzyme kinetic isotope effect was independent of temperature suggesting that the enzyme's active site is optimally configured for the use of the tested substrates; the physiological substrate is therefore likely to be structurally similar (19).

The crystal structure of Cbei_3974 revealed a typical AKR structure based around a TIM barrel fold and is essentially the same as that of STM2406 (21). Docking of 4-pyridinecarboxaldehyde and furfural revealed residues that may be involved in substrate binding. Although no enzyme variants were generated in this study, these residues represent targets for future work to generate an improved enzyme for more efficient detoxification of furfural.

Materials and Methods

Material.

A *pET-14b* vector harboring a codon optimized gene encoding Cbei_3974 was purchased from GenScript (sequence in Figure S6). This also encodes a 6xHis tag and thrombin cleavage site upstream of Cbei_3974.

NADPH was obtained from Fisher or Apollo Scientific. (4*R*)-[4-(2)H]NADPD was prepared from $NADP^+$ (Melford) and d_8 isopropanol (Acros) according to the published protocol (34).

^{15}N-NH_4Cl_2 and ^{13}C-glucose were obtained from Goss Scientific, Cheshire, UK.

Furfural was obtained from VWR; methylglyoxal from Apollo Scientific; 3-pyridine carboxaldehyde from Acros. All other chemicals were obtained from Sigma-Aldrich.

Crystallography.

Cbei_3974 was overproduced in BL21(DE3) cells and purified by Ni-affinity chromatography as previously described (7). The protein was dialyzed against 10 mM HEPES-NaOH, pH 7.5, 300 mM NaCl and concentrated to 10 mg/mL. Crystallization trials were performed using the sitting drop vapor diffusion method by mixing 0.5 µL protein stock solution and 0.5 µL of a seed stock with 0.5 µL reservoir solution. The seed stock was obtained from microcrystals grown in 100 mM MOPS, pH 7.3, 13% PEG 8000, 750 mM NH_4Cl with 2 mM NADPH and 2 mM 4-pyridine methanol. Diffracting crystals were obtained from 90 mM MOPS, pH 7.6, 271 mM NH_4Cl, 2.7% PEG 8000, with 10 mM NADPH and 2 mM 4-pyridine methanol added to the protein prior to crystallization. The crystals were transferred to cryoprotectant (90 mM MOPS, pH 7.6, 271 mM NH_4Cl, 2.7% PEG 8000, 25% ethylene glycol) and flash-frozen with liquid nitrogen.

The X-ray diffraction data was collected at 100K at Diamond Light Source (Oxfordshire, U.K.) on Beamline I04-1 and integrated with XDS(35) in the xia2 package (36). The data were scaled, reduced and analyzed with AIMLESS and TRUNCATE in the CCP4i (37). The structure was solved by molecular replacement with PHASER (38) using coordinates from PDB 5T79 as a searching model (21). The structure model was adjusted with COOT (39) and refined with REFMAC5 (40). Graphical representations were prepared in Chimera(41), PyMOL (The PyMOL Molecular Graphics System, Version 1.8.X Schrödinger, LLC), YASARA View(42) and LigPlot+ (22).

Accession Numbers.

The X-ray structure solved in this study was deposited into the Protein Data Bank (http://www.rcsb.org/pdb/), with accession code 6HG6.

Molecular Docking.

Ligand structures were downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/) (43) as SDF files and converted into mol2 format using Chimera (41). Ligand and protein were converted into PDBQT format using AudoDockTools1.5.6 and docked using AutoDock Vina (29). Graphical representations were prepared in PYMOL (The PyMOL Molecular Graphics System, Version 1.8.X Schrödinger, LLC), Chimera (41), and LigPlot+ (22).

Heavy and Natural Abundance Enzyme Production

Arctic Express DE3 cells harboring *pET-14b_Cbei3974* were grown in 20 mL M9 minimal medium containing 100 μg/mL carbenicillin overnight at 37 °C with shaking. This was diluted 1:50 into 0.5 L M9 media containing either natural abundance isotopes or ^{13}C-glucose and ^{15}N-NH_4Cl_2 for heavy enzyme production. Cultures were grown to OD_{600} 1.0 at 37 °C, 220 rpm. Cultures were cooled to 12 °C and 0.4 mM IPTG was added to induce gene expression. Cells were harvested at 4,000 rpm after 16 h growth at 12 °C, 200 rpm. The protein was purified as previously described (7).

Substrate Screen

Cbei_3974 (0.585 μM), 0.4 mM NADPH, NADH or $NADP^+$ and 2 mM putative substrate was mixed in 20 mM K_iPO_4 pH 7.0. The subsequent change in NADPH concentration was monitored at 340 nm (e_{340} = 6220 M^{-1} cm^{-1}) for 1 minute, using a Shimadzu UV-2401PC spectrophotometer in 5 mm quartz cuvettes, to give the reaction rate.

315 **Enzyme Kinetics.**

316 Kinetic parameters were determined with 117 nM Cbei_3974 with one of the two substrates;
317 one held at saturating level while the other was varied across a concentration range of 0-75 mM
318 (aldehydes) or 0-0.2 mM (NADPH). Rates were measured as above and the data fitted to the
319 Michaelis-Menten equation using GraphPad Prism version 7.00 for Windows, GraphPad
320 Software, La Jolla California USA, www.graphpad.com. Each datapoint is the average of 3
321 repeats. For heavy enzyme kinetics, a minimum of two datasets of triplicates were collected.
322
323 **Mass Spectrometry.**

324 Liquid chromatography mass spectrometry (LC-MS) was performed on a Waters Synapt G2-Si
325 quadrupole time of flight mass spectrometer coupled to a Waters Acquity H-Class UPLC system.
326 The column was an Acquity UPLC protein BEH C4 (300 Å 1.7 µm x 2.1 mm x 100 mm)
327 operated in reverse phase and held at 60 °C. The gradient employed was 95% A to 35% A over
328 50 minutes, where A is H_2O with 0.1% HCO_2H and B is acetonitrile (ACN) with 0.1% HCO_2H.
329 Data was collected in positive ionisation mode and analyzed using the *Waters MassLynx*
330 software version 4.1. Deconvolution of protein charged states was obtained using the maximum
331 entropy 1 processing software.
332
333 **GC-MS analysis of reaction products.**

334 Mixtures containing 5 µM Cbei_3974, 4.6 mM NADPH or (4*R*)-[4-^{2}H]-NADPD and 28 mM
335 aldehyde in a final volume of 40 µl 20 mM potassium phosphate pH 7.0 were incubated at 40 °C
336 for 1 hour. Aliquots (4 µL) were removed at time zero and at 5 or 7 hours and quenched with 1
337 ml of chloroform, which also served to extract the organic molecules. A 5 µl aliquot of the

organic layer was injected onto a PerkinElmer Clarus® 680 Gas Chromatograph. The initial temperature was 40 °C, held for 1 minute and elution was with a gradient rising to 150 °C at 15 deg/min, holding at 150 °C for 1 minute. After a 3 minute solvent delay, mass spectra were collected over the range 45 – 200 E+. Controls omitting either enzyme or NADPH were also performed and a standard of furfuryl alcohol was run on the GC-MS.

^{1}H NMR of Nicotinamide Cofactors.

Cbei_3974 (1 mM), 2 mM (4*R*)-[4-^{2}H]-NADPD and 25 mM 4-pyridine carboxaldehyde were incubated at 37 °C for 3 hours. $NADP^+$ was purified from the reaction on a SAX-10 column using published methodology (34), freeze dried and re-dissolved in D_2O. Commercial standards of NADPH and $NADP^+$ were also dissolved in D_2O. NMR spectra of standards and reaction product were collected on Bruker 400.

Circular Dichroism.

Circular dichroism (CD) measurements were performed on an Applied Photophysics Chirascan spectrometer using 7 μM protein in 20 mM K_iPO_4 20 % glycerol. Spectra was recorded over a temperature range of 5 - 85 °C from 200 nm to 400 nm. The melting temperature was calculated by fitting the data in SigmaPlot (Systat Software, San Jose, CA).

Supporting Information.

A PDF file with 6 pages containing: sequence of synthetic *cbei_3974* gene, SDS-PAGE showing purity of proteins, mass spectrometry and circular dichroism of natural abundance and heavy enzyme, and X-ray crystallography data collection and refinement statistics.

Acknowledgements.

We thank Diamond Light Source for access to beamline I04-1 (beamtime code mx1484-11). Helpful discussions with Dr. Enas Behiry are gratefully acknowledged. This work was supported by the UK's Biotechnology and Biological Sciences Research Council through grants BB/J005266/1 and BB/L020394/1.

References

1. Isikgor FH, Becer CR. 2015. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. Polym Chem 6:4497–4559.

2. Kumar P, Barrett DM, Delwiche MJ, Stroeve P. 2009. Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. Ind Eng Chem Res 48:3713–3729.

3. Liu ZL, Blaschek HP. 2010. Biomass Conversion Inhibitors andIn Situ Detoxification, p. 233–259. *In* Biomass to Biofuels. Blackwell Publishing Ltd., Oxford, UK.

4. Ezeji T, Qureshi N, Blaschek HP. 2007. Butanol production from agricultural residues: Impact of degradation products onClostridium beijerinckii growth and butanol fermentation. Biotechnol Bioeng 97:1460–1469.

5. Zhang Y, Han B, Ezeji TC. 2012. Biotransformation of furfural and 5-hydroxymethyl furfural (HMF) by Clostridium acetobutylicum ATCC 824 during butanol fermentation. N Biotechnol 29:345–351.

6. Ujor V, Agu CV, Gopalan V, Ezeji TC. 2014. Glycerol supplementation of the growth

382 medium enhances in situ detoxification of furfural by Clostridium beijerinckii during
383 butanol fermentation. Appl Microbiol Biotechnol 98:6511–6521.

384 7. Zhang Y, Ujor V, Wick M, Ezeji TC. 2015. Identification, purification and
385 characterization of furfural transforming enzymes from Clostridium beijerinckii NCIMB
386 8052. Anaerobe 33:124–131.

387 8. Zhang Y, Ezeji TC. 2013. Transcriptional analysis of Clostridium beijerinckii NCIMB
388 8052 to elucidate role of furfural stress during acetone butanol ethanol fermentation.
389 Biotechnol Biofuels 6:66.

390 9. Desai KK, Miller BG. 2008. A Metabolic Bypass of the Triosephosphate Isomerase
391 Reaction. Biochemistry 47:7983–7985.

392 10. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a
393 reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462.

394 11. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new
395 perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–
396 D361.

397 12. Penning TM. 2015. The aldo-keto reductases (AKRs): Overview. Chem Biol Interact
398 234:236–46.

399 13. Kholodar SA, Ghosh AK, Kohen A. 2017. Measurement of Enzyme Isotope Effects.
400 Methods Enzymol 596:43–83.

401 14. Silva RG, Murkin AS, Schramm VL. 2011. Femtosecond dynamics coupled to chemical
402 barrier crossing in a Born-Oppenheimer enzyme. Proc Natl Acadamy Sci U S A

108:18661–18665.

15. Kipp DR, Silva RG, Schramm VL. 2011. Mass-dependent bond vibrational dynamics influence catalysis by HIV-1 protease. J Am Chem Soc 133:19358–19361.

16. Pudney CR, Guerriero A, Baxter NJ, Johannissen LO, Waltho JP, Hay S, Scrutton NS. 2013. Fast Protein Motions Are Coupled to Enzyme H-Transfer Reactions. J Am Chem Soc 135:2512–2517.

17. Antoniou D, Ge X, Schramm VL, Schwartz SD. 2012. Mass Modulation of Protein Dynamics Associated with Barrier Crossing in Purine Nucleoside Phosphorylase. J Phys Chem Lett 3:3538–3544.

18. Luk LYP, Loveridge EJ, Allemann RK. 2014. Different Dynamical Effects in Mesophilic and Hyperthermophilic Dihydrofolate Reductases. J Am Chem Soc 136:6862.

19. Behiry E, Ruiz-Pernia JJ, Luk L, Tunon I, Moliner V, Allemann RK. 2018. Isotope substitution of promiscuous alcohol dehydrogenase reveals origin of substrate preference in transition state. Angew Chem Int Ed 57:3128–3131.

20. Ruiz-Pernía JJ, Behiry E, Luk LYP, Loveridge EJ, Tuñón I, Moliner V, Allemann RK. 2016. Minimization of dynamic effects in the evolution of dihydrofolate reductase. Chem Sci 7:3248–3255.

21. Kim T, Flick R, Brunzelle J, Singer A, Evdokimova E, Brown G, Joo JC, Minasov GA, Anderson WF, Mahadevan R, Savchenko A, Yakunin AF. 2017. Novel Aldo-Keto Reductases for the Biocatalytic Conversion of 3-Hydroxybutanal to 1,3-Butanediol: Structural and Biochemical Studies. Appl Environ Microbiol 83:e03172-16.

22. Laskowski RA, Swindells MB. 2011. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. J Chem Inf Model 51:2778–2786.

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. Tramontina R, Franco Cairo JPL, Liberato M V., Mandelli F, Sousa A, Santos S, Rabelo SC, Campos B, Ienczak J, Ruller R, Damásio ARL, Squina FM. 2017. The Coptotermes gestroi aldo–keto reductase: a multipurpose enzyme for biorefinery applications. Biotechnol Biofuels 10:4.

25. Barski OA, Tipparaju SM, Bhatnagar A. 2008. The aldo-keto reductase superfamily and its role in drug metabolism and detoxification. Drug Metab Rev 40:553–624.

26. Ehrensberger AH, Wilson DK. 2004. Structural and Catalytic Diversity in the Two Family 11 Aldo-keto Reductases. J Mol Biol 337:661–673.

27. Bohren KM, Grimshaw CE, Gabbay KH. 1992. Catalytic effectiveness of human aldose reductase. Critical role of C-terminal domain. J Biol Chem 267:20965–70.

28. Wilson DK, Bohren KM, Gabbay KH, Quiocho FA. 1992. An unlikely sugar substrate site in the 1.65 A structure of the human aldose reductase holoenzyme implicated in diabetic complications. Science 257:81–4.

29. Trott O, Olson AJ. 2009. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:NA-NA.

30. Huang K, Rudolph FB, Bennett GN. 1999. Characterization of methylglyoxal synthase

445 from Clostridium acetobutylicum ATCC 824 and its use in the formation of 1, 2-
446 propanediol. Appl Environ Microbiol 65:3244–7.

447 31. Liyanage H, Kashket S, Young M, Kashket ER. 2001. Clostridium beijerinckii and
448 Clostridium difficile Detoxify Methylglyoxal by a Novel Mechanism Involving Glycerol
449 Dehydrogenase. Appl Environ Microbiol 67:2004–2010.

450 32. Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO. 2009.
451 Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-
452 resistant ethanologenic Escherichia coli. Appl Environ Microbiol 75:4315–23.

453 33. Brown NL, Stoyanov J V, Kidd SP, Hobman JL. 2003. The MerR family of
454 transcriptional regulators. FEMS Microbiol Rev 27:145–63.

455 34. Loveridge EJ, Allemann RK. 2011. Effect of pH on Hydride Transfer by Escherichia coli
456 Dihydrofolate Reductase. ChemBioChem 12:1258–1262.

457 35. Kabsch W. 2010. XDS. Acta Crystallogr D Biol Crystallogr 66:125–32.

458 36. Winter G, IUCr. 2010. *xia2* : an expert system for macromolecular crystallography data
459 reduction. J Appl Crystallogr 43:186–190.

460 37. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM,
461 Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS,
462 Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the *CCP* 4
463 suite and current developments. Acta Crystallogr Sect D Biol Crystallogr 67:235–242.

464 38. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007.
465 Phaser crystallographic software. J Appl Crystallogr 40:658–674.

466 39. Emsley P, Cowtan K. 2004. *Coot* : model-building tools for molecular graphics. Acta
467 Crystallogr Sect D Biol Crystallogr 60:2126–2132.

468 40. Murshudov GN, Vagin AA, Dodson EJ, IUCr. 1997. Refinement of Macromolecular
469 Structures by the Maximum-Likelihood Method. Acta Crystallogr Sect D Biol Crystallogr
470 53:240–255.

471 41. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE.
472 2004. UCSF Chimera?A visualization system for exploratory research and analysis. J
473 Comput Chem 25:1605–1612.

474 42. Krieger E, Vriend G. 2014. YASARA View—molecular graphics for all devices—from
475 smartphones to workstations. Bioinformatics 30:2981–2982.

476 43. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S,
477 Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. 2016. PubChem Substance and
478 Compound databases. Nucleic Acids Res 44:D1202-13.

TABLES:

Table 1. Substrate Screen

Activity of Cbei_3974 with putative substrates (2 mM) and 0.4 mM NADPH. Errors indicate standard deviation of three repeats. n.d. = no detectable activity

Substrate	Specific Activity μmols/min/mg
furfural	416.86 ± 4.81
benzaldehyde	n.d.
4-pyrridine carboxaldehyde	4,652.70 ± 671.05
3-pyrridine carboxaldehyde	1,519.52 ± 61.48
4-nitrobenzaldehyde	4,390.44 ± 61.61
isatin	n.d.
methylglyoxal	2,852.59 ± 335.72
formaldehyde	n.d.
acetaldehyde	n.d.
propionaldehyde	n.d.
butyraldehyde	307.37 ± 23.65

valeraldehyde	n.d.
furfural alcohol	n.d.
benzyl alcohol	n.d.
4-pyridine methanol	n.d.
methanol	n.d.
ethanol	n.d.
isopropyl alcohol	n.d.
butanol	n.d.
2,4-pentanedione	n.d.
4-acetylpyridine	n.d.
2-butanone	n.d.
L-arabinose	n.d.
D-glucose	n.d.
β-lactose	n.d.
L-glyceraldehyde-3-phosphate	269.71 ± 34.05

484

485 Table 2. Kinetic measurements for Cbei_3974.

486 All measurements at 19 °C. Errors show the standard error from fitting the data to the

487 Michaels-Menten equation in GraphPad Prism.

488

Substrate	k_{cat} /s^{-1}	K_M /mM	k_{cat}/K_M s^{-1} M^{-1}
furfural	2.72 ± 0.27	34.9 ± 7.6	78
butyraldehyde	5.16 ±1.04	40.79 ± 4.5	127
3-pyridine carboxaldehyde	4.97 ± 1.38	15.7 ± 2.6	317
methylglyoxal	8.52 ± 0.26	12.9 ± 1.5	660
4-pyridine carboxaldehyde	10.60 ± 0.2	3.87 ± 0.34	2,739
NADPH (with 4-pyridine carboxaldehyde)	10.0 ± 1.02	0.032 ± 0.02	312,500
NADPH (with furfural)	2.59 ± 0.145	0.015 ± 0.003	172,666

FIGURES:

Figure 1. GC-MS analysis of the Cbei_3974 reaction product

(a) mixture of enzyme, NADPH and furfural at time zero, (b) incubation of furfural and NADPH without enzyme after 5 hours, (c) incubation of furfural and enzyme without NADPH for 5 hours, (d) incubation of furfural, NADPH and enzyme after 5 hours, (e) as d but doped with furfuryl alcohol, (f) standard of furfuryl alcohol, (g) mass spectrum of reaction product from incubation of furfural, NADPH and enzyme after 5 hours, (h) mass spectrum of furfuryl alcohol standard. The full length GC traces can be found in figure S1.

Figure 2. MS analysis of the alcohol product from an incubation of 4-pyridine carboxaldehyde, enzyme and either (A) NADPH or (B) (4*R*)-[4-(^{2})H]NADPD. The corresponding GC traces can be seen in S2. (C) NMR spectra showing the purified nucleotide reaction product obtained from incubations of Cbei_3974 with (4*R*)-[4-(^{2})H]NADPD and 4-pyridine carboxaldehyde (red) against a standard of $NADP^+$ (blue).

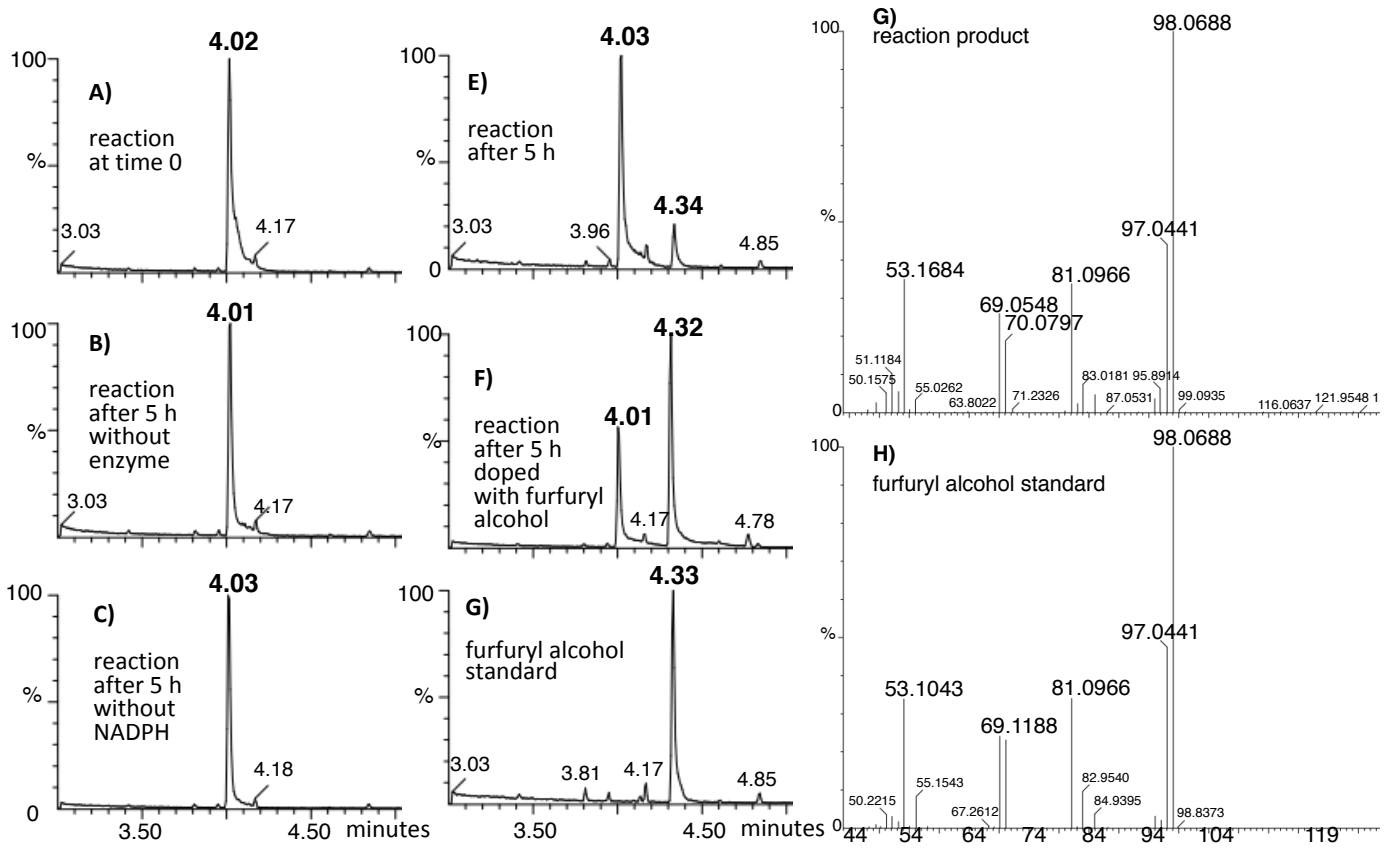
Figure 3. Kinetic isotope effects. Grey circles: k_{cat} values for natural abundance enzyme with NADPH; blue circles: enzyme KIEs; red circles: substrate KIE (NADPH). All measurements were made at 19 °C. Errors are standard deviations of three repeats.

Figure 4. The temperature dependency of the enzyme kinetic isotope effect (k_{cat} light enzyme / k_{cat} heavy enzyme) with 3-pyridine carboxaldehyde as a substrate. Error bars show the standard deviation of three repeats.

Figure 5. Crystal structure of Cbei_3974. **(A)** Cartoon representation showing the TIM barrel fold. NADPH can be seen in gold, with the nicotinamide ring in the central cavity. **(B)** Surface representation showing the exposed nature of the active site. NADPH shown in gold. **(C)** Residues involved in binding NADPH. Red semicircles identify hydrophobic interactions, residues involved in hydrogen bonds are shown in blue. The green numbers indicate hydrogen bond distance. Figure prepared with LigPlot+ (22).

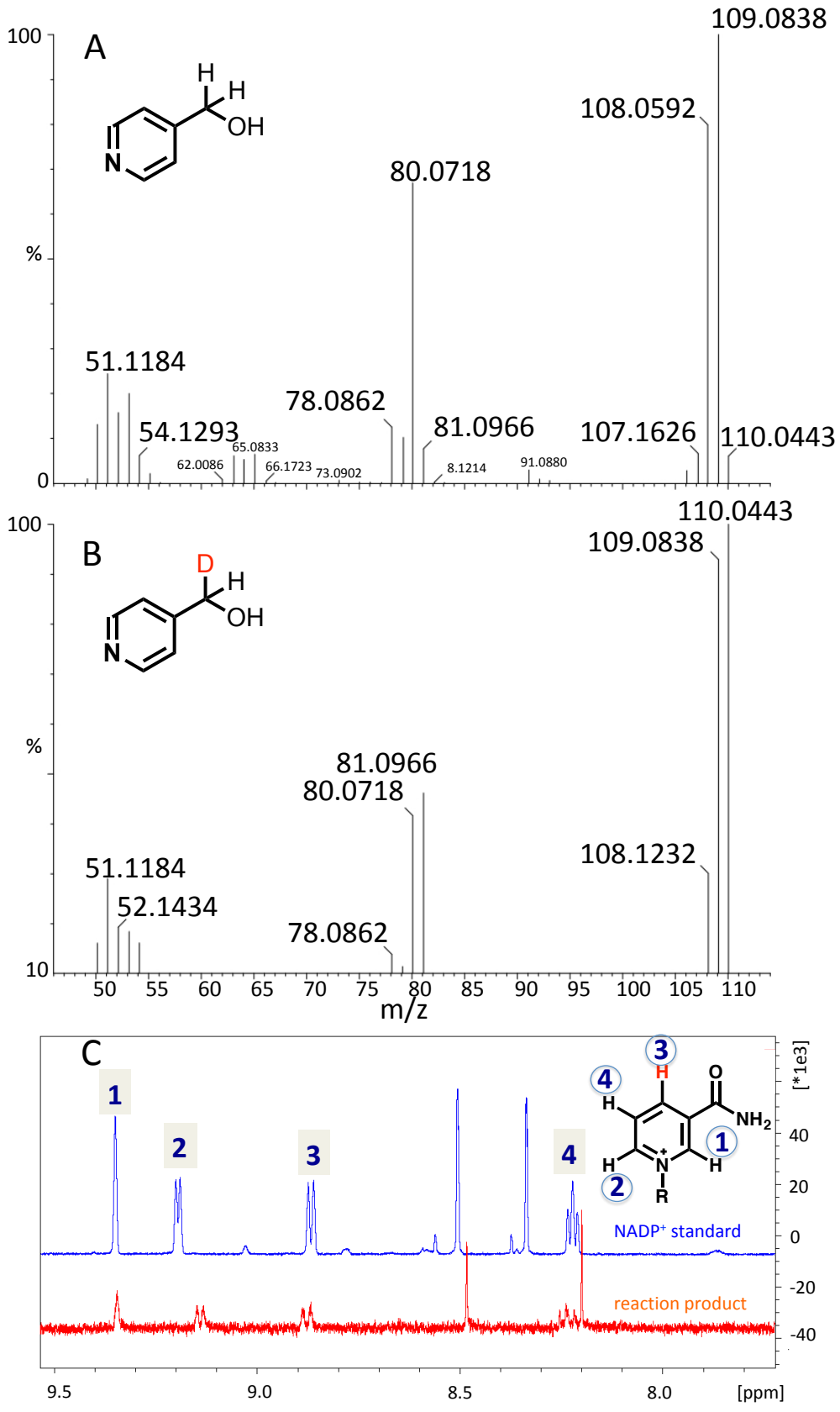
Figure 6. Docking of 4-pyridine carboxaldehyde **(A and B)** and furfural **(C and D)** into the active site. Substrates are shown in orange. **(A and C)** show a hydrophobic surface rendering; **(B and D)** identifies residues involved in binding (grey) and catalysis (green).

522

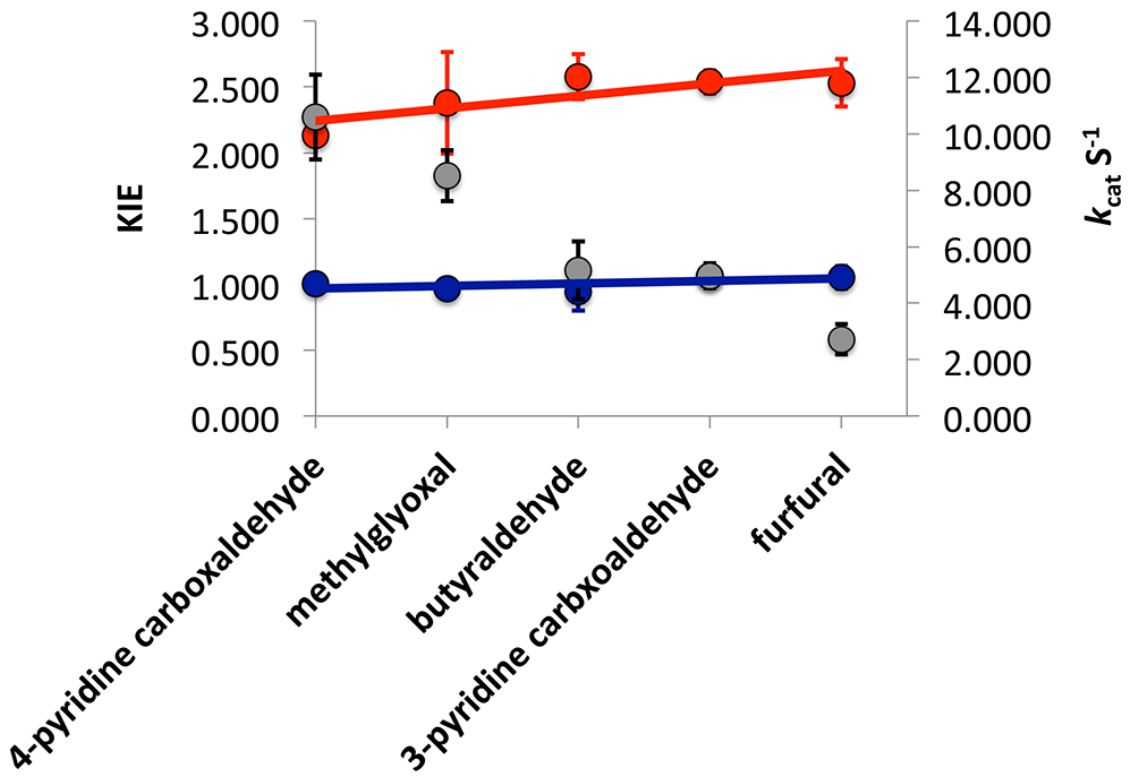


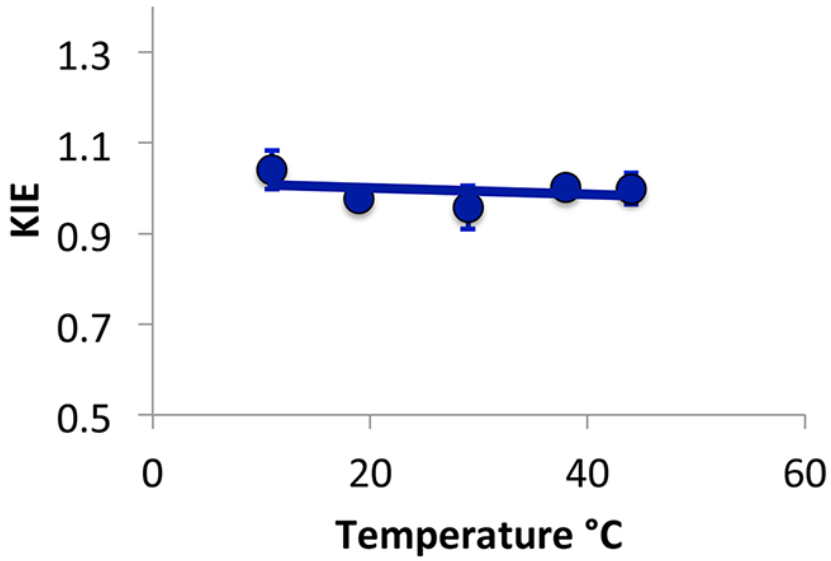
523
524

525


A
100
%
0
109.0838
108.0592
80.0718
51.1184
78.0862
54.1293
81.0966
107.1626
110.0443
65.0833
62.0086
66.1723
73.0902
8.1214
91.0880
H
H
OH
N
B
100
%
10
D
H
OH
N
110.0443
109.0838
81.0966
80.0718
108.1232
51.1184
52.1434
78.0862
50 55 60 65 70 75 80 85 90 95 100 105 110
m/z
C
1
2
3
4
3
H
4
H
O
NH2
1
H
H
2
N
R
[*1e3]
40
20
0
-20
-40
NADP+ standard
reaction product
9.5
9.0
8.5
8.0
[ppm]

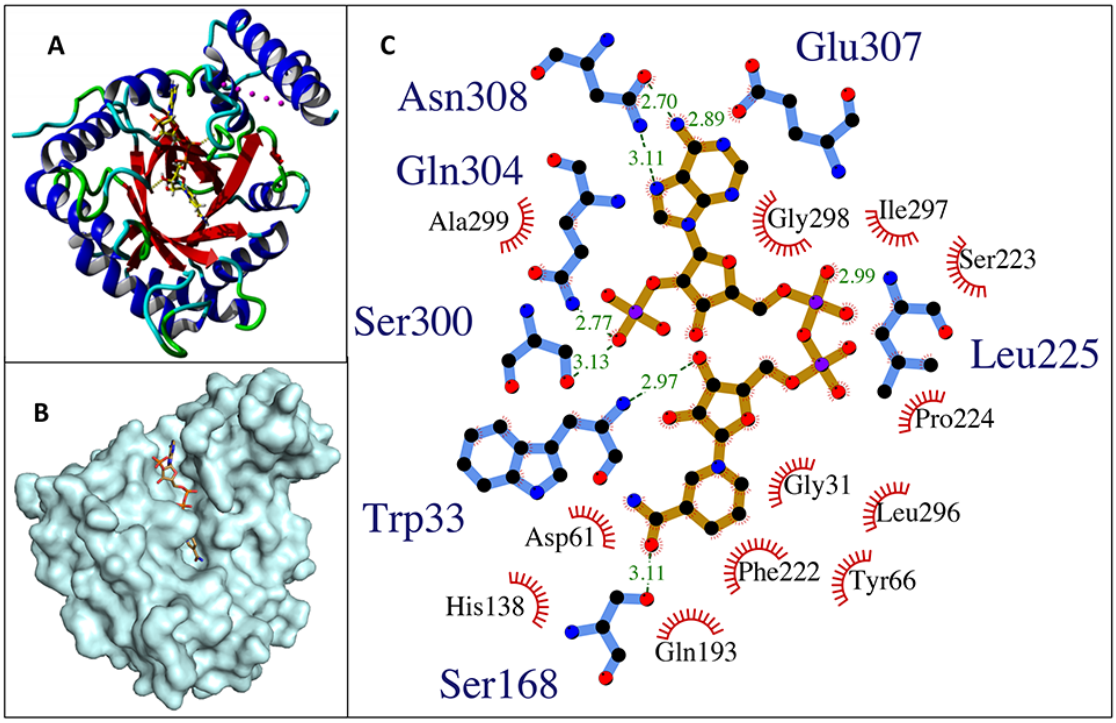
526


3.000
2.500
2.000
1.500
1.000
0.500
0.000
KIE
14.000
12.000
10.000
8.000
6.000
4.000
2.000
0.000
k_{cat} S^{-1}
4-pyridine carboxaldehyde
methylglyoxal
butyraldehyde
3-pyridine carbxoaldehyde
furfural



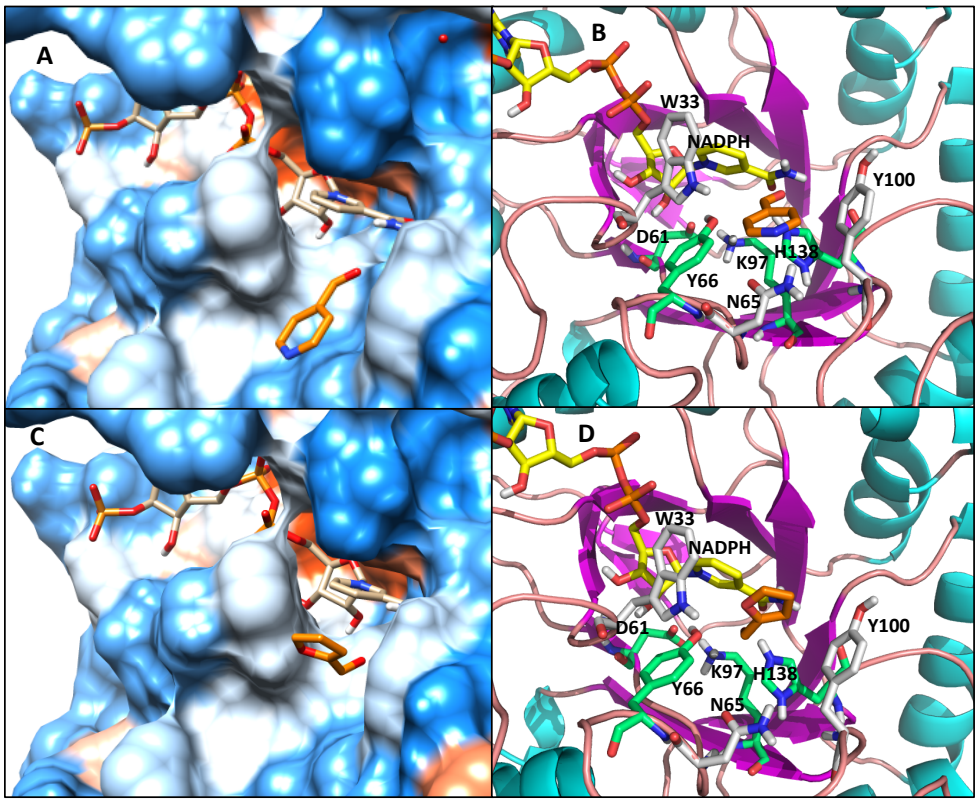
531

532


A
B
C
Glu307
Asn308
Gln304
Ala299
Gly298
Ile297
Ser223
Ser300
Leu225
Pro224
Trp33
Gly31
Leu296
Asp61
Phe222
Tyr66
His138
Gln193
Ser168
2.70
2.89
3.11
2.99
2.77
3.13
2.97
3.11

533

534


A
B
W33
NADPH
Y100
D61
K97
H138
Y66
N65
C
D
W33
NADPH
D61
Y100
K97
H138
Y66
N65

535

536