Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d3ob01937c

A comparative study of two aldehyde dehydrogenases from *Sphingobium* sp.: the substrate spectrum and catalytic mechanism†

Siyi Chen, Jieyu Zhou,* Xiangyuan Gu and Ye Ni *

Received 29th November 2023,
Accepted 12th January 2024

DOI: 10.1039/d3ob01937c

rsc.li/obc

Biocatalytic oxidation is one of the most important and indispensable organic reactions for the development of green and sustainable biomanufacturing processes. NAD(P)$^+$-dependent aldehyde dehydrogenase (ALDH) catalyzes the oxidation of aldehydes to carboxylic acids. Here, two ALDHs, *Sp*ALDH1 and *Sp*ALDH2, were identified from *Sphingobium* sp. SYK-6. They belong to different ALDH families and share only 32.30% amino acid identity. Interestingly, *Sp*ALDH1 and *Sp*ALDH2 exhibit significantly different enzymatic properties and substrate profiles. *Sp*ALDH2 has better thermostability than *Sp*ALDH1. *Sp*ALDH1 is a metalloenzyme and is activated by potassium ions, while *Sp*ALDH2 is not metallic-dependent. Compared with *Sp*ALDH1, *Sp*ALDH2 has a relatively broad substrate spectrum toward aromatic aldehydes. Based on homology modeling and molecular docking analysis, mechanisms underlying the substrate specificity of ALDHs were elucidated. For both ALDHs, hydrophobicity of substrate binding pockets is important for the catalytic properties, especially substrate specificity. Notably, optimization of the flexible loop 444–457 reforms a hydrogen bond between pyridine substrates and *Sp*ALDH1, contributing to the high catalytic activity. Finally, a coupling reaction catalyzed by ALDHs and NOX was constructed for efficient production of aromatic carboxylic acids.

Introduction

Aromatic acids are essential chemical materials widely utilized in the manufacture of pharmaceuticals, pesticides, fragrances, and fine chemicals. The oxidation of the corresponding aldehydes represents a significant approach for synthesizing aromatic carboxylic acids. These aldehyde substrates can be derived from lignin, offering a cost-effective option through biorefinery processes. Traditional chemical oxidation methods involve the use of strong acids, strong bases, and substantial quantities of metal oxidants.[1–5] However, due to the lack of regioselectivity, chemical processes often result in the formation of undesired by-products for substrates containing multiple oxidizable functional groups.[6] Moreover, the excessive use of organic solvents such as dipropylene glycol dimethyl ether and harsh reaction conditions[7] are not in line with the principles of "Green Chemistry" in modern industries.

Biooxidation of aldehydes can be accomplished either using aldehyde dehydrogenases or aldehyde oxidases. Aldehyde oxidases have limited regioselectivity and may oxidize hydroxyl substituent groups of aromatic aldehydes in the presence of oxygen.[8] In contrast, NAD(P)$^+$-dependent aldehyde dehydrogenase (ALDH, EC 1.2.1.5), a superfamily of enzymes with wide substrate specificity, provides a more environmentally friendly and versatile approach. These enzymes, exemplified by their compelling ability to catalyze the irreversible oxidation of a chemically diverse range of aldehydes to their corresponding acid metabolites, play a pivotal role in the biotransformation or detoxification of various exogenous and physiologically important endogenous aldehydes.[9] ALDH commonly employs NAD$^+$ (although several use NAD(P)$^+$) as a cofactor, ensuring efficient and selective biooxidation of aldehydes while maintaining superior regioselectivity.[10] Although ALDHs have been well characterized since the early 1980s, their performance in biocatalytic synthesis has only gained attention recently. Knaus *et al.*[11] applied a combination of Ene-reductases (ERs) and ALDHs to convert α-substituted α,β-unsaturated aldehydes into related optically active saturated carboxylic acids. Similarly, Wu *et al.*[12] co-expressed styrene monooxygenase (SMO), styrene isomerase (SOI), and acetaldehyde dehydrogenase (*Ec*ALDH) in *Escherichia coli* for the regioselective oxidation of styrene and its analogues to produce the corresponding acids in high conversion. These studies emphasize the versatility of ALDH in enzyme cascade

Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi 214122, Jiangsu, China.
E-mail: zhoujieyu@jiangnan.edu.cn, yni@jiangnan.edu.cn

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3ob01937c

reactions and highlight the necessity to explore and obtain more ALDHs with high stability and adaptability to different reaction conditions.

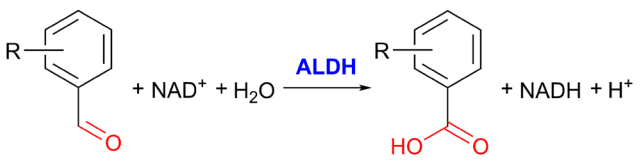
ALDH monomers typically consist of three domains, including the oligomerization domain, NAD^+-binding domain, and substrate-binding domain.[13,14] The unique geometry and chemical characteristics of the substrate-binding domains confer distinct substrate-binding specificity. A meticulous study conducted by Knaus *et al.*[1] compared the substrate profiles of three ALDHs derived from bovine lens (*Bov*-ALDH), *Escherichia coli* (*Ec*-ALDH), and *Pseudomonas putida* (*Pp*-ALDH). Significant differences in catalytic efficiency toward hydroxyl- and halogen-substituted benzaldehydes, as well as heteroaromatic ring-substituted aldehyde substrates were observed with *Bov*-ALDH and *Ec*-ALDH, whereas *Pp*-ALDH demonstrated substrate preference towards aliphatic aldehydes. This study provides valuable insights into the substrate specificity of ALDHs concerning different properties of substituents such as the position and electron-donating/withdrawing nature. However, no information on the structural features of ALDHs and their interactions with substrates has been reported. An in-depth study on the structural and catalytic properties of ALDHs could shed more light on the molecular mechanisms, further advancing their applications.

Here, two NAD^+-dependent ALDHs derived from *Sphingobium* sp., *Sp*ALDH1 and *Sp*ALDH2, were identified and characterized. *Sp*ALDH1 and *Sp*ALDH2 exhibited notable stability and could catalyze the formation of aromatic carboxylic acids from aromatic aldehydes (Scheme 1). A comprehensive comparison was conducted to assess their enzymatic properties, metal-dependency, substrate profiles, *etc*. Furthermore, structural modeling and interaction analysis were performed to understand the mechanisms underlying their substrate specificity toward aromatic aldehydes. This study provides valuable insights into the catalytic mechanisms of *Sp*ALDH1 and *Sp*ALDH2, providing new prospects for their potential application in the synthesis of aromatic carboxylic acids.

Experimental

Reagents

All reagents used were purchased from Macklin Inc. (Shanghai, China) and Sinopharm Chemical Reagent (Wuhan, China). The nucleotide sequences of *Sp*ALDH1, *Sp*ALDH2 and *Sp*NOX (derived from *Streptococcus pyogenes*, FAD-dependent oxidoreductase) were commercially synthesized by Yixin Biotechnology (Wuxi, China). Coenzymes NAD^+ and $NADP^+$ were purchased from Bangtai Biological Engineering Co., Ltd (Shenzhen, China).



Scheme 1 ALDHs catalyzed the oxidation of aromatic aldehydes to aromatic acids.

Protein expression

Recombinant plasmids pET-28a (+)-*Sp*ALDH1 and pET-28a (+)-*Sp*ALDH2 were transformed into *E. coli* BL21 (DE3). Recombinant strains were cultivated in LB broth supplemented with 50 μg ml^{-1} kanamycin at 37 °C and 180 rpm. When OD_{600} reached around 0.8–1.0, protein expression was induced using β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.2 mM. Cells were then incubated for 16–18 h at 16 °C. Afterward, the cells were harvested by centrifugation at 8000 rpm for 5 min at 4 °C and suspended in 50 mM PBS (pH 7.0). Then, cell suspensions were sonicated on ice using an ultrasonic cell pulverizer for 10 min at 300 W. The resultant was centrifuged again at 8000 rpm for 30 min at 4 °C, and the supernatant was collected as a cell-free extract. The expression level of target proteins was assessed by SDS-PAGE, and the protein concentration was determined by the Bradford method.

Protein purification

Cell-free extracts were purified using an AKTA Avant protein purification system (GE Healthcare). The supernatant was loaded onto a 5 mL HisTrap HP affinity column that was pre-equilibrated with equilibration buffer (20 mM imidazole, 25 mM PBS pH 7.4, 500 mM NaCl, and 5% glycerol). The bound protein was eluted with a linear gradient of imidazole ranging from 20 to 500 mM. The fractions containing target proteins were then concentrated by ultrafiltration using Amicon-Ultra-15, 30-KMWCO filters (Millipore, USA). The purified proteins were verified by SDS-PAGE and preserved at −80 °C with 20% (v/v) glycerol for further study.

Activity assay of ALDHs

The activity of *Sp*ALDH1 and *Sp*ALDH2 was determined using the phenazine methosulfate (PMS)-nitroblue tetrazolium (NBT) method. In the presence of PMS, NBT reacts with the NADH produced by ALDHs to produce insoluble blue-purple formazan, which has absorbance at 562 nm. To calculate the corresponding enzyme activity under different pH conditions, the molar absorbance coefficient (ε) was determined for each pH condition (Fig. S1 and Table S1†). In this study, a 200 μL reaction system included 180 μl buffer containing 1 mM syringaldehyde (**1a**), 1 mM NAD^+, 0.07% (w/v) gelatin, 0.3 mM NBT and 0.03 mM PMS, and an appropriate amount of the purified enzyme. The reaction was started by adding an appropriate amount of purified enzymes into the reaction system, and OD_{562} was monitored for the first 5 min. One unit (U) of ALDH activity is defined as the amount of enzyme required to produce 1 μmol of NADH per min.

Effect of pH on ALDH activity

The pH profiles of purified *Sp*ALDH1 and *Sp*ALDH2 were determined at 30 °C using the standard activity assay protocol in the following buffer solutions: citric acid/sodium citrate buffer (citrate buffer) (pH 5.0–6.5, 100 mM), phosphate buffer (pH 6.0–8.0, 50 mM), Tris–HCl (pH 7.0–9.0, 50 mM), and glycine–NaOH (pH 8.5–11.0, 50 mM). The protein quantity was kept constant across all reaction systems. All the activities were assayed in triplicate.

Effect of temperature on ALDH activity

The optimal temperature for *Sp*ALDH1 and *Sp*ALDH2 was determined using the standard activity assay protocol at various temperatures ranging from 28 to 50 °C. The purified enzyme solution was incubated on ice, while the buffer was maintained at a series of test temperatures in a water bath. Citrate buffer (100 mM, pH 6.5) was used to test *Sp*ALDH1, while glycine–NaOH buffer (50 mM, pH 10.5) was employed for *Sp*ALDH2. The thermostability of *Sp*ALDH1 and *Sp*ALDH2 was investigated by incubating 0.1 mg mL^{-1} purified enzymes and buffer at 30 and 40 °C. The microplate reader was set at 45 °C for activity measurements and residual activity was measured to calculate the half-life. All activities were assayed in triplicate.

Effect of metal ions on ALDH activity

Effects of various metal ions and additives (including Ni^{2+}, Mg^{2+}, Fe^{2+}, Zn^{2+}, Cu^{2+}, Ca^{2+}, Mn^{2+}, Co^{2+} and EDTA) on the activity of *Sp*ALDH1 and *Sp*ALDH2 were examined by adding each metal ion to the purified enzyme and incubating at 30 °C for 20 min. Afterwards, the enzyme activities were measured under the standard conditions. Control was performed in the absence of any metal ions. All the activities were assayed in triplicate.

Effect of K^+ on the activity of *Sp*ALDH1

Effect of K^+ on the activity of *Sp*ALDH1 was investigated by conducting enzyme activity assays in the presence of 0.0–1.0 M KCl in the reaction mixture. Control was performed in the absence of KCl, and all the activities were assayed in triplicate.

Substrate spectrum and kinetic parameters

The specific activities of *Sp*ALDH1 and *Sp*ALDH2 toward 29 aldehyde substrates were determined using the standard activity assay protocol. Kinetic parameters were measured by employing a standard activity protocol at different substrate concentrations (0.02–2 mM). K_m, K_i, and V_{max} were calculated by fitting the substrate inhibition equation $v = V_{max}[S]/(K_m + [S](1 + [S]/K_i))$ using Origin Pro 8.5.

Homology modeling

The overall structures of apo-*Sp*ALDH1 and apo-*Sp*ALDH2 were predicted using AlphaFold2. Simultaneously, ALDH structures containing NAD^+ and K^+ in the PDB database were selected for multi-template homologous modeling by employing Discovery Studio 2021. Subsequent addition of the coenzyme into apo-forms of two enzymes was achieved *via* structural alignment. For *Sp*ALDH1, PDB templates sharing 39–43% amino acid sequence identity, namely 4O5H and 1WNB, were utilized. In the case of *Sp*ALDH2, PDB structures 3B4 W, 5GTK, and 4OU2 were employed with 34–45% identity.

Molecular docking and *in situ* ligand minimization

The homology structures of *Sp*ALDH1 and *Sp*ALDH2 were docked separately with various aldehydes using the CDOCKER module, applying a CHARMm force field. Prior to docking, the homology structures and substrates were prepared using the "prepare protein" and "prepare ligand" protocols, respectively. The active pocket was defined from receptor cavities. The number of hits for the ligand was set as 100. The number of starting random conformations generated from equilibration and minimization of the starting ligand structure was set as 10. Maximum bad orientation was set as 800, and orientation van der Waals energy threshold was set as 300.

Ligand minimization was performed using an *in situ* ligand minimization tool based on the enzyme–substrate complex obtained by molecular docking (Discovery Studio 2021).[15] Atoms within the active pocket sphere defined by the receptor cavities were set to move freely. The sphere center coordinates for *Sp*ALDH1 are (−13.65, −1.93, 97.73), and (−37.24, 8.99, 42.07) for *Sp*ALDH2, and the radius of the sphere is 6.7 Å. The residues within 6.0 Å around the substrate were allowed to move. CHARMm was chosen for atom typing, the acceptor hydrogen atom near the ligand was allowed to be flexible, and substructure constraints were set as false. A smart minimizer algorithm was chosen for performing the minimization and the number of minimization steps was set as 1000. The RMS gradient, indicating the tolerance (kcal $(mol Å)^{-1}$) that should be applied to the average gradient over a minimization period, was set to 0.001. Minimization Energy Change, representing the tolerance that should be applied to the change in total energy during a cycle of minimization, was set to 0. After calculation, the simulation results were converged to a single structural output for force analysis. After molecular docking and *in situ* minimization, conformations meeting the criteria of nucleophilic attack angle ($S_{cys}\cdots C{=}O$) between 95° and 105° and a nucleophilic attack distance (d) of approximately 3.5 Å are considered to conform to the catalytic mechanism.[16,17]

Preparation of syringic acid catalyzed by coupling ALDHs with *Sp*NOX

The reaction catalyzed by ALDH coupled with *Sp*NOX was conducted to produce syringic acid (**1b**). The consumption of **1a** and the formation of **1b** were analyzed by HPLC using a VWD detector, and the conversion ratio was calculated. Reactions with 20 mM and 50 mM **1a** were carried out, respectively (Tables S2 and S3†). For reaction with 50 mM **1a**, the substrate was added in 3 batches (20 mM at 0 h, 20 mM at 2 h, and 10 mM at 6 h), and *Sp*NOX and ALDHs were added at triple dosage. Samples were collected at 0, 2, 4, 6, and 8 h, and the reactions were terminated by mixing with the same volume of

0.2 N HCl. Supernatants obtained by centrifugation (12 000 rpm for 5 min) were filtered and analyzed using an Agilent 1260 HPLC using a ZORBAX SB-C18 column (250 × 4.6 mm, 5 μm) at 279 nm. The HPLC mobile phase consisted of water (75%) and acetonitrile (25%) supplemented with formic acid (0.1%), and was applied at a flow rate of 1.0 ml min^{-1}.

Results and discussion

Identification of *Sp*ALDH1 and *Sp*ALDH2

Sphingobium sp. SYK-6 is a Gram-negative bacterium with the genome sequenced, and it is capable of degrading aromatic compounds such as benzene, toluene, and xylene. Numerous studies have demonstrated its high biodegradability.[18] From SYK-6, nine ALDHs responsible for the catabolism of **1a** have been identified by Kamimura *et al.*[19] Here, two ALDHs, WP_014075087.1 (*Sp*ALDH1) and WP_014077148.1 (*Sp*ALDH2), were chosen for a comparison study. In a previous report, *Sp*ALDH1 exhibited the highest specific activity against syringaldehyde, while *Sp*ALDH2 has been shown to play a key role in the metabolism of syringaldehyde.[19] In this study, we aim to provide a deeper insight into aldehyde dehydrogenases from the perspectives of catalytic mechanism and structure–activity relationship through careful characterization and comparison based on enzymatic properties. *Sp*ALDH1 and *Sp*ALDH2 share only 32.30% amino acid identity, yet both have the potential to catalyze the oxidation of aromatic aldehydes into the corresponding carboxylic acids.

Sequence similarity searches for both enzymes were performed using the BLAST algorithm (**https://www.ncbi.nlm.nih.gov/BLAST/**), followed by multiple sequence alignment and phylogenetic analysis using MEGA 7.0 software. The results were displayed and annotated by iTOL (**https://itol.embl.de/**), which utilized differences in branch lengths as a basis for coloring (Fig. S2†). The analysis indicated that *Sp*ALDH1 and *Sp*ALDH2 are located in different branches, suggesting that they may belong to distinct ALDH families (Fig. 1A). *Sp*ALDH1 exhibited better solubility than *Sp*ALDH2, and the purified protein was observed as a single band of 57 kDa in SDS-PAGE, consistent with its calculated molecular weight of 56.15 kDa (Fig. 1B). On the other hand, the soluble fraction of *Sp*ALDH2 accounted for approximately 50% of the total protein, and the purified protein showed a band of 58 kDa, which is in good agreement with its molecular weight of 57.25 kDa (Fig. 1C).

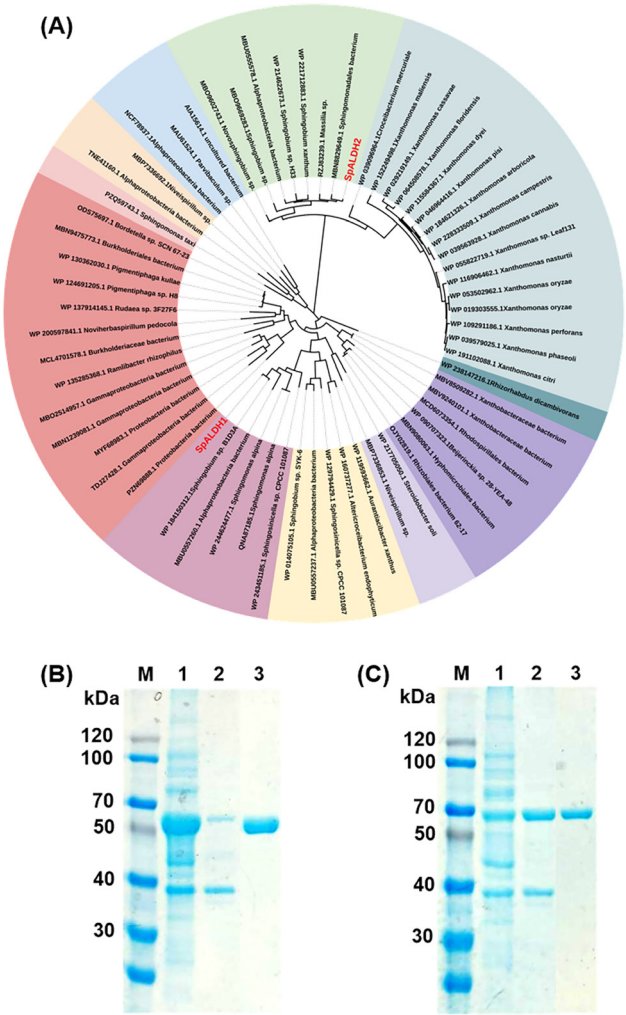


Fig. 1 (A) Phylogenetic analysis of *Sp*ALDH1, *Sp*ALDH2 and other bacterial ALDHs. The phylogenetic tree was constructed with MEGA 7.0 software by the neighbor-joining method; (B) SDS-PAGE analysis of the recombinant expression of *Sp*ALDH1 and (C) *Sp*ALDH2. Lane M: marker, lane 1: supernatant, lane 2: precipitate, lane 3: purified ALDH.

Effect of pH on the activity of *Sp*ALDH1 and *Sp*ALDH2

Maintaining an optimal pH is critical for efficient biocatalytic reactions. The pH preferences of purified *Sp*ALDH1 and *Sp*ALDH2 were assessed using different buffer systems with a pH range of 5.0–11.0. Significantly different pH preferences were observed. Interestingly, the optimum pH for *Sp*ALDH1 was pH 6.5 in citrate buffer with a specific activity of 2.06 U mg^{-1}, while *Sp*ALDH1 showed a specific activity of merely 0.14 U mg^{-1} in pH 6.5 PBS buffer, indicating that various buffer systems have significant impacts on the catalytic activity of ALDHs (Fig. 2A). In contrast, *Sp*ALDH2 has an obvious preference for an alkaline environment, and showed the highest activity of 3.91 U mg^{-1} in glycine–NaOH buffer at pH 11.0. Considering the buffering capacity, pH 10.5 was selected for *Sp*ALDH2, in which over 90% of the highest activity (3.59 U mg^{-1}) was retained (Fig. 2B).

Effect of temperature on the activity of *Sp*ALDH1 and *Sp*ALDH2

The reaction temperature is critical for biocatalytic reactions. The optimum temperature of *Sp*ALDH1 was determined to be 35 °C with a specific activity of 2.38 U mg^{-1} (Fig. 2C), and 80% activity could be maintained over the range of 30–45 °C. At temperatures above 45 °C, the activity decreases rapidly, and only 20% of the maximum activity was determined at 50 °C. It is interesting to note that the activity of *Sp*ALDH2 increases linearly from 28 °C to 45 °C, reaching an optimal temperature

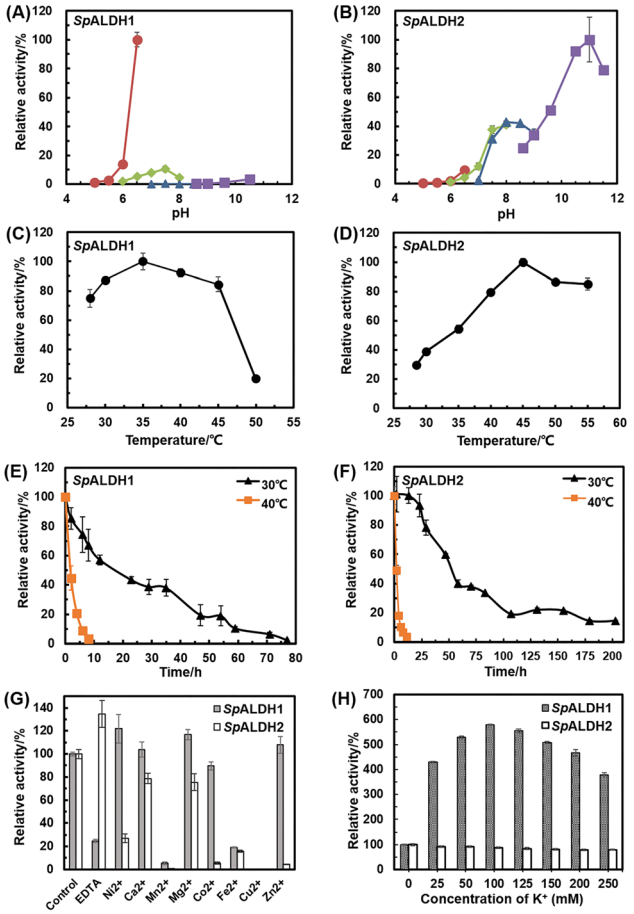


Fig. 2 Effect of pH on the activity of *Sp*ALDH1 (A) and *Sp*ALDH2 (B). (-●-): citrate buffer (pH 5.0–6.5, 100 mM), (-◆-): phosphate buffer (pH 6.0–8.0, 50 mM), (-▲-): Tris–HCl (pH 7.0–9.0, 50 mM), (-■-): glycine–NaOH (pH 8.5–11.0, 50 mM). Optimum temperature of *Sp*ALDH1 (C) and *Sp*ALDH2 (D). Thermostability of *Sp*ALDH1 (E) and *Sp*ALDH2 (F). Effects of divalent metal ions on *Sp*ALDH1 and *Sp*ALDH2 (G). Effect of K^+ on *Sp*ALDH1 (H).

of 45 °C (12.42 U mg^{-1}), after which the activity began to decrease (Fig. 2D). However, the enzyme retains over 85% of activity at a high temperature of 55 °C. The thermostability of both enzymes was examined at 30 °C and 40 °C. At 40 °C, both enzymes demonstrated similar profiles, with *Sp*ALDH1 retaining only 44.60% residual activity after 2 h (Fig. 2E), compared with the 49.20% of *Sp*ALDH2 (Fig. 2F). Notably, *Sp*ALDH2 exhibited better thermostability at 30 °C, with a calculated half-life of 55.8 h as opposed to the 28.2 h of *Sp*ALDH1. Taken together, *Sp*ALDH2 features a higher optimum temperature and better thermostability at moderate temperatures than *Sp*ALDH1.

Effect of metal ions and EDTA on the activity of *Sp*ALDH1 and *Sp*ALDH2

Metal ions have been found to play a key role in the catalytic activity of many enzymes by affecting their conformation and the stability of reaction intermediates.[20–22] Metal ions can also be involved in charge transfer and redox processes, which ultimately impacts the efficiency of enzymatic reactions. The effects of divalent metal ions on the catalytic activity of *Sp*ALDH1 and *Sp*ALDH2 were examined (Fig. 2G). Interestingly, significant differences were observed. The addition of Ni^{2+}, Ca^{2+}, Mg^{2+} and Zn^{2+} resulted in an increased catalytic activity of *Sp*ALDH1. In contrast, the addition of metal ion chelator EDTA resulted in a residual activity of only 20%. Conversely, the addition of all metal ions inhibited the activity of *Sp*ALDH2 more or less. It is worth noting that metal ions such as Ni^{2+}, Zn^{2+}, and Mg^{2+} showed a significant inhibitory effect on *Sp*ALDH2, although they are beneficial for *Sp*ALDH1. Moreover, Mn^{2+}, Fe^{2+}, and Cu^{2+} strongly inhibited the enzymatic activity of both enzymes, with less than 20% residual activity for both enzymes. Taken together, *Sp*ALDH1 is a metalloenzyme, while for *Sp*ALDH2, the above metal ions should be avoided in the reaction systems.

Several previous studies have reported the activation of ALDHs by K^+. Garza-Ramos *et al.* conducted structural analysis of K^+-dependent *Pa*BADH and *So*BADH, in which a significant change in the tertiary structure of *Pa*BADH in the absence of K^+ ions or at low ion strengths was observed.[23] Valenzuela-Soto *et al.* compared the NMR spectra of K^+-dependent *pk*BADH in the presence and absence of NAD^+ and K^+. A possible mechanism was proposed whereby high K^+ concentrations could influence the chemical environment in the catalytic center, leading to increased affinity for NAD^+.[24,25] Initially, addition of 500 mM KCl was attempted. The activity of *Sp*ALDH2 remained unchanged while the activity of *Sp*ALDH1 increased by approximately 5-fold, indicating that *Sp*ALDH1 is K^+-activated. In further experiments with 25–250 mM KCl, the highest activity of 13.77 U mg^{-1} was determined at 100 mM, representing 5.8-times improvement (Fig. 2H), which was consistent with the above report. The beneficial effects of K^+ on ALDHs are likely associated with their conformational changes elicited by potential potassium-binding sites.

Substrate specificity of *Sp*ALDH1 and *Sp*ALDH2

The substrate spectrum of *Sp*ALDH1 and *Sp*ALDH2 was evaluated toward a total of 29 aromatic and alkyl aldehydes, using **1a** as the model substrate. Both enzymes showed similar activities for **1a**. However, *Sp*ALDH1 demonstrated strong specificity towards several aromatic aldehydes including **3a** (3,5-dimethoxybenzaldehyde), **8a** (3-methoxybenzaldehyde), **12a** (4-nitrobenzaldehyde), **13a** (4-methylsulphonyl benzaldehyde), **14a** (4-chlorobenzaldehyde), **16a** (3-chlorobenzaldehyde), **17a** (4-bromobenzaldehyde), **23a** (3-pyridinecarboxaldehyde), and **24a** (4-pyridinecarboxaldehyde), exhibiting 5.2–10.8 fold higher activities than that for **1a** (Fig. 3A). In contrast, *Sp*ALDH2 exhibited a broader substrate spectrum, except for **21a** (indole-3-carboxaldehyde) (Fig. 3B). Notably, the highest activity of *Sp*ALDH2 was determined with **6a** (3,4-dihydroxybenzaldehyde), which is around twice that of **1a**. Interestingly, no activity was detected toward **6a** for *Sp*ALDH1.

Furthermore, substrates **1a–11a** were selected and sufficient NAD^+ and reaction time were applied to determine the conver-

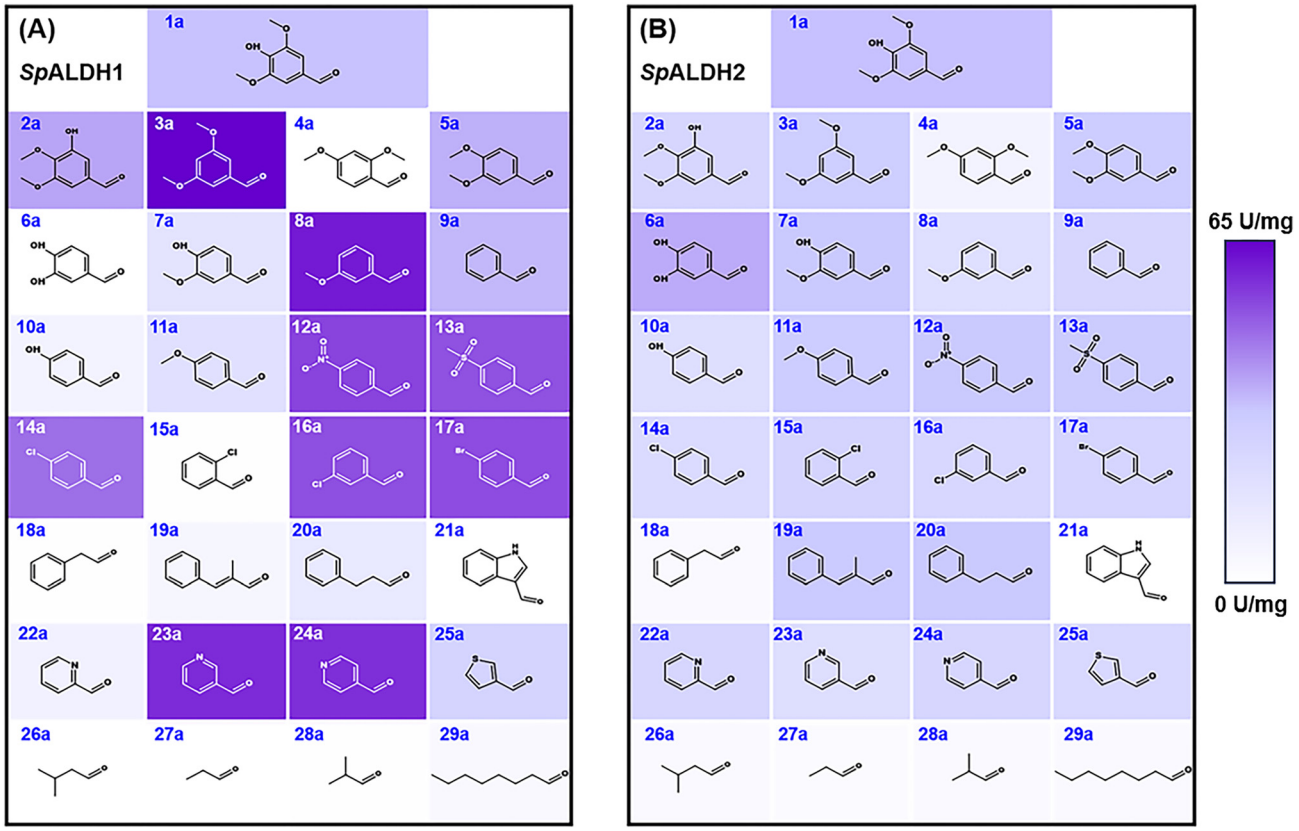


Fig. 3 Substrate spectrum of *Sp*ALDH1 (A) and *Sp*ALDH2 (B). The shades of purple color represent specific activity values of ALDHs.

sion ratios. Fig. S3† demonstrates that both enzymes exhibit excellent chemoselectivity in catalysing the model substrate **1a**. Neither methoxy nor hydroxy substituents on the phenyl ring were catalyzed by the enzymes, and therefore no undesirable by-product was generated. For *Sp*ALDH2, over 99% conversion was achieved for all the substrates after 18 h. However, for *Sp*ALDH1, over 99% conversions were attained with all the substrates except **4a** and **6a**, for which only 3.2% and 5.3% conversions were observed, respectively (Table S4†).

Homology modeling and molecular docking of syringaldehyde

To gain insights into the molecular mechanisms underlying the different catalytic performances of *Sp*ALDH1 and *Sp*ALDH2 towards various substituted aromatic aldehydes, homology models of both ALDHs containing coenzymes were constructed. The catalytic activity of ALDHs relies on a conserved quadruplet, Asn–Cys–Gly–Cys. The catalytic Cys, under Glu activation, undertakes a nucleophilic attack upon the aldehyde carbon in the substrate. Meanwhile, Asn and Gly participate in substrate and coenzyme binding, respectively. In the catalytic process, a nucleophilic attack on aldehyde substrates is initiated by a thiolate anion of Cys, forming a tetrahedral thiohemiacetal oxyanion intermediate. Subsequently, a hydride ion is transferred to the pyridinium moiety of NAD, thereby an intermediate thioester is produced alongside the generation of NADH. Finally, through the hydrolysis of thioester, the reconstitution of the thiolate entity occurs, concomitantly giving rise to the product carboxylic acid (Fig. S4†). Here, the conformation of nucleophilic attack that satisfies the catalytic mechanism is defined as a nucleophilic attack angle ($S_{cys}\cdots C{=}O$) of $95° \leq \theta \leq 105°$ and a nucleophilic attack distance (d) of approximately 3.5 Å.[16,17]

First, the model substrate **1a** was docked into *Sp*ALDH1 and *Sp*ALDH2 separately, followed by interaction analysis. For *Sp*ALDH1, the conserved catalytic tetramer comprises N161-E257-G288-C291. The conformation aligns with the catalytic mechanism, featuring a nucleophilic attack angle of 97° and a nucleophilic attack distance of 3.5 Å (Fig. 4A). The phenyl ring of the substrate forms a pi–pi stack with F455, and two methoxy substituents engage in hydrophobic pi–alkyl interactions with M165 and L166 individually. Additionally, the *para*-hydroxy group interacts with Q467 to form a hydrogen bond. Collectively, **1a** is well anchored in the substrate binding pocket of *Sp*ALDH1.

For *Sp*ALDH2, the catalytic tetramer is N166-E263-G294-C297. The catalytic conformation presents a nucleophilic attack angle of 101° and a distance of 3.2 Å (Fig. 4B). F298 engages in pi–pi stacking with the benzene ring of **1a**. The residues including F461, Y174, V455, W102, and F112 form hydrophobic interactions with the methoxy groups. Unlike *Sp*ALDH1, the *para*-hydroxyl group does not participate in hydrogen bonding with neighboring residues.

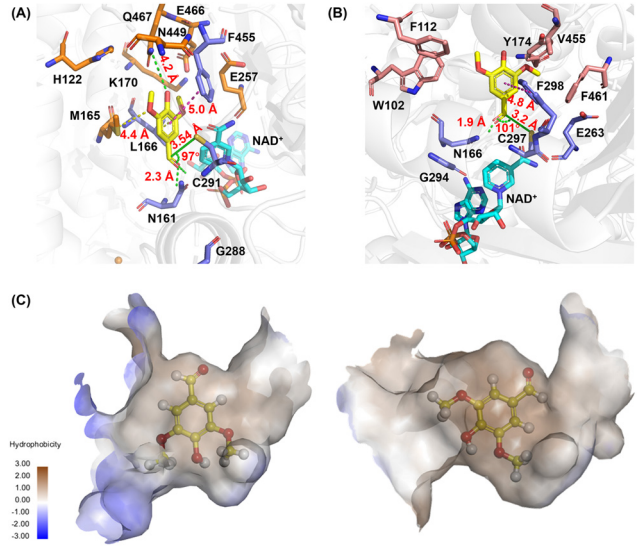


Fig. 4 Molecular docking of ALDHs with **1a**. (A) *Sp*ALDH1; (B) *Sp*ALDH2. Active site residues, **1a** substrate and NAD^+ cofactor are shown as sticks, and carbon atoms are represented by slate, yellow and cyan, respectively. Hydrophobic amino acids are represented by pink and hydrophilic amino acids are represented by orange. ALDHs are depicted as grey. Oxygen, red; nitrogen, blue; sulfur, yellow orange; potassium ions, purple ball; sodium ions, orange ball. Hydrogen bonds: green dotted lines; alkyl interactions: yellow dotted lines; pi–pi stacking interactions: light magenta dotted lines. (C) Hydrophobicity of protein binding pockets of *Sp*ALDH1 (left) and *Sp*ALDH2 (right). **1a** is shown as a ball and stick model. Dark blue indicates higher hydrophilicity, and dark brown indicates higher hydrophobicity.

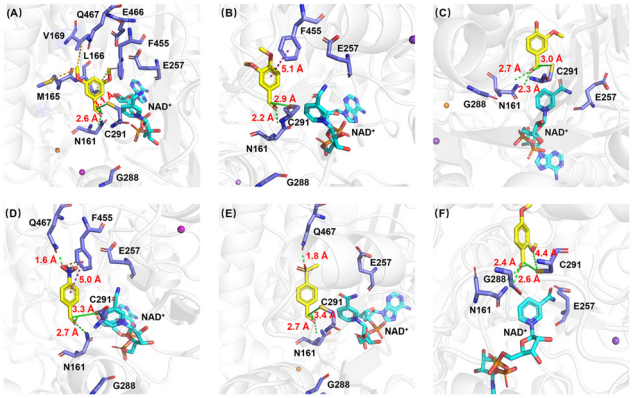


Fig. 5 Molecular docking of *Sp*ALDH1 with aromatic aldehydes. (A) **3a**; (B) **5a**; (C) **7a**; (D) **12a**; (E) **13a**; (F) **4a**. Active site residues, substrates and NAD^+ cofactor are shown as sticks, and carbon atoms are represented by slate, yellow and cyan, respectively. ALDHs are depicted as grey. Oxygen, red; nitrogen, blue; sulfur, yellow orange; potassium ions, purple ball; sodium ions, orange ball. Hydrogen bonds: green dotted lines; pi–alkyl/alkyl interactions: yellow dotted lines; carbon hydrogen bonds: pale green dotted lines; pi–pi stacking interactions: light magenta dotted lines; pi–cation/pi–anion interactions: bright orange dotted lines.

Analysis of solvent hydrophobicity surfaces reveals a pronounced hydrophobic nature of the binding pocket of *Sp*ALDH2 (Fig. 4C). This characteristic property enables the accommodation of a wide range of hydrophobic aromatic aldehydes, leading to a broad substrate spectrum. In the case of *Sp*ALDH1, the hydrophilic bottom of its substrate binding pocket is characterized by a number of polar amino acids, including H122, M165, K170, E257, E466, Q467, and N449. These residues could interact with the substituents of aromatic aldehydes, thereby imparting *Sp*ALDH1 with a distinctive substrate specificity.

Interaction analysis reveals the substrate specificity of *Sp*ALDH1

*Sp*ALDH1 exhibits the highest activity towards **3a** and **8a**, both of which share the common feature of having a methoxy substituent at either one or both of the *meta*-positions on the benzene ring. Interaction analysis between *Sp*ALDH1 and substrate **3a** reveals that the *meta*-methoxy group on one side of **3a** engages in hydrophobic alkyl/pi–alkyl interactions with F455 and L166, while the other *meta*-methoxy group forms alkyl interaction with V169 and hydrogen bonds with M165 and Q467 (Fig. 5A). Therefore, the presence of *meta*-methoxy groups stabilizes the substrate, thereby facilitating the nucleophilic attack initiated by the catalytic residue C291.

For substrates **5a** (3,4-dimethoxybenzaldehyde) and **7a** (4-hydroxy-3-methoxybenzaldehyde) with one *meta*-methoxy group, the presence of an additional *para*-substituted methoxyl and hydroxyl groups resulted in over 80% decreased activity. However, docking conformations of **5a** and **7a** are consistent with the catalytic mechanism. It is speculated that the incompatibility between these two substrates and *Sp*ALDH1 might be caused by the electronic effect of the substituents. In **5a**, two *para*-methoxy groups introduce a conjugation effect with increased electron density, which stabilizes the aldehyde group and enhances the specificity of oxidation (Fig. 5B). For **7a**, the *para*-hydroxy group creates lone pair electrons on the oxygen through a resonance effect (Fig. 5C). The lone pair electrons can transfer electron density through conjugation and resonance effects, resulting in an increased electron density of the aldehyde group. Thus, the electron-rich aldehyde group is less susceptible to oxidation. It is worth noting that the methoxy group has a weaker electron-donating ability than the hydroxyl group, which makes the enzyme more active toward **5a** than **7a**. Meanwhile, a similar trend was observed for **10a** (4-hydroxybenzaldehyde) (0.79 U mg^{-1}) and **11a** (4-methoxybenzaldehyde) (1.82 U mg^{-1}) with only one *para*-hydroxyl or methoxy substituent.

Remarkably, a significantly enhanced activity of *Sp*ALDH1 was observed with benzaldehyde substrates harboring electron-withdrawing groups (such as nitro and methylsulfonyl) at the *para*-position, namely **12a** and **13a**. These groups can draw electron density from the neighboring aromatic ring through electron conjugation, thereby elevating the electron affinity of the aldehyde group and its susceptibility to oxidation. However, this phenomenon is less pronounced in *Sp*ALDH2. Therefore, it is presumable that the electronic effect is not the

key factor, and other interactions between *Sp*ALDH1 and the substrates might have more significant effects. Docking analysis reveals that Q467 can form hydrogen bonds with nitro and sulfone groups, favoring substrate anchoring in the catalytic conformation (Fig. 5D–E). Furthermore, for **12a**, F455 is also capable of forming pi–cation and pi–anion interactions with the nitro group.

However, enzyme–ligand interactions can also contribute to unfavorable substrate binding. For **4a** (2,4-dimethoxybenzaldehyde), the *ortho*-substituted methoxy group could potentially form a stable hydrogen bond with the catalytic residue C291 (Fig. 5F). However, *Sp*ALDH1 showed no activity toward **4a**, suggesting that the newly formed hydrogen bond could interfere with the catalytic conformation of C291 and its nucleophilic attack on the substrate.

The unfavorable substrate binding is more evident in the docking conformation of **6a** with *Sp*ALDH1. The carbonyl oxygen on the side chain at position 467 forms a hydrogen bond with the hydroxyl group at the adjacent position of **6a**, leading to an unfavorable donor–donor interaction with the amino group at the side chain. In this case, the nucleophilic attack angle is found to be 92.3° (Fig. 6A). The observed incompatibility between **6a** and *Sp*ALDH1 in molecular docking corresponds to the wet experimental results that no oxidation activity was determined. In contrast, *Sp*ALDH2 exhibits relatively high activity towards **6a**. Interaction analysis indicates that although no interaction is formed with the hydroxyl substituent of **6a**, residues A167, V170, L171, and F298 effectively stabilize the substrate through pi–pi/pi–alkyl interactions with the benzene ring of **6a**, facilitating its catalytic orientation (Fig. 6B).

*Sp*ALDH1 exhibits similar trends towards aldehyde substrates bearing chloro moiety and pyridine substitutions at different positions (such as **14a** and **24a**), wherein only substituents at the *meta*- or *para*-position resulted in higher activity. However, such patterns were not observed in the reactions catalyzed by *Sp*ALDH2. Molecular docking and the following *in situ* minimization were conducted for both sets of substrates. Concerning **14a**, the *para*-chloro substituent forms a stable halogen bond with Q467, and the benzene ring is stabilized through pi–alkyl/pi–pi interactions with F455, L166, and V162, and these similar hydrophobic interactions are retained in the case of **16a** (Fig. 7A and C). Notably, for **16a**, the *meta*-chloro substituent assists in substrate anchoring *via* alkyl interactions with M165 instead of halogen bonding. Under the cooperative effect of these interactions, the substrate is oriented toward catalytic conformation. In contrast, for **15a** (2-chlorobenzaldehyde), both the coenzyme and catalytic residue C291 engage in alkyl/π–alkyl interactions with the *meta*-chloro substituent (Fig. 7B). This interference disrupts the proper substrate anchoring, resulting in an undesirable nucleophilic attack angle of 68°. Correspondingly, *Sp*ALDH1 exhibited specific activities of 29.60 and 38.70 U mg^{-1} towards **14a** and **16a**, while activity towards **15a** is merely 0.19 U mg^{-1}.

For pyridine-substituted substrates, unexpected complex conformations were obtained. Interaction analysis indicates that the pyridine ring could be anchored through hydrophobic interactions involving V162, M165, and L166. However, regardless of the position of the nitrogen atom, it failed to establish interactions with surrounding residues (Fig. S5A–C†). This poses a challenge to explain the excellent activity of *Sp*ALDH1 towards **23a** and **24a**. Re-examination of the overall structure of *Sp*ALDH1 revealed the potential for loop 444–457 to engage with the substrate, facilitated by the flexibility of this loop. Upon loop refinement[26] (Discovery Studio 2021), the optimized loop structure exhibits a swing towards NAD, bringing it closer toward the substrate (Fig. 8A).

Then, molecular docking was conducted using the above optimized structure, followed by *in situ* ligand minimization. As expected, interaction analysis was confirmed by our hypothesis. Loop 444–457 does not interact with the pyridine ring of **22a** before and after optimization, which also explains its low catalytic activity (Fig. 8B and Fig. S5A†). In the case of **23a**, the upswing of loop 444–457 enables the nitrogen of *meta*-pyridine and the hydrogen on *para*-carbon to establish stable hydrogen bonds with G450 (Fig. 8C). In contrast, for **24a**, although *para*-nitrogen could not directly interact with G450, the electron density of the 3-position carbon is enhanced through electron conjugation. Consequently, the hydrogen atom at the 3-posi-

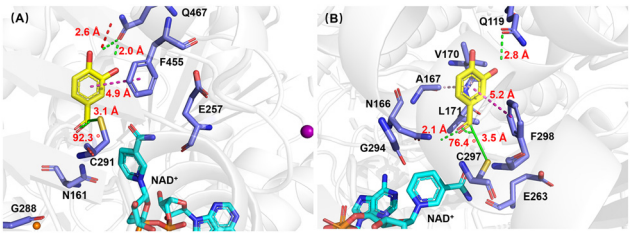


Fig. 6 Molecular docking of ALDHs with **6a**. (A) *Sp*ALDH1; (B) *Sp*ALDH2. Active site residues, substrates and NAD^+ cofactor are shown as sticks, and carbon atoms are represented by slate, yellow and cyan, respectively. ALDHs are depicted as grey. Oxygen, red; nitrogen, blue; sulfur, yellow orange; potassium ions, purple ball; sodium ions, orange ball. Hydrogen bonds: green dotted lines; pi–alkyl/alkyl interactions: yellow dotted lines; pi–pi stacking interactions: light magenta dotted line.

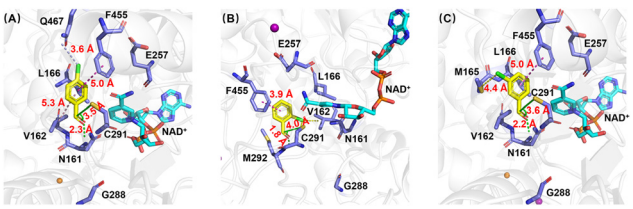


Fig. 7 Molecular docking of *Sp*ALDH1 with chlorine-substituted benzaldehydes. (A) **14a**; (B) **15a**; (C) **16a**. Active site residues, substrates and NAD^+ cofactor are shown as sticks, and carbon atoms are represented by slate, yellow and cyan, respectively. ALDHs are depicted as grey. Oxygen, red; nitrogen, blue; sulfur, yellow orange; potassium ions, purple ball; sodium ions, orange ball. Hydrogen bonds: green dotted lines; pi–alkyl/alkyl interactions: yellow dotted lines; pi–pi stacking interactions: light magenta dotted lines; halogen interactions: light blue dotted lines.

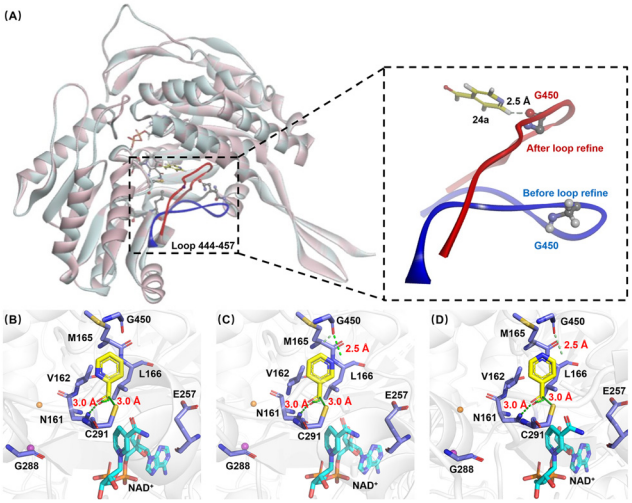


Fig. 8 Molecular docking of *Sp*ALDH1 with pyridine-formaldehydes. (A) Structure optimization of loop 444–457. Red: optimized loop 444–457, blue: loop 444–457 before optimization. (B) **22a** (2-pyridinecarboxaldehyde); (C) **23a**; (D) **24a**. Active site residues, substrates and NAD^+ cofactor are shown as sticks, and carbon atoms are represented by slate, yellow and cyan, respectively. ALDHs are depicted as grey. Oxygen, red; nitrogen, blue; sulfur, yellow orange; potassium ions, purple ball; sodium ions, orange ball. Hydrogen bonds: green dotted lines; carbon hydrogen bonds: pale green dotted lines.

tion becomes more electrophilic, contributing the hydrogen bond with G450 (Fig. 8D). These additional hydrogen-bond interactions effectively stabilize the substrate in a catalytically competent conformation, thereby facilitating the reaction.

Preparation of syringic acid by coupling ALDHs with *Sp*NOX

For preparative scale synthesis of syringic acid, it is imperative to establish an NAD^+ regeneration system to sustain the aldehyde oxidation reaction continuously. NADH oxidase emerges as a commonly used strategy, catalyzing the reduction of oxygen to either H_2O_2 or H_2O while concurrently oxidizing NADH to NAD^+.[1,27–30] Acknowledging the potential harm posed by H_2O_2 to ALDH, a water-forming NADH-oxidase derived from *Streptococcus pyogenes* was employed for NAD^+ regeneration, facilitating the preparative-scale bio-oxidation of **1a** coupled with ALDHs (Fig. S6†). *Sp*ALDH2 exhibited better catalytic efficiency than *Sp*ALDH1 (Fig. S7†). At 20 mM **1a**, 100% conversion of **1a** to **1b** was achieved by *Sp*ALDH2 in 2 h, while 51.9% and 87.4% conversions were observed for *Sp*ALDH1 at 2 h and 4 h. *Sp*ALDH2 showed better performance at an increased substrate concentration of 50 mM. After 16 h, 90.5% of **1a** was converted by *Sp*ALDH2, while 55.8% conversion was observed for *Sp*ALDH1. It is noteworthy that *Sp*ALDH1 exhibits a preference for acidic reaction conditions and tolerates metal ions such as Ni^{2+}, Ca^{2+}, Mg^{2+}, and Zn^{2+}. Despite the lower catalytic efficiency of *Sp*ALDH1 than that of *Sp*ALDH2, it may find utility in scenarios where *Sp*ALDH2 faces challenges, highlighting the diverse applicability and complementary nature of these two enzymes in bio-oxidation processes.

Conclusions

In summary, two aldehyde dehydrogenases, *Sp*ALDH1 and *Sp*ALDH2, identified from *Sphingobium* sp. strain SYK-6, exhibited significant differences in their enzymatic properties, substrate spectrum, and metal dependence, making them potential candidates for different catalytic scenarios, such as multi-enzyme cascade or chemo-enzymatic reactions. *Sp*ALDH2, exhibiting better thermostability, could catalyze the oxidation of 50 mM syringaldehyde with over 90% conversion when coupled with *Sp*NOX. Interestingly, markedly different substrate profiles were observed for two ALDHs. *Sp*ALDH1 exhibits higher substrate specificity for *meta*-methoxybenzaldehyde, while *Sp*ALDH2 shows broader substrate promiscuity toward aromatic aldehydes. Based on molecular docking and interaction analysis, the composition and polarity differences of residues at the bottom of the substrate binding pocket are mainly responsible for the distinct substrate specificity of *Sp*ALDH1 and *Sp*ALDH2. Notably, the flexible loop 444–457 could contribute to a stronger interaction between pyridine aldehyde substrates and ALDHs.

Author contributions

Siyi Chen: conceptualization, data curation, investigation, visualization, and writing – original draft. Jieyu Zhou: conceptualization, formal analysis, funding acquisition, supervision, and writing – review & editing. Xiangyuan Gu: investigation, software, and resources. Ye Ni: conceptualization, funding acquisition, resources, supervision, project administration, and writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the National Key R&D Program (2021YFC2102700), the National Natural Science Foundation of China (22377040 and 21907040), the Fundamental Research Funds for the Central Universities (JUSRP122038).

References

1 T. Knaus, V. Tseliou, L. D. Humphreys, N. S. Scrutton and F. G. Mutti, *Green Chem.*, 2018, **20**, 3931–3943.
2 S. O. Nwaukwa and P. M. Keehn, *Tetrahedron Lett.*, 1982, **23**, 3131–3134.

3 G. Tojo and M. Fernández, in oxidation of alcohols to carboxylic acids via isolated aldehydes, *Oxidation of primary alcohols to carboxylic acids*, Springer, New York, 2007, pp. 105–109.
4 J. Sedelmeier, S. V. Ley, I. R. Baxendale and M. Baumann, *Org. Lett.*, 2010, **12**, 3618–3621.
5 I. R. L. Barker and R. H. Dahm, *J. Chem. Soc. B*, 1970, 650–653.
6 E. Dalcanale and F. Montanari, *J. Org. Chem.*, 1986, **51**, 567–569.
7 K. J. Liu, Y. L. Fu, L. Y. Xie, C. Wu, W. B. He, S. Peng, Z. Wang, W. H. Bao, Z. Cao, X. Xu and W. M. He, *ACS Sustainable Chem. Eng.*, 2018, **6**, 4916–4921.
8 S. M. McKenna, S. Leimkühler, S. Herter, N. J. Turner and A. J. Carnell, *Green Chem.*, 2015, **17**, 3271–3275.
9 V. Vasiliou, D. C. Thompson and D. R. Petersen, in *Comprehensive Toxicology*, ed. C. A. McQueen, Elsevier, Jordan, second edition, 2010, aldehyde dehydrogenases, pp. 131–147.
10 H. Morrison, in *Enzyme Active Sites and their Reaction Mechanisms*, Elsevier, United States, 2021, aldehyde dehydrogenase, pp. 21–26.
11 T. Knaus, F. G. Mutti, L. D. Humphreys, N. J. Turner and N. S. Scrutton, *Org. Biomol. Chem.*, 2015, **13**, 223–233.
12 S. Wu, Y. Zhou, D. Seet and Z. Li, *Adv. Synth. Catal.*, 2017, **359**, 2132–2141.
13 A. S. Halavaty, R. L. Rich, C. Chen, J. C. Joo, G. Minasov, I. Dubrovska, J. R. Winsor, D. G. Myszka, M. Duban, L. Shuvalova, A. F. Yakunin and W. F. Anderson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2015, **71**, 1159–1175.
14 B. Ahvazi, R. Coulombe, M. Delarge, M. Vedadi, L. Zhang, E. Meighen and A. Vrielink, *Biochem. J.*, 2000, **349**, 853–861.
15 J. Zhou, X. Gu, Y. Zhu, Z. Tao and Y. Ni, *ChemBioChem*, 2023, **24**, e202300066.
16 H. B. Burgi, J. D. Dunitz, J. M. Lehn and G. Wipff, *Tetrahedron*, 1974, **30**, 1563–1572.
17 H. Rzepa, *The Winnower*, 2015, DOI: **10.15200/winn.143149.99424**.
18 E. Masai, Y. Katayama and M. Fukuda, *Biosci. Biotechnol. Biochem.*, 2007, **71**, 1–15.
19 N. Kamimura, T. Goto, K. Takahashi, D. Kasai, Y. Otsuka, M. Nakamura, Y. Katayama, M. Fukuda and E. Masai, *Sci. Rep.*, 2017, **7**, 44422.
20 J. E. Jo, S. M. Raj, C. Rathnasingh, E. Selvakumar, W. C. Jung and S. Park, *Appl. Microbiol. Biotechnol.*, 2008, **81**, 51–60.
21 R. Malik and R. E. Viola, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 673–684.
22 N. E. Rosli, M. S. M. Ali, N. H. A. Kamarudin, M. Masomian, W. Latip, S. Saadon and R. N. Z. R. A. Rahman, *Microorganisms*, 2022, **10**, 1444.
23 A. Pastore, G. Garza-Ramos, C. Mújica-Jiménez and R. A. Muñoz-Clares, *PLoS One*, 2013, **8**, e54899.
24 E. M. Valenzuela-Soto, R. Velasco-García, C. Mújica-Jiménez, L. L. Gaviria-González and R. A. Muñoz-Clares, *Chem.-Biol. Interact.*, 2003, **143–144**, 139–148.
25 C. Muñoz-Bacasehua, H. Santacruz-Ortega and E. M. Valenzuela-Soto, *Cell Biochem. Biophys.*, 2022, **80**, 39–44.
26 J. Zhou, Y. Wang, G. Xu, L. Wu, R. Han, U. Schwaneberg, Y. Rao, Y. L. Zhao, J. Zhou and Y. Ni, *J. Am. Chem. Soc.*, 2018, **140**, 12645–12654.
27 A. Švarc, Z. F. Blažević, Đ. Vasić-Rački, S. J. Charnock and A. V. Presečki, *Chem. Eng. Res. Des.*, 2020, **164**, 35–45.
28 B. Geueke, B. Riebel and W. Hummel, *Enzyme Microb. Technol.*, 2003, **32**, 205–211.
29 G. Rehn, A. T. Pedersen and J. M. Woodley, *J. Mol. Catal. B: Enzym.*, 2016, **134**, 331–339.
30 X. Y. Zhang, X. Wang, N. W. Li, Z. W. Guo, M. H. Zong and N. Li, *ChemCatChem*, 2020, **12**, 3257–3264.