WARWICK
THE UNIVERSITY OF WARWICK

Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/136753

How to cite:
Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

© 2020 Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

warwick.ac.uk/lib-publications

3,6-anhydro-L-galactose dehydrogenase *Vv*AHGD is a member of a new aldehyde dehydrogenase family and catalyzes by a novel mechanism with conformational switch of two catalytic residues cysteine 282 and glutamate 248

Ping-Yi Li[1#], Yue Wang[1#], Yi Zhang[1], Hai-Yan Cao[1], Yan-Jun Wang[1], Chun-Yang Li[2,3], Peng Wang[2], Hai-Nan Su[1], Yin Chen[4], Xiu-Lan Chen[1,3]*, Yu-Zhong Zhang[1,2,3]*

[1]State Key Laboratory of Microbial Technology, Marine Biotechnology Research Center, Shandong University, Qingdao 266237, China;

[2]College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China;

[3]Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China

[4]School of Life Sciences, University of Warwick, Coventry, CV4 7AL, United Kingdom

[#] Ping-Yi Li and Yue Wang contributed equally to this work.

*Corresponding author:

Yu-Zhong Zhang, State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, China. Tel & Fax: +86-532-58632578; E-mail: zhangyz@sdu.edu.cn;

Xiu-Lan Chen, State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, China. Tel & Fax: +86-532-58632568; E-mail: cxl0423@sdu.edu.cn

ABSTRACT

3,6-anhydro-α-L-galactose (L-AHG) is one of the main monosaccharide constituents of red macroalgae. In the recently discovered bacterial L-AHG catabolic pathway, L-AHG is firstly oxidized by an NAD(P)$^+$-dependent dehydrogenase (AHGD), which is a key step of this pathway. However, the catalytic mechanism(s) of AHGDs is still unclear. Here, we identified and characterized an AHGD from marine bacterium *Vibrio variabilis* JCM 19239 (*Vv*AHGD). The NADP$^+$-dependent *Vv*AHGD could efficiently oxidize L-AHG. Phylogenetic analysis suggested that *Vv*AHGD and its homologs represent a new aldehyde dehydrogenase (ALDH) family with different substrate preferences from reported ALDH families, named the L-AHGDH family. To explain the catalytic mechanism of *Vv*AHGD, we solved the structures of *Vv*AHGD in the apo form and in complex with NADP$^+$ and modeled its structure with L-AHG. Based on structural, mutational, and biochemical analyses, the cofactor channel and the substrate channel of *Vv*AHGD are identified and the key residues involved in the binding of NADP$^+$ and L-AHG and in the catalysis are revealed. *Vv*AHGD performs catalysis by controlling the consecutive connection and interruption of the cofactor channel and the substrate channel via the conformational changes of its two catalytic residues Cys282 and Glu248. Comparative analyses of structures and enzyme kinetics revealed that differences in the substrate channels (in shape, size, electrostatic surface and residue composition) lead to the different substrate preferences of *Vv*AHGD from other ALDHs. This study on *Vv*AHGD sheds light on the diversified catalytic mechanisms and evolution of NAD(P)$^+$-dependent ALDHs.

Keywords: Red macroalgae; 3,6-anhydro-L-galactose; Aldehyde dehydrogenase; Catalysis; Substrate binding

Abbreviations: L-AHG, 3,6-anhydro-α-L-galactose; GH, glycoside hydrolase; AHGD, L-AHG dehydrogenase; L-AHGA, 3,6-L-anhydrogalactonate; ALDH, aldehyde dehydrogenase; PDB, Protein Data Bank; *Vv*AHGD-NADP, *Vv*AHGD complexed with NADP$^+$; DLS, dynamic light scattering; CD, circular dichroism; GC-MS, gas chromatography-mass spectrometry.

INTRODUCTION

Marine algae generate approximately a half of the global primary production (1). Compared to terrestrial biomass, algae contain higher carbohydrates and lower lignin, thus making them a good target for the production of biofuels and biochemicals (2,3). In red macroalgae, the most abundant polysaccharides are agarose and carrageenan. Agarose is a linear polymer with repeating disaccharide subunits composed of 3-linked β-D-galactose and 4-linked 3,6-anhydro-α-L-galactose (L-AHG) (4). L-AHG is a recalcitrant bicyclic sugar, which is unique to red macroalgae (4). Marine bacteria play an important role in the degradation and cycling of algal polysaccharides in the ocean. Many bacterial enzymes have been reported to be involved in the degradation of agarose into its monomeric units, including agarases from the hallmark glycoside hydrolase (GH) families GH50, GH86 and GH117 (5-7). In contrast, there are only a few of marine bacteria reported to be capable of utilizing the L-AHG monosaccharides, including *Postechiella marina* M091 (8), *Vibrio* sp. strain EJY3 (9) and *Raoultella ornithinolytica* B6-JMP12 (10).

Recently, the L-AHG catabolic pathway is found in marine bacteria *P. marina* M091 (8) and *Vibrio* sp. strain EJY3 (9). L-AHG is firstly oxidized to 3,6-L-anhydrogalactonate (L-AHGA) by an $NAD(P)^+$-dependent L-AHG dehydrogenase (AHGD, EC 1.2.1.92) before its assimilation into the central metabolism (8,9). The oxidation of L-AHG is a key step due to its initiation of the L-AHG pathway. Up to date, only several AHGDs are biochemically characterized, including the M091_0723 from *P. marina* M091 (8), the Patl_2553 from *Pseudoalteromonas atlantica* T6c (8), and the AHGD from *Vibrio* sp. strain EJY3 (9). These three enzymes share high sequence identities ranging from 56% to 66%. Among the three enzymes, the AHGD from *Vibrio* sp. strain EJY3 (*Vej*AHGD) is characterized in detail (11,12). The recombinant *Vej*AHGD displays high substrate specificity to L-AHG, but hardly oxidizes galactose, D-AHG, L-lactaldehyde or other aldehydes (11). The kinetic parameters of *Vej*AHGD against L-AHG are also unaffected by the presence of other sugars such as galactose, neoagarobiose, and agaro-oligosaccharides (12). In addition, an AHGD from *R. ornithinolytica* B6-JMP12 is also reported (10), the sequence of

which, however, has not been determined.

AHGDs belong to the $NAD(P)^+$-dependent aldehyde dehydrogenase (ALDH) superfamily. ALDHs play a key role in the metabolism of endogenous and exogenous aldehydes by oxidizing aldehydes to corresponding carboxylic acids with $NAD(P)^+$ as an electron acceptor (13). ALDHs exist in all three domains of life. Based on sequence identities, ALDHs have been classified into more than 19 families with different substrate preferences (14). Until now, many apo and holo structures of ALDHs from different families have been reported. ALDHs generally function as homodimers or homotetramers, with each monomer containing a catalytic domain, a cofactor binding domain, and an oligomerization domain within their structures (13). Monomeric ALDHs contain two channels in opposite sites, a cofactor channel and a substrate channel, and the active site is located at the junction of these two channels (15,16). ALDHs utilize a mechanism involving the nucleophilic residue Cys and the general base Glu (17,18). During the catalytic process, these two catalytic residues usually undergo conformational changes, which play a key role in the association/release of the cofactor/substrate/product (19-21). While the substrate channel and the cofactor channel in most ALDHs are always connected throughout the catalytic cycle (22-24), the binding of cofactor and/or substrate in some ALDHs results in the two separate channels to be connected due to the conformational change of the catalytic residue Cys (25). In addition to the catalytic residues, the conformational changes of non-catalytic residues in the substrate channel upon the binding of NAD^+ also induce the connection of two separate channels in the betaine aldehyde dehydrogenase from *Staphylococcus aureus* (21). For AHGDs, only the structure of the apo Patl_2553 was deposited in the Protein Data Bank (PDB) by the New York SGX Research Center for Structural Genomics team, and this structure was preliminarily analyzed by Sun et al. (26). However, the catalytic mechanism(s) of AHGDs is still unclear. In addition, the AHGDs so far reported share relatively high sequence identities (43%-44%) with the lactaldehyde dehydrogenase from *Escherichia coli* (*Ec*ALDH for short) (24), but have quite different substrate specificity from *Ec*ALDH. The structural basis for their different substrate specificities is still unknown.

In this study, the crystal structures of an AHGD from marine bacterium *Vibrio variabilis* JCM 19239 (*Vv*AHGD) are solved in the apo form and in complex with $NADP^+$. Based on structural and mutational analyses combined with substrate docking, the catalytic mechanism of *Vv*AHGD is explained. Moreover, via detailed comparative analyses of structures and enzyme kinetics, the structural basis for the substrate recognition of *Vv*AHGD is illustrated. Based on our results combined with phylogenetic analysis, *Vv*AHGD and its homologs are suggested to represent a new ALDH family.

RESULTS AND DISCUSSION

*Vv*AHGD is an $NADP^+$-dependent L-AHG dehydrogenase

Based on blasting analysis, a gene encoding a putative AHGD (GenBank accession No. GAL27243) was identified from the genome sequence of *Vibrio variabilis* JCM 19239, which was named *VvAHGD*. The *Vv*AHGD protein consists of 480 amino acid residues with a calculated molecular mass of 52.6 kDa. *Vv*AHGD lacks an N-terminal signal peptide sequence based on SignalP 4.1 prediction, consistent with its intracellular location. Among the reported AHGDs, *Vv*AHGD shares the highest sequence identity (79%) to *Vej*AHGD (9) and the lowest identity (57%) to Patl_2553 (8). Multiple sequence alignment suggested that the catalytic residues of *Vv*AHGD are Cys282 and Glu248 (Fig. 1).

*Vv*AHGD was then over-expressed in *E. coli* BL21(DE3) and purified. Because no commercial L-AHG is available, the crude extract of L-AHG is prepared from neoagarobioses and used as substrate for enzyme activity assay of recombinant *Vv*AHGD (Fig. 2). *Vv*AHGD could oxidize L-AHG using both $NADP^+$ and NAD^+ as the cofactor (Fig. 2b). Like *Vej*AHGD (9), *Vv*AHGD preferred $NADP^+$, showing an activity 3.6-fold higher than that with NAD^+. Among the known AHGDs, *Vv*AHGD showed the highest activity towards L-AHG with a specific activity of 57.4 U/mg (Fig. 2b), approximately 4.0-fold higher than that of *Vej*AHGD (Fig. 2c). Similar to *Vej*AHGD (11), *Vv*AHGD hardly oxidized lactaldehyde, D-AHG, galactose or glucose. The optimal temperature for *Vv*AHGD activity towards L-AHG was 40°C (Fig. 3a) and the optimal pH was 7.0 (Fig. 3b). The intracellular *Vv*AHGD showed bad salt tolerance

(Fig. 3c).

Overall structural analysis of *Vv*AHGD

To study the catalytic mechanism of *Vv*AHGD, we solved the crystal structures of *Vv*AHGD in the apo form (2.70 Å resolution) and in complex with $NADP^+$ (2.37 Å resolution), by molecular replacement using the structure of Patl_2553 (PDB code 3K2W) as the starting model (56% sequence identity). The statistics for refinement are summarized in Table 1. Crystals of the apo *Vv*AHGD belong to the $P2_12_12_1$ space group, with four molecules per asymmetric unit (Fig. 4a). Crystals of *Vv*AHGD complexed with $NADP^+$ (*Vv*AHGD-NADP) belong to the $P22_12_1$ space group, with two molecules per asymmetric unit. Gel filtration analysis showed that the apo *Vv*AHGD and the *Vv*AHGD-NADP complex tend to form large oligomers in solution and that the binding of $NADP^+$ has no effect on the oligomerization state of *Vv*AHGD (Fig. S1). Gel filtration analysis is applicable in measuring the oligomerization states of globular proteins but not for nonglobular proteins. Therefore, dynamic light scattering (DLS) analysis was conducted to further determine the oligomerization state of *Vv*AHGD, which suggested that both the apo *Vv*AHGD and its complex form stable dimers in solution (Fig. 4b). The overall structure of *Vv*AHGD monomer is similar to those of other ALDHs (Fig. 4c), most closely resembling the structures of Patl_2553 and *Ec*ALDH (PDB code 2HG2), with the root mean square deviations of 0.60 Å (421 monomer Cα atoms) and 0.87 Å (434 monomer Cα atoms), respectively. Like other ALDHs, monomeric *Vv*AHGD contains three domains, a catalytic domain (Gly250-Gly440), a cofactor binding domain (Gln3-Ala116, His441-Tyr469, and Arg140-Leu249 in the Rossmann fold), and an oligomerization domain (Arg117-Pro139 and Leu470-Tyr479) (Fig. 4d). Both the catalytic domain and the cofactor binding domain are built on a topologically related βαβ fold. The catalytic residues Cys282 and Glu248 are located in the catalytic domain and the cofactor binding domain, respectively (Fig. 4d). The oligomerization domain is composed of a three-stranded antiparallel β-sheet, mainly contributing to the formation of the dimerization surface. The overall structure of the *Vv*AHGD-NADP complex is similar to that of the apo enzyme with a root mean square deviation of 0.33 Å for 436 monomer Cα atoms.

The $NADP^+$-binding mode in *Vv*AHGD

In the crystal structure of the *Vv*AHGD-NADP complex, one molecule of $NADP^+$ is bound in the active site of each monomer, occupying the cofactor channel (Fig. 5a and 5b). The substrate channel is on the opposite side of the cofactor channel. The catalytic residues Cys282 and Glu248 are located at the junction of these two channels. These two channels are disconnected in the apo structure of *Vv*AHGD, but connected in the structure of the *Vv*AHGD-NADP complex (Fig. 5c). The cofactor channel has a wide opening on the protein surface, which is mostly positively charged to hold the adenosine monophosphate (AMP) moiety and the pyrophosphate group of $NADP^+$ (Fig. 5a). The internal part of this channel is negatively charged to accommodate the nicotinamide riboside moiety of $NADP^+$ (Fig. 5a). In the complex, the cofactor $NADP^+$ adopts an extended conformation, which is typical for oxidized $NADP^+$ (Fig. 5b). Among the reported structures of ALDHs, the bound cofactors are found to have three different conformations, the hydride transfer (27,28), the hydrolysis (22,24) and the out conformation (24,29). Structural superposition with other ALDHs reveals that the $NADP^+$ bound in the structure of *Vv*AHGD adopts the hydride transfer conformation.

A detailed comparative structural analysis indicates that residues in the cofactor channel and the substrate channel have similar conformations in both the apo enzyme and the complex, except for the two catalytic residues Cys282 and Glu248 (Fig. 5d and 5e). Upon the binding of the cofactor, the nucleophilic Cys282 of *Vv*AHGD is reoriented towards the substrate channel to be in the attacking conformation (Fig. 5d), and the side chain of the catalytic Glu248 moves away from the active site to leave room for the nicotinamide ring of $NADP^+$ (Fig. 5d). The conformational changes of these two catalytic residues result in the connection of the cofactor channel and the substrate channel in *Vv*AHGD, which is essential for the nicotinamide ring of $NADP^+$ to be close to the substrate in the hydride transfer state. Mutation of each of these two catalytic residues to Ala inactivated the enzyme completely (Fig. 6), supporting their key roles in the catalysis of *Vv*AHGD.

In the *Vv*AHGD-NADP complex, $NADP^+$ is stabilized mainly by hydrogen bond interactions (Fig. 5e). For the AMP moiety of $NADP^+$, the adenine base forms hydrogen

bonds with the hydroxyl group of Ser233 and the main-chain CO group of Gly206, the hydroxyl group of the ribose ring is hydrogen bonded to the side chain of Lys173 and the main-chain CO group of Thr147, and the phosphate oxygen atoms are hydrogen bonded to the side chains of Lys173 and Ser176 and the main-chain NH group of Gly206. The pyrophosphate group of $NADP^+$ is hydrogen bonded to the hydroxyl group and the main-chain NH group of Ser227 and to the side chain and the main-chain NH group of Trp149. For the nicotinamide riboside moiety of $NADP^+$, the hydroxyl groups of the ribose ring form hydrogen bonds with the carboxyl group of Glu383, and the nicotinamide ring are hydrogen bonded to the side chain of His449 and the main-chain CO group of Leu249. Except for residues Ser233 and Leu249, mutation of all the other residues above to Ala led to a complete loss or significant reduction of the *Vv*AHGD activity (Fig. 6), suggesting that these residues play key roles in the binding of $NADP^+$. Residues Ser233 and Leu249 possibly contribute little to the binding of $NADP^+$, because mutation of these two residues to Ala had no or small impact on both the specific activity and the kinetic parameters of *Vv*AHGD (Fig. 6 and Table 2). Mutations S176A and S227A dramatically reduced the k_{cat} values of *Vv*AHGD against both $NADP^+$ and L-AHG and its affinity to $NADP^+$, but had little impact on its affinity to L-AHG (Table 2), further confirming that residues Ser176 and Ser227 play roles in the catalysis of *Vv*AHGD through interacting with the cofactor. In addition to hydrogen bond interactions, hydrophobic interactions also contribute to the binding of the cofactor involving residues such as Ile146, Gly210, Val214, Met224 and Ile234.

The substrate-binding mode in *Vv*AHGD

To investigate the substrate-binding mode in *Vv*AHGD, we tried to obtain the structure of *Vv*AHGD binding a substrate. We cocrystallized the wild-type *Vv*AHGD and its inactive mutant E248A in the presence of $NADP^+$ and the prepared L-AHG. Unfortunately, no electron density was observed for the substrate. Under aqueous condition, L-AHG was found to exist in three forms, the open-chain aldehyde, the open-chain hydrate and the fused-ring pyranose (11). Therefore, by molecular docking, we modeled the structures of the *Vv*AHGD-NADP complex bound with L-AHG in three different forms. In all the modeled structures, L-AHG is bound in the substrate channel

(Figs. 6a and S2). In this channel, the aldehyde carbonyl of L-AHG in any form is close to the thiol group of the catalytic Cys282 and the nicotinamide ring of $NADP^+$ (Figs. 6a and S2), which is important for the hydride transfer from the substrate to the nicotinamide ring of $NADP^+$. L-AHG in all three forms is hydrogen bonded to the side chain of Glu443 (Figs. 6b and S2). Mutation of this residue to Ala abolished the enzymatic activity completely, indicating the key role of Glu443 in the binding of L-AHG (Fig. 6c). Circular dichroism (CD) spectroscopy analysis showed that the secondary structures of the mutants exhibit little deviation from that of wild-type *Vv*AHGD (Fig. 6d), indicating that the changes in enzymatic activity and kinetic parameters of the mutants result from residue substitution rather than structural changes.

The catalytic mechanism of *Vv*AHGD

Based on our structural and biochemical results on *Vv*AHGD combined with reported catalytic mechanisms of other ALDHs (13,30), we propose the mechanism for *Vv*AHGD to catalyze the oxidation of L-AHG (Fig. 7). The proposed catalytic mechanism of *Vv*AHGD involves four steps. Firstly, in the absence of cofactor and substrate, *Vv*AHGD is in the resting state, in which the catalytic residues Cys282 and Glu248 adopt the apo conformation, and the cofactor channel and the substrate channel are disconnected due to the blocking by residues Cys282 and Glu248. Secondly, when the cofactor and the substrate enter their respective channels, the cofactor channel and the substrate channel are connected due to the conformational changes of Cys282 and Glu248. The enzyme is in the hydride transfer state. Glu248 moves away from the active site (the hydride transfer conformation) to leave room for the accommodation of the nicotinamide ring of $NADP^+$, and Cys282 is oriented towards the substrate channel (the attacking conformation), so that the nucleophilic Cys282 could attack the aldehyde carbonyl of L-AHG to form a tetrahedral intermediate with concomitant hydride transfer from the intermediate to the nicotinamide ring of $NADP^+$. The binding of the cofactor leads to the conformational changes of these two catalytic residues, which is required for the nicotinamide ring of $NADP^+$ to be close to both the substrate and the catalytic Cys282 in this state. Thirdly, the resulting NADPH leaves the active site for the catalytic Glu248 moving back to adopt the apo-like conformation. Then Glu248

acts as a general base to activate a water molecule to attack the thioester intermediate. The enzyme is in the hydrolysis state. The reposition is important for Glu248 to be situated in an appropriate distance from the substrate and Cys282 to catalyze the deacylation process. In this state, the cofactor channel and the substrate channel are still connected, but the opening between these two channels possibly becomes narrow due to the conformational change of Glu248. Lastly, NADPH and the product L-AHGA are released from the enzyme. Cys282 is reoriented in the apo conformation resulting in the interruption of the cofactor channel and the substrate channel. Therefore, *Vv*AHGD is back to the resting state, ready for the next catalytic cycle (Fig. 7).

During the catalytic process of reported ALDHs, it is requisite for the catalytic residues Cys and Glu to transform between different conformations (19-22). In *Ec*ALDH (24) and most other ALDHs (22,23), the substrate channel and the cofactor channel are always connected throughout the catalytic cycle, which is uninfluenced by the conformational changes of their catalytic residues. In the succinic semialdehyde dehydrogenase PpALDH21 from *Physcomitrella patens* (25) and the phosphonoacetaldehyde dehydrogenase PhnY from *Sinorhizobium meliloti* (31), the distantly related homologs of *Vv*AHGD with sequence identities below 30%, their two channels in the apo enzymes are found to be separated by the catalytic Cys whose conformational change upon the binding of the cofactor leads to the connection of the two separate channels. However, in the catalytic process of *Vv*AHGD, both the two catalytic residues, Cys282 and Glu248, not only undergo conformational changes, but also function as gatekeepers between the cofactor channel and the substrate channel. In *Vv*AHGD, the conformational changes of Cys282 and Glu248 lead to the connection and interruption of the cofactor channel and the substrate channel, which promotes the productive binding of $NADP^+$/L-AHG and the efficient release of NADPH and L-AHGA during the catalysis and therefore leads to the high catalytic activity of *Vv*AHGD. Notably, the two channels in the apo Patl_2553 are connected, suggesting that the catalytic process of Patl_2553 is different from *Vv*AHGD and that this difference may result in the enzymatic activity of Patl_2553 different from that of *Vv*AHGD. In summary, *Vv*AHGD performs catalysis by controlling the consecutive

connection and interruption of the cofactor channel and the substrate channel through the conformational changes of its two catalytic residues Cys282 and Glu248.

Phylogenetic analysis of L-AHG dehydrogenases

To reveal the evolutionary relationship between *Vv*AHGD and other ALDHs, a phylogenetic tree was constructed including *Vv*AHGD and its homologs, lactaldehyde dehydrogenases, succinic semialdehyde dehydrogenases from family ALDH5, betaine aldehyde dehydrogenases from family ALDH25, cytosolic aldehyde dehydrogenases from family ALDH2, and retinaldehyde dehydrogenases from family ALDH1 (Fig. 8). The sequence identities between *Vv*AHGD-like sequences and other ALDHs are in a range of 33% to 44%. Except for the ALDH2 family cytosolic aldehyde dehydrogenases with a wide substrate specificity (32), *Vv*AHGD and other ALDHs (24,33-35) show a relatively narrow substrate specificity. As shown in the tree, different ALDH branches in the phylogenetic tree display different substrate preferences. *Vv*AHGD and its homologs form a separate branch in the phylogenetic tree, as a sister group of lactaldehyde dehydrogenases, indicating that *Vv*AHGD and its homologs represent a new ALDH family with a different substrate specificity from other ALDHs, which is named the L-AHGDH family in this study.

Structure comparison of aldehyde dehydrogenases

*Vv*AHGD and its homologs are similar to other ALDHs in both sequence (Fig. 1) and overall structures (Fig. 4c). However, *Vv*AHGD and other ALDHs display different substrate preferences towards aldehydes. To reveal the structural basis for their different substrate specificities, a detailed comparative structural analysis of the active sites involving the cofactor channel and the substrate channel was carried out. No matter whether ALDHs utilize $NADP^+$ or NAD^+ as a cofactor, their cofactor channels are similar in shape and size (Fig. 9a). Although the cofactor channels of *Vv*AHGD and *Ec*ALDH are coated by a larger patch of positively charged surface than those of the ALDH5 and ALDH25 family enzymes, most of the key residues in *Vv*AHGD involved in the binding of the cofactor are highly conserved in these ALDHs (Fig. 9b and 9c), which is consistent with a previous report on ALDHs from families ALDH2 and ALDH7 (32). However, Ser176 in *Vv*AHGD close to the phosphate group of the AMP

moiety of $NADP^+$ is replaced by Glu and other residues with a larger side chain in other ALDHs using NAD^+ as a cofactor. It has been documented that the residue Glu at this position sterically hinders the binding of the phosphate group of $NADP^+$ in the plant ALDH3 family enzymes (36). Compared to the bound NAD^+ in the ALDH25 family enzyme, the adenine moiety of $NADP^+$ in *Vv*AHGD is closer to the residue Ser233 in space due to the presence of an additional phosphate group (Fig. 9b). The counterparts of Ser233 in *Vv*AHGD are larger residues in NAD^+-dependent ALDHs such as Lys, Ile or His (Fig. 9b), suggesting that this residue might also be involved in determining the cofactor specificity of ALDHs. Based on the residue types of these two positions (Fig. 9c), we can hypothesize that lactaldehyde dehydrogenases and other ALDHs from the ALDH5 and ALDH25 families are dominated by NAD^+-dependent ALDHs, whereas a large proportion of AHGDs tend to prefer $NADP^+$ as a cofactor.

In contrast to the conserved cofactor channels, the substrate channels are quite different in shape, size and electrostatic surface for ALDHs from different families (Fig. 10). Among the ALDHs analyzed, retinaldehyde dehydrogenases in the ALDH1 family harbor the largest substrate channels to hold the largest aldehyde, retinaldehyde (33,37). The ALDH2 family enzymes have a wide substrate specificity covering aliphatic and aromatic aldehydes (32), and therefore their substrate channels are not discussed here. The substrate channels of succinic semialdehyde dehydrogenases in the ALDH5 family are positively charged to accommodate negatively charged succinic semialdehydes, whereas those of betaine aldehyde dehydrogenases in the ALDH25 family are negatively charged to accommodate positively charged betaine aldehydes (Fig. 10a). Different from the charged substrate channels of the ALDH5 and ALDH25 enzymes, *Vv*AHGD and *Ec*ALDH have approximately neutral substrate channels (Fig. 10a). *Vv*AHGD has a large substrate channel with 26 residues lining on the channel surface, while *Ec*ALDH has a small substrate channel with only 17 residues on the channel surface (Fig. 10b). In *Ec*ALDH, the lactate product is stabilized through forming hydrogen bonds with residues Arg161, Glu251, Glu443 and His449 (24), which correspond to residues Arg158, Glu248, Glu443 and His449 in *Vv*AHGD (Fig. 10c). These residues are conserved in *Vv*AHGD-like and *Ec*ALDH-like sequences (Fig. 10d).

Residue Glu443 has been shown to be involved in the binding of L-AHG in *Vv*AHGD (Fig. 6). Although key residues involved in the substrate/product binding are conserved in L-AHG dehydrogenases and lactaldehyde dehydrogenases, *Vv*AHGD shows a limited activity towards lactaldehyde and *vice versa* (Fig. 2b and 2d). The substrate channel of *Ec*ALDH contains several conserved residues with large side chains including Leu100, Phe107 and Phe442 at the channel mouth, and Phe157 and Asn286 at the channel neck (Fig. 10c and 10d), which make the channels of *Ec*ALDH-like ALDHs too narrow for the entry of larger substrates such as L-AHG to the active site. In contrast, *Vv*AHGD is occupied by small residues corresponding to Val97, Ala104, Gly442, Ala154, and Thr283 at the equivalent positions of the substrate channel (Fig. 10c), which are conserved in *Vv*AHGD homologs (Fig. 10d), and therefore, *Vv*AHGD and its homologs have a channel with a wide opening to allow the entry of L-AHG. The k_{cat} value (0.74 ± 0.02 s^{-1}) of *Vv*AHGD and its affinity to L-lactaldehyde (K_m of 3.74 ± 0.22 mM) were much lower than those for L-AHG (Table 2), suggesting that the large substrate channel of *Vv*AHGD is unfavorable for the binding of the small substrate L-lactaldehyde and therefore the enzyme activity towards L-lactaldehyde is severely decreased.

CONCLUSION

*Vv*AHGD and its homologs form a separate group in the ALDH superfamily based on phylogenetic analysis. Biochemical characterization showed that *Vv*AHGD could efficiently oxidize L-AHG by using $NADP^+$ as cofactor. Structural and mutational analysis revealed key residues involved in the binding of the cofactor and the substrate L-AHG and in the catalysis in *Vv*AHGD. During the catalytic process, *Vv*AHGD performs catalysis by controlling the consecutive connection and interruption of the cofactor channel and the substrate channel via the conformational changes of its two catalytic residues Cys282 and Glu248. Differences in the substrate channels (in shape, size, electrostatic surface and residue composition) lead to the different substrate preferences of *Vv*AHGD from other ALDHs. Based on our results, we suggest AHGDs to be a new family of the ALDH superfamily, named the L-AHGDH family. This study reveals the structural basis for catalysis and substrate recognition of a L-AHG

dehydrogenase, which provides a better understanding of bacterial AHGDs in the degradation and cycling of algal polysaccharides in the ocean.

EXPERIMENTAL PROCEDURES

Gene synthesis and mutagenesis

The genes encoding *Vv*AHGD (GenBank accession No. GAL27243), *Vej*AHGD (GenBank accession No. WP_014232205), and *Ec*ALDH (GenBank accession No. P25553) were synthesized by Sangon (Shanghai) co., Ltd, and cloned into the expression vector pET22b, respectively. To study the roles of some residues in catalysis, the codons for all the selected residues of *Vv*AHGD were individually replaced with that for Ala. All of the point mutations in *Vv*AHGD were introduced by a QuikChange^{TM} site-directed mutagenesis method (38) using plasmid pET22b-*VvAHGD* as the template. All of the recombinant plasmids were verified by DNA sequencing.

Protein expression and purification

All the recombinant plasmids were transformed into *E. coli* BL21 (DE3), respectively. *E. coli* cells were cultured at 37°C in Luria-Bertani medium to OD_{600} at 0.6, and then the expression of all recombinant proteins in the cells was induced at 20°C by the addition of 1 mM isopropyl-D-thiogalactopyranoside. After 20 h induction, cells were collected and disrupted by high pressure cell cracker (JNBIO) in a binding buffer (50 mM Tris-HCl buffer pH 8.0, 100 mM NaCl and 5 mM imidazole). The recombinant proteins in the resulting extract were first purified by nickel-nitrilotriacetic acid resin (Qiagen), and then by gel filtration chromatography on a Superdex- 200 column (GE Healthcare) with 10 mM Tris-HCl buffer (pH 8.0) containing 100 mM NaCl. Protein concentrations were determined by Pierce BCA Protein Assay Kit (Thermo Scientific, USA).

L-AHG preparation

In bacterial agar degradation pathway, NABH is a key enzyme that degrades neoagarobiose into L-AHG and D-galactose (39). The gene of NABH was cloned into the expression vector pET22b. The expression and purification of protein NABH was

done by the same method as for *Vv*AHGD. To prepare L-AHG, 1.5 mM neoagarobiose was hydrolyzed by 5 nM NABH in 100 mL tri-distilled water at 30℃ and 180 rpm for 2 h. After reaction, the mixture was centrifuged at 8228 g for 20 min, and the resulting supernatant was filtrated in a 5 kDa MWCO ultrafiltration centrifuge tube (Sartorius). The filtration was then freeze dried, which was used as the crude extract of L-AHG. Thin layer chromatography analysis of the crude extract of L-AHG was performed by following the method described by Yun et al. (40). The quantitation of L-AHG in the crude extract was determined by gas chromatography-mass spectrometry (GC-MS) using D-AHG (Sigma) as the standard. The derivatization of samples and the implementation of GC-MS was performed by following the method described by Yun et al. (40).

Biochemical characterization

The crude extract of L-AHG was used as the substrate for enzymatic activity assays. The standard reaction system (1 mL) contained 0.99 nM *Vv*AHGD, 3.52 mM L-AHG and 1.5 mM $NADP^+$ in 20 mM Tris-HCl (pH 7.0). After incubation at 40°C for 20 min, the reaction was terminated by heating at 95°C for 1 min, and then the absorbance of the reaction mixture was measured at 340 nm. One unit of enzyme (U) is defined as the amount of enzyme required to produce 1 μmol NADPH per min. The optimum temperature was measured in the temperature range of 0℃ to 60℃ at pH 7.0. The optimum pH of *Vv*AHGD was measured at 40℃ in Britton-Robinson buffers ranging from pH 5.0 to pH 10.0. The effect of NaCl on *Vv*AHGD activity was determined at NaCl concentrations ranging from 0 to 2.0 M under optimal conditions. L-AHG, L-lactaldehyde, D-AHG, D-galactose and glucose were used as substrates for the enzyme specificity assays under different cofactors. Enzyme kinetic assays of *Vv*AHGD and its mutants against $NADP^+$ were carried out in 20 mM Tris-HCl buffer (pH 7.0) using $NADP^+$ at concentrations from 0 to 4.8 mM and L-AHG at a fixed concentration of 3.52 mM. Enzyme kinetic assays of *Vv*AHGD and its mutants against L-AHG were carried out using L-AHG at concentrations from 0 to 4.4 mM and $NADP^+$ at a fixed concentration of 1.5 mM. Kinetic parameters were calculated by nonlinear regression fit directly to the Michaelis-Menten equation using the Origin8.5 software.

The activities of *Vej*AHGD and *Ec*ALDH towards L-AHG or L-lactaldehyde were assayed with different cofactors, at 30℃ in 20 mM Tris-HCl (pH 7.0) buffer (11) and at 25℃ in 20 mM Tris-HCl (pH 8.0) buffer (26,41), respectively.

DLS and CD spectroscopy assays

The DLS experiment was carried out on Dynapro Titan TC (Wyatt Technology, America) at 4℃. The *Vv*AHGD protein concentration used was 1 mg/mL in 10 mM Tris-HCl (pH 8.0) buffer containing 100 mM NaCl, and the data were analyzed by the dynamics 7. 0. 1 software. The secondary structures of *Vv*AHGD and its mutants were detected at 25°C using a J-810 CD spectropolarimeter (JASCO). CD spectra were collected from 200 to 250 nm at a scanning rate of 200 nm/min with a path length of 0.1 cm. The concentration of all proteins was 0.5 mg/mL in 10 mM Tris-HCl buffer (pH 8.0) containing 100 mM NaCl.

Crystallization, data collection, and structure determination

The protein concentration of *Vv*AHGD for crystallization was 5.0 mg/ml in 10 mM Tris-HCl (pH 8.0) containing 100 mM NaCl. To obtain the crystals of the *Vv*AHGD-NADP complex, *Vv*AHGD was mixed with $NADP^+$ at a molar ratio of 1: 4 and then incubated at 4°C for 1 h before crystallization. The crystals of both the apo *Vv*AHGD and the *Vv*AHGD-NADP complex suitable for x-ray diffraction were obtained at 4°C after 2 weeks in the buffer containing 0.2 M $NaAC{\cdot}3H_2O$, 0.1 M sodium cacodylate trihydrate (pH 6.5), and 30% (w/v) PEG 8,000 by the hanging-drop vapor diffusion method. All the X-ray diffraction data of crystals were collected on the BL18U1 beamline at Shanghai Synchrotron Radiation Facility. The initial diffraction data sets were integrated by the HKL3000 program (42). The crystal structure of *Vv*AHGD was solved by molecular replacement using the CCP4 suite, with the structure of Patl_2553 (PDB code 3K2W) as the staring model. Subsequent refinement was performed using Coot (43) and Phenix (44). Structure of the *Vv*AHGD-NADP complex was determined using *Vv*AHGD as a staring model. All structure figures were generated using PyMOL.

Molecular docking

Schrödinger software (https://www.schrodinger.com/) was used to conduct the *Vv*AHGD-NADP complex and L-AHG docking. L-AHG was preprocessed by the

LigPrep (LigPrep, Schrödinger, LLC, New York, NY, 2019) to obtain its low energy three dimensional conformers. The crystal structure of the *Vv*AHGD-NADP complex was optimized by the Protein Preparation Wizard in Maestro to remove crystallographic water molecules, add hydrogen atoms, and assign partial charges and protonation states. L-AHG was then docked into the binding site of the minimised *Vv*AHGD-NADP complex using the Glide (45) with the standard precision scoring mode. The docking grid box for L-AHG was generated using the binding site identified by the SiteMap. In molecular docking, the candidate solution with the lowest binding energy was chosen.

Accession codes

The structures of the apo *Vv*AHGD and the *Vv*AHGD-NADP complex have been deposited in PDB under the accession numbers 6J75 and 6J76, respectively.

Acknowledgements

We thank the staffs from BL18U1 & BL19U1 beamlines of National Facility for Protein Sciences Shanghai (NFPS) and Shanghai Synchrotron Radiation Facility, for assistance during data collection. This work was supported by the National Key Research and Development Program of China (2018YFC1406700 and 2018YFC0310704), the National Science Foundation of China (grants 91851205, 31870052, U1706207, 91751101, 31728001, 31670038, 41676180, and 31630012), the Program of Shandong for Taishan Scholars (tspd20181203), AoShan Talents Cultivation Program Supported by Qingdao National Laboratory for Marine Science and Technology (2017ASTCP-OS14 and QNLM2016ORP0310), and the Young Scholars Program of Shandong University (2017WLJH57).

Conflict of interest

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

Y-Z.Z. and X.C. designed the research. X.C. and P.L. directed the experiments. Y.W., P.L. and Y.Z performed all experiments. P.L. and H.C. solved the structures. Y-J.W., P.W. and C.L. helped in data analysis. P.L. and Y.W. wrote the manuscript. X.C., Y.C. and H.S. edited the manuscript.

REFERENCES

1. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281, 237-240.
2. Wargacki, A. J., Leonard, E., Win, M. N., Regitsky, D. D., Santos, C. N., Kim, P. B., Cooper, S. R., Raisner, R. M., Herman, A., Sivitz, A. B., Lakshmanaswamy, A., Kashiyama, Y., Baker, D. & Yoshikuni, Y. (2012). An engineered microbial platform for direct biofuel production from brown macroalgae. Science 335, 308-313.
3. Yun, E. J., Choi, I. G. & Kim, K. H. (2015). Red macroalgae as a sustainable resource for bio-based products. Trends. Biotechnol. 33, 247-249.
4. Duckworth, M. & Yaphe, W. (1971). The structure of agar: Part I. Fractionation of a complex mixture of polysaccharides. Carbohydr. Res. 16, 189-197.
5. Hehemann, J. H., Boraston, A. B. & Czjzek, M. (2014). A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. Curr. Opin. Struct. Biol. 28, 77-86.
6. Yun, E. J., Yu, S. & Kim, K. H. (2017). Current knowledge on agarolytic enzymes and the industrial potential of agar-derived sugars. Appl. Microbiol. Biotechnol. 101, 5581-5589.
7. Pluvinage, B., Grondin, J. M., Amundsen, C., Klassen, L., Moote, P. E., Xiao, Y., Thomas, D., Pudlo, N. A., Anele, A., Martens, E. C., Inglis, G. D., Uwiera, R. E. R., Boraston, A. B. & Abbott, D. W. (2018). Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. Nat. Commun. 9, 1043.
8. Lee, S. B., Cho, S. J., Kim, J. A, Lee, S. Y., Kim, S. M. & Lim, H. S. (2014). Metabolic pathway of 3,6-anhydro-L-galactose in agar-degrading microorganisms. Biotechnol. Bioproc. Eng. 19, 866-878.
9. Yun, E. J., Lee, S., Kim, H. T., Pelton, J. G., Kim, S., Ko, H. J., Choi, I. G. & Kim, K. H. (2015). The novel catabolic pathway of 3,6-anhydro-L-galactose, the main component of red macroalgae, in a marine bacterium. Environ.

Microbiol. 17, 1677-1688.

10. Oh, Y. R., Jung, K. A., Lee, H. J., Jung, G. Y. & Park, J. M. (2018). A novel 3,6-anhydro-L-galactose dehydrogenase produced by a newly isolated *Raoultella ornithinolytica* B6-JMP12. Biotechnol. Bioproc. Eng. 23, 64-71.

11. Yu, S., Choi, I.-G., Yun, E. J. & Kim, K. H. (2018). High substrate specificity of 3, 6-anhydro-L-galactose dehydrogenase indicates its essentiality in the agar catabolism of a marine bacterium. Process Biochem. 64, 130-135.

12. Pathiraja, D., Kim, K. H. & Choi, I. G. (2017). Rapid and robust enzymatic sensing and quantitation of 3,6-Anhydro-L-galactose in a heterogeneous sugar mixture. Carbohydr. Res. 446, 13-18.

13. Koppaka, V., Thompson, D. C., Chen, Y., Ellermann, M., Nicolaou, K. C., Juvonen, R. O., Petersen, D., Deitrich, R. A., Hurley, T. D. & Vasiliou, V. (2012) Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application. Pharmacol. Rev. 64, 520-539.

14. Black, W. & Vasiliou, V. (2009). The aldehyde dehydrogenase gene superfamily resource center. Hum. genomics 4, 136-142.

15. Cobessi, D., Tete-Favier, F., Marchal, S., Branlant, G. & Aubry, A. (2000). Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. J. Mol. Biol. 300, 141-152.

16. D'Ambrosio, K., Pailot, A., Talfournier, F., Didierjean, C., Benedetti, E., Aubry, A., Branlant, G. & Corbier, C. (2006). The first crystal structure of a thioacylenzyme intermediate in the ALDH family: new coenzyme conformation and relevance to catalysis. Biochemistry 45, 2978-2986.

17. Hempel, J., Perozich, J., Chapman, T., Rose, J., Boesch, J. S., Liu, Z. J., Lindahl, R. & Wang, B. C. (1999). Aldehyde dehydrogenase catalytic mechanism. A proposal. Adv. Exp. Med. Biol. 463, 53-59.

18. Pohl, E., Brunner, N., Wilmanns, M. & Hensel, R. (2002). The crystal structure of the allosteric non-phosphorylating glyceraldehyde-3-phosphate

562 dehydrogenase from the hyperthermophilic archaeum *Thermoproteus tenax*. J.
563 Biol. Chem. 277, 19938-19945.
564 19. Hammen, P. K., Allali-Hassani, A., Hallenga, K., Hurley, T. D. & Weiner, H.
565 (2002). Multiple conformations of NAD and NADH when bound to human
566 cytosolic and mitochondrial aldehyde dehydrogenase. Biochemistry 41, 7156-
567 7168.
568 20. Perez-Miller, S. J. & Hurley, T. D. (2003). Coenzyme isomerization is integral
569 to catalysis in aldehyde dehydrogenase. Biochemistry 42, 7100-7109.
570 21. Halavaty, A. S., Rich, R. L., Chen, C., Joo, J. C., Minasov, G., Dubrovska, I.,
571 Winsor, J. R., Myszka, D. G., Duban, M., Shuvalova, L., Yakunin, A. F. &
572 Anderson, W. F. (2015). Structural and functional analysis of betaine aldehyde
573 dehydrogenase from *Staphylococcus aureus*. Acta Crystallogr. D Biol.
574 Crystallogr. 71, 1159-1175.
575 22. Langendorf, C. G., Key, T. L., Fenalti, G., Kan, W. T., Buckle, A. M., Caradoc-
576 Davies, T., Tuck, K. L., Law, R. H. & Whisstock, J. C. (2010). The X-ray crystal
577 structure of *Escherichia coli* succinic semialdehyde dehydrogenase; structural
578 insights into NADP+/enzyme interactions. PloS one 5, e9280.
579 23. Morgan, C. A. & Hurley, T. D. (2015). Development of a high-throughput in
580 vitro assay to identify selective inhibitors for human ALDH1A1. Chem. Biol.
581 Interact. 234, 29-37.
582 24. Di Costanzo, L., Gomez, G. A. & Christianson, D. W. (2007). Crystal structure
583 of lactaldehyde dehydrogenase from *Escherichia coli* and inferences regarding
584 substrate and cofactor specificity. J. Mol. Biol. 366, 481-493.
585 25. Kopecna, M., Vigouroux, A., Vilim, J., Koncitikova, R., Briozzo, P., Hajkova,
586 E., Jaskova, L., von Schwartzenberg, K., Sebela, M., Morera, S. & Kopecny, D.
587 (2017). The ALDH21 gene found in lower plants and some vascular plants
588 codes for a NADP(+)-dependent succinic semialdehyde dehydrogenase. Plant
589 J. 92, 229-243.
590 26. Lee, S. B., Lee, S. Y. & Lim, H. S. (2015). Aldehydic nature and conformation
591 of 3,6-anhydro-L-galactose monomer. Biotechnol. Bioproc. Eng. 20, 878-886.

27. Lorentzen, E., Hensel, R., Knura, T., Ahmed, H. & Pohl, E. (2004). Structural Basis of allosteric regulation and substrate specificity of the non-phosphorylating glyceraldehyde 3-Phosphate dehydrogenase from *Thermoproteus tenax*. J. Mol. Biol. 341, 815-828.

28. Cobessi, D., Tete-Favier, F., Marchal, S., Azza, S., Branlant, G. & Aubry, A. (1999). Apo and holo crystal structures of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*. J. Mol. Biol. 290, 161-173.

29. Gruez, A., Roig-Zamboni, V., Grisel, S., Salomoni, A., Valencia, C., Campanacci, V., Tegoni, M. & Cambillau, C. (2004). Crystal structure and kinetics identify *Escherichia coli* YdcW gene product as a medium-chain aldehyde dehydrogenase. J. Mol. Biol. 343, 29-41.

30. Habenicht, A. (1997). The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase: biochemistry, structure, occurrence and evolution. Biol. Chem. 378, 1413-1419.

31. Agarwal, V., Peck, S. C., Chen, J. H., Borisova, S. A., Chekan, J. R., van der Donk, W. A. & Nair, S. K. (2014). Structure and function of phosphonoacetaldehyde dehydrogenase: the missing link in phosphonoacetate formation. Chem. Biol. 21, 125-135.

32. Koncitikova, R., Vigouroux, A., Kopecna, M., Andree, T., Bartos, J., Sebela, M., Morera, S. & Kopecny, D. (2015). Role and structural characterization of plant aldehyde dehydrogenases from family 2 and family 7. Biochem. J. 468, 109-123.

33. Moretti, A., Li, J., Donini, S., Sobol, R. W., Rizzi, M. & Garavaglia, S. (2016). Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD(+) and retinoic acid. Sci. Rep. 6, 35710.

34. Kim, Y. G., Lee, S., Kwon, O. S., Park, S. Y., Lee, S. J., Park, B. J. & Kim, K. J. (2009). Redox-switch modulation of human SSADH by dynamic catalytic loop. EMBO J. 28, 959-968.

35. Chen, C., Joo, J. C., Brown, G., Stolnikova, E., Halavaty, A. S., Savchenko, A., Anderson, W. F. & Yakunin, A. F. (2014). Structure-based mutational studies of

substrate inhibition of betaine aldehyde dehydrogenase BetB from *Staphylococcus aureus*. Appl. Environ. Microbiol. 80, 3992-4002.

36. Stiti, N., Podgorska, K. & Bartels, D. (2014). Aldehyde dehydrogenase enzyme ALDH3H1 from *Arabidopsis thaliana*: Identification of amino acid residues critical for cofactor specificity. Biochim. Biophys. Acta. 1844, 681-693.

37. Sobreira, T. J., Marletaz, F., Simoes-Costa, M., Schechtman, D., Pereira, A. C., Brunet, F., Sweeney, S., Pani, A., Aronowicz, J., Lowe, C. J., Davidson, B., Laudet, V., Bronner, M., de Oliveira, P. S., Schubert, M. & Xavier-Neto, J. (2011). Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans. Proc. Natl. Acad. Sci. U.S.A. 108, 226-231.

38. Xia, Y., Chu, W., Qi, Q. & Xun, L. (2015). New insights into the QuikChange process guide the use of Phusion DNA polymerase for site-directed mutagenesis. Nucleic Acids Res. 43, e12.

39. Kim, H. T., Yun, E. J., Wang, D., Chung, J. H., Choi, I. G. & Kim, K. H. (2013). High temperature and low acid pretreatment and agarase treatment of agarose for the production of sugar and ethanol from red seaweed biomass. Bioresour. Technol. 136, 582-587.

40. Yun, E. J., Shin, M. H., Yoon, J.-J., Kim, Y. J., Choi, I.-G. & Kim, K. H. (2011). Production of 3,6-anhydro-l-galactose from agarose by agarolytic enzymes of *Saccharophagus degradans* 2-40. Process Biochem. 46, 88-93.

41. Hidalgo, E., Chen, Y. M., Lin, E. C. & Aguilar, J. (1991). Molecular cloning and DNA sequencing of the *Escherichia coli* K-12 ald gene encoding aldehyde dehydrogenase. J. Bacteriol. 173, 6118-6123.

42. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallogr. D Biol. Crystallogr. 62, 859-866.

43. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126-2132.

44. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. 58, 1948-1954.

45. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C. & Mainz, D. T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J. Med. Chem. 49, 6177-6196.

FIGURE LEGENDS

Fig. 1. Multiple-sequence alignment of *Vv*AHGD and its homologs and *Ec*ALDH. Using ESPript, the secondary structures of *Vv*AHGD are shown above alignment and those of *Ec*ALDH under alignment. Helices are indicated by springs, strands by arrows, turns by TT letters, and 3_{10} helices by η letters. Identical residues are shown in white on a black background, and similar residues are in bold black. Green circles indicate key catalytic residues conserved in ALDHs. Red triangles indicate key residues of *Vv*AHGD involved in the binding of the cofactor $NADP^+$ and the substrate L-AHG.

Fig. 2. Substrate specificity analyses of *Vv*AHGD, *Vej*AHGD and *Ec*ALDH with different cofactors. (a) Thin layer chromatography analysis of the crude extract of L-AHG. The crude extract of L-AHG prepared from neoagarobioses comprises L-AHG and D-galactose. Lane 1, the standards of D-AHG, neoagarobiose and D-galactose; lanes 2 and 3, the prepared crude extract of L-AHG. (b) Substrate specificity analysis of *Vv*AHGD at 40℃ in 20 mM Tris-HCl (pH 7.0) buffer. (c) Substrate specificity analysis of *Vej*AHGD at 30℃ in 20 mM Tris-HCl (pH 7.0) buffer. (d) Substrate specificity analysis of *Ec*ALDH at 25℃ in 20 mM Tris-HCl (pH 8.0) buffer. The data shown in (b)-(d) are from triplicate experiments (mean ± S.D.).

Fig. 3. Effect of temperature (a), pH (b) and NaCl (c) on the activity of *Vv*AHGD. In (a)-(c), the highest activity of *Vv*AHGD at 40℃ (a), at pH 7.0 (b) and in 0 M NaCl (c) was taken as 100%, respectively. The data shown in (a)-(c) are from triplicate experiments (mean ± S.D.).

Fig. 4. Structural analysis of *Vv*AHGD. (a) The structure of tetrameric *Vv*AHGD in one asymmetric unit. (b) DLS analysis of the molecular mass of *Vv*AHGD. *Vv*AHGD monomer has a calculated molecular mass of 52.6 kDa. DLS analysis indicated that *Vv*AHGD forms dimers in solution. (c) Superimposition of *Vv*AHGD and other ALDHs. *Vv*AHGD is colored in cyan, *Ec*ALDH (PDB code 2HG2) in magenta, the succinic semialdehyde dehydrogenase (PDB code 2W8R) in yellow, and the betaine aldehyde dehydrogenase (PDB code 4MPY) in green. (d) Monomeric structure of *Vv*AHGD. The cofactor-binding domain is colored in yellow, the catalytic domain in green, and the oligomerization domain in red. The potential catalytic residues Cys282 and Glu248 are

shown in sticks. The loop containing the catalytic residue Cys282 is colored in magenta.

Fig. 5. Analyses of the *Vv*AHGD structure in complex with $NADP^+$ and the important residues involved in $NADP^+$ binding. (a) Electrostatic surface view of the *Vv*AHGD-NADP complex. $NADP^+$ is shown as yellow sticks. (b) Cartoon view of the overall structure of *Vv*AHGD-NADP complex. $NADP^+$ is shown as yellow sticks and the corresponding $2F_o$-F_c electron density map is shown at 1.0 σ. (c) Comparison of the catalytic tunnels in the apo and the $NADP^+$-bound *Vv*AHGD. The catalytic residues Cys282 and Glu248 are shown as cyan sticks. The binding of $NADP^+$ induces structural changes of *Vv*AHGD in the active site, leading to a connection between the cofactor channel and the substrate channel. (d) Superimposition of the active sites of the apo (grey) and $NADP^+$-bound (cyan) *Vv*AHGD. The catalytic residues Cys282 and Glu248 are shown as magenta sticks in the $NADP^+$-bound structure, and as grey sticks in the apo structure. (e) Superimposition of the potential $NADP^+$-binding residues in the apo (grey) and the $NADP^+$-bound (cyan) *Vv*AHGD.

Fig. 6. Analysis of the key residues in *Vv*AHGD for substrate binding and mutational analysis of the identified key residues in *Vv*AHGD. (a) Electrostatic surface view of the $NADP^+$-bound *Vv*AHGD docked with the substrate L-AHG in open-chain aldehyde form. The substrate is shown as green sticks, and the cofactor as yellow sticks. (b) Detailed structure of the $NADP^+$-bound *Vv*AHGD docked with L-AHG. Residues involved in the binding and catalysis of L-AHG are shown as cyan sticks. The substrate is shown as green sticks, and the cofactor as yellow sticks. (c) Enzymatic activities of the mutants of *Vv*AHGD. The activity of wild-type *Vv*AHGD was defined as 100%. The graph shows data from triplicate experiments (mean ± S.D.). (d) CD spectra of wild-type *Vv*AHGD and its mutants.

Fig. 7. A proposed mechanism for *Vv*AHGD to catalyze the oxidation of L-AHG with $NADP^+$ as a cofactor. Electron transfer is shown as red arrows. The cofactor channel and the substrate channel are located on the opposite sides of the catalytic residues Cys282 and Glu248. The borders of these two channels are shown as dashed lines. The gap within the dashed line represents the opening of the corresponding channel, and a lager gap represents a larger opening of the channel. (a) The resting state.

The cofactor channel and the substrate channel are disconnected due to the blocking by the catalytic residues Cys282 and Glu248. (b) The hydride transfer state. When the cofactor and the substrate enter their respective channels, Cys282 and Glu248 undergo conformational changes, and thus the nucleophilic Cys282 could attack the L-AHG substrate to form a tetrahedral intermediate with concomitant hydride transfer from the intermediate to $NADP^+$. In this state, the cofactor channel and the substrate channel are connected due to the conformational changes of Cys282 and Glu248. (c) The hydrolysis state. The resulting NADPH moves away from the active site, leaving room for the catalytic Glu248 to hydrolyze the thioester intermediate. In this state, the cofactor channel and the substrate channel are still connected, but the opening between these two channels possibly becomes narrow due to the conformational change of Glu248. (d) Back to the resting state. NADPH and the product are released, and the enzyme molecule returns to the resting state.

Fig. 8. Phylogenetic analysis of *Vv*AHGD and other ALDHs. The tree was constructed by the neighbor-joining method with a Jones-Taylor-Thornton (JTT) matrix-based model using 428 amino acid positions. Bootstrap analysis of 1,000 replicates is conducted and values above 50% are shown. The structural formula(s) of the main substrate(s) catalyzed by ALDHs is also shown for each family. Sequence identities shared by *Vv*AHGD and other ALDHs range from 33% to 79%.

Fig. 9. Structural comparison of the cofactor channels of ALDHs from different families. (a) Comparison of the electrostatic potential surfaces of the cofactor channels of ALDHs from different families. The NAD^+/$NADP^+$ cofactor is shown as sticks. In the structure of the selected ALDH5 succinic semialdehyde dehydrogenase (PDB code 2W8R), only the ADP moiety of the NAD^+ cofactor is observed. (b) Superposition of the cofactor channels of ALDHs from different families. Residues that form hydrogen bonds with the cofactor are shown as sticks, and colored in cyan for *Vv*AHGD, in magenta for *Ec*ALDH (PDB code 2IMP), in yellow for the ALDH5 succinic semialdehyde dehydrogenase (PDB code 2W8R), and in green for the ALDH25 betaine aldehyde dehydrogenase (PDB code 4MPY). $NADP^+$ bound in *Vv*AHGD is shown as cyan lines, and NAD^+ from the ALDH25 betaine aldehyde dehydrogenase in green lines.

For residues conserved in all enzymes, only the residue positions of *Vv*AHGD are shown in black. Residues that recognize NAD^+ and $NADP^+$ are boxed. (c) An overview of conservation of residues forming the cofactor channels of ALDHs from different families. Residues are numbered according to *Vv*AHGD. Residues surrounding the nicotinamide riboside, diphosphate and adenosine/adenylate moieties of the NAD^+/$NADP^+$ cofactor are marked by solid circles, open squares and open circles, respectively. Sequence logos were made using WebLogo (http://weblogo.threeplusone.com).

Fig. 10. Structural comparison of the substrate channels of ALDHs from different families. (a) Comparison of the electrostatic potential surfaces of the substrate channels (marked by yellow circles) of selected ALDHs with different substrate specificities. (b) Close-ups of the entrances of the substrate channels of *Vv*AHGD and *Ec*ALDH. Residues forming the substrate channels of *Vv*AHGD and *Ec*ALDH are shown as cyan and magenta sticks, respectively. For *Vv*AHGD, the modeled L-AHG substrate is shown as yellow sticks. For *Ec*ALDH, the lactate product is shown as green sticks. (c) Superposition of the substrate channels of *Vv*AHGD and *Ec*ALDH. The modeled L-AHG substrate is shown as yellow sticks. (d) An overview of conservation of residues forming the substrate channels of AHGDs and lactaldehyde dehydrogenases. Residues are numbered according to *Vv*AHGD. Residues forming the substrate channels of lactaldehyde dehydrogenases are marked by solid circles. Sequence logos were made using WebLogo (http://weblogo.threeplusone.com).

773 **TABLE 1** Diffraction data and refinement statistics of the apo *Vv*AHGD and the

774 *Vv*AHGD-NADP complex.

Parameters	*Vv*AHGD	*Vv*AHGD-NADP
Diffraction data		
Space group	$P2_12_12_1$	$P22_12_1$
Unit cell		
a, b, c (Å)	103.39, 132.36, 159.38	76.38, 90.74, 149.05
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00-2.70 (2.79-2.70)[a]	50.00-2.37 (2.45-2.37)
Redundancy	12.2 (11.9)	13.3 (12.8)
Completeness (%)	99.0 (100.0)	99.0 (97.6)
R_{merge} (%)[b]	17.4 (43.1)	16.5 (34.0)
$I/\sigma I$	27.5 (7.2)	44.3 (15.3)
Refinement statistics		
Resolution (Å)	47.24-2.70 (2.74-2.70)	45.98-2.37 (2.42-2.37)
Number of reflections	60214	42552
R-factor (%)	16.1	16.3
Free R-factor (%)	21.1	20.5
Number of atoms		
Protein	14684	7350
Solvent	407	560
Ligands		96
Average B-factor (Å^2)		
Protein	24.82	20.00
Solvent	25.14	27.01
Ligands		17.24
RMSD from ideal geometry		
Bond lengths (Å)	0.010	0.007
Bond angles (°)	1.15	0.86
Ramachandran plot (%)		
Favoured	95.84	96.21
Allowed	4.00	3.68
Outliers	0.16	0.11

775 [a] Numbers in parentheses refer to data in the highest resolution shell.

776 [b] $R_{merge}=\sum_{hkl}\sum_i|I(hkl)_i - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i\langle I(hkl)_i\rangle$

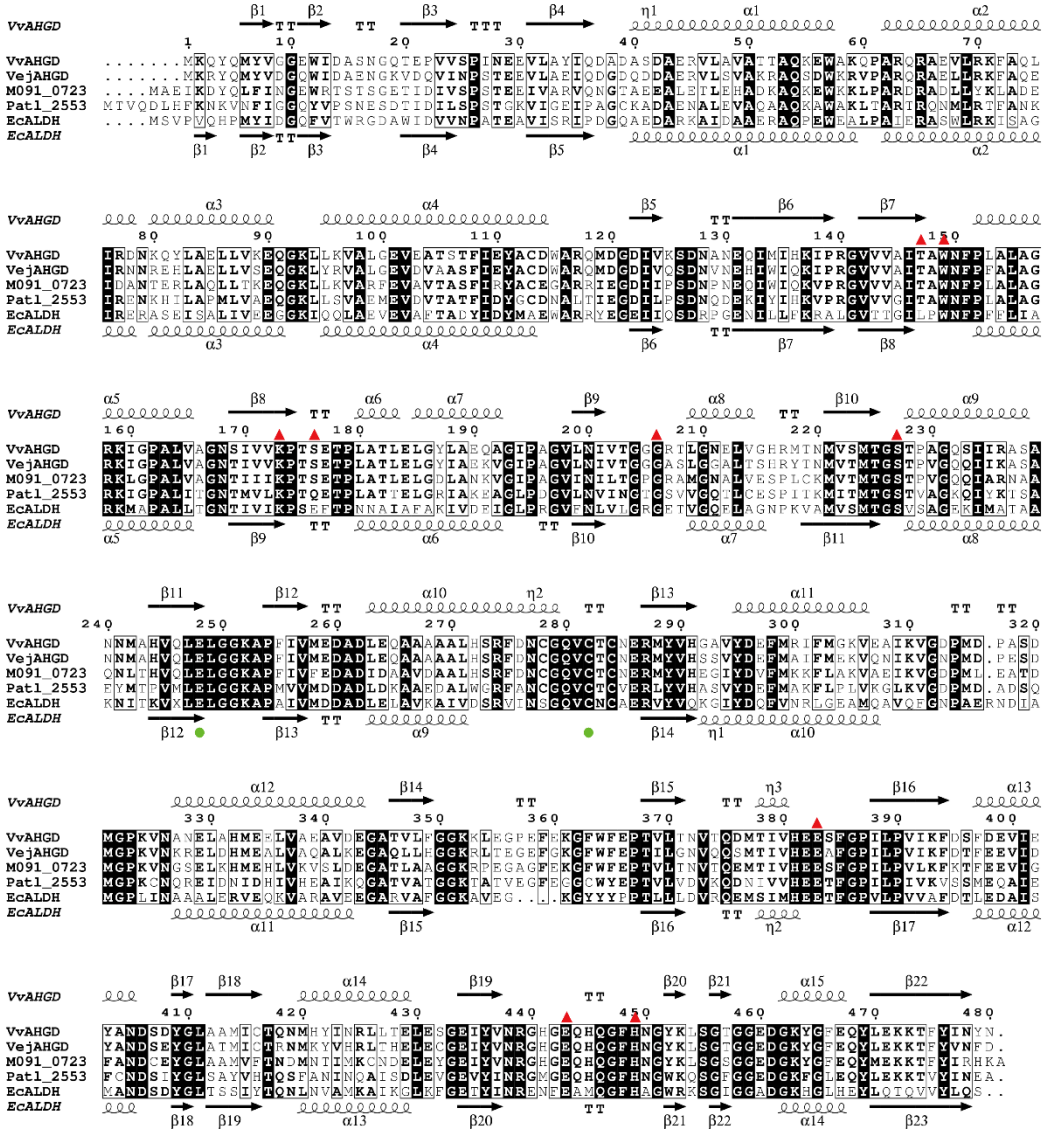
777

778 **TABLE 2** Kinetic parameters of *Vv*AHGD and its mutants against the cofactor
779 $NADP^+$ and the substrate L-AHG.

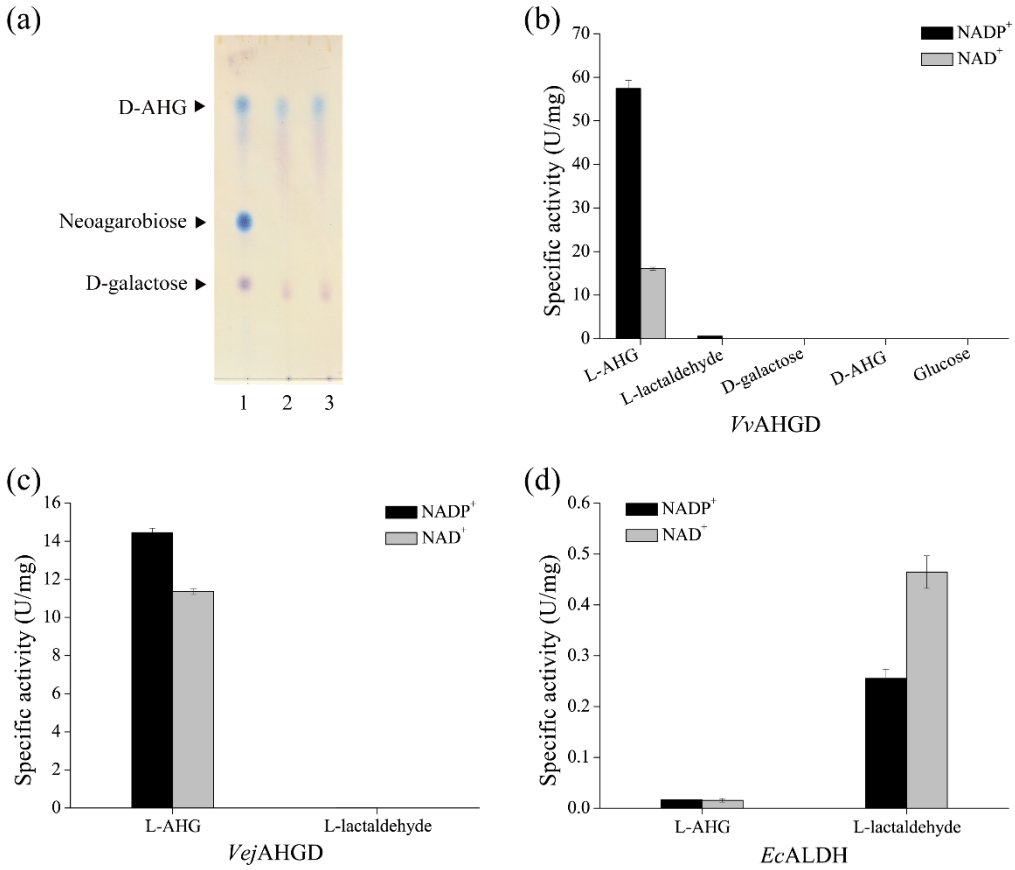
Enzyme	$NADP^{+a}$			L-AHG[b]		
	K_m	k_{cat}	k_{cat}/K_m	K_m	k_{cat}	k_{cat}/K_m
	mM	s^{-1}	$mM^{-1}\ s^{-1}$	mM	s^{-1}	$mM^{-1}\ s^{-1}$
Wild Type	0.33 ± 0.03	69.09 ± 3.04	209.36	1.20 ± 0.08	58.29 ± 2.63	48.58
S176A	1.11 ± 0.16	19.15 ± 1.61	17.25	1.45 ± 0.11	10.67 ± 0.68	7.36
S227A	0.58 ± 0.02	44.55 ± 0.88	76.81	1.26 ± 0.04	35.99 ± 0.86	28.56
S233A	0.38 ± 0.05	75.9 ± 4.01	199.74	1.35 ± 0.03	63.18 ± 1.26	46.80
L249A	0.41 ± 0.11	63.34 ± 5.42	154.49	1.11 ± 0.05	52.77 ± 9.31	47.54

780 [a] To measure the kinetic parameters of *Vv*AHGD and its mutants against the cofactor
781 $NADP^+$, reactions were conducted in triplicate in 20 mM Tris-HCl buffer (pH 7.0) using $NADP^+$
782 over a concentration range of 0-4.8 mM and L-AHG at a fixed concentration of 3.52 mM.
783 [b] To measure the kinetic parameters of *Vv*AHGD and its mutants against L-AHG, reactions
784 were conducted in triplicate in 20 mM Tris-HCl buffer (pH 7.0) using L-AHG over a
785 concentration range of 0-4.4 mM and $NADP^+$ at a fixed concentration of 1.5 mM.
786

Fig. 1



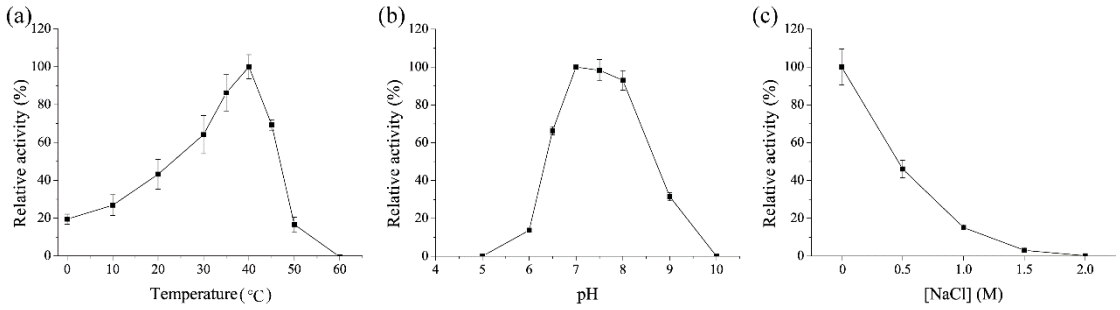
790 Fig. 2



791

792

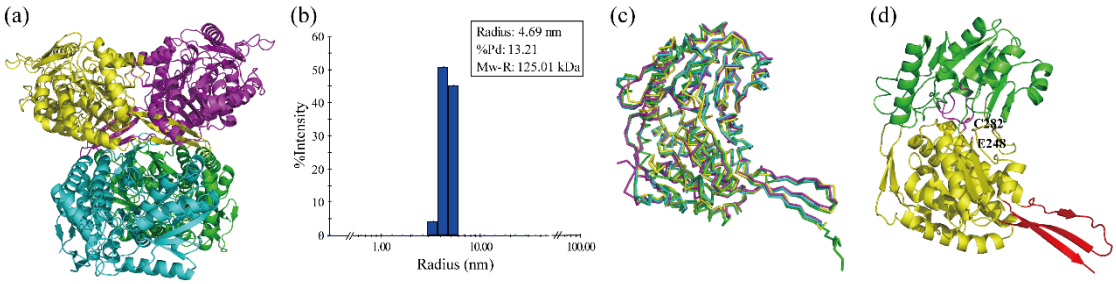
793 Fig. 3



794

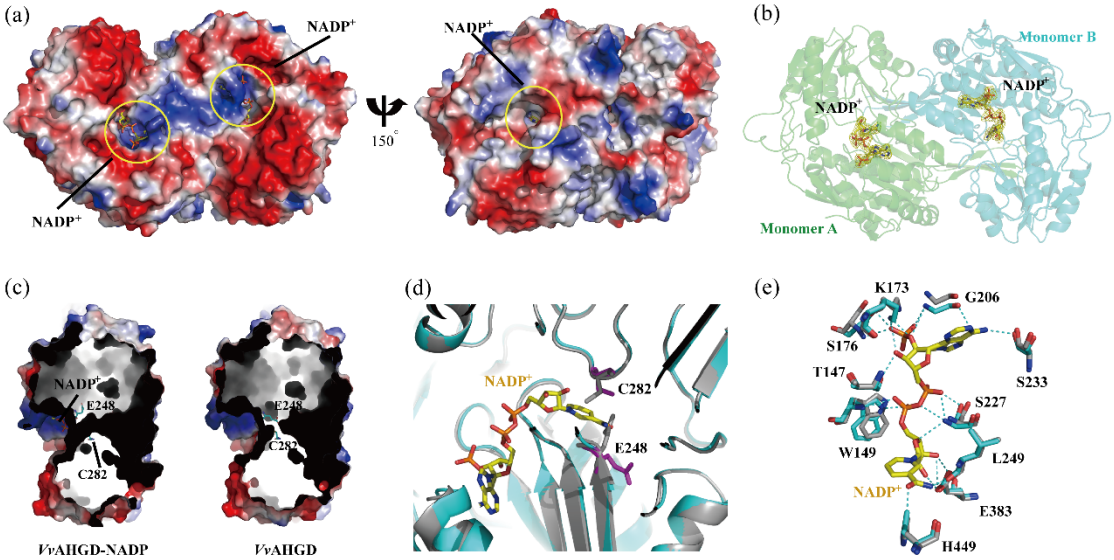
795

Fig. 4


(a)
(b)
Radius: 4.69 nm
%Pd: 13.21
Mw-R: 125.01 kDa
%Intensity
Radius (nm)
(c)
(d)
C282
E248

799

Fig. 5

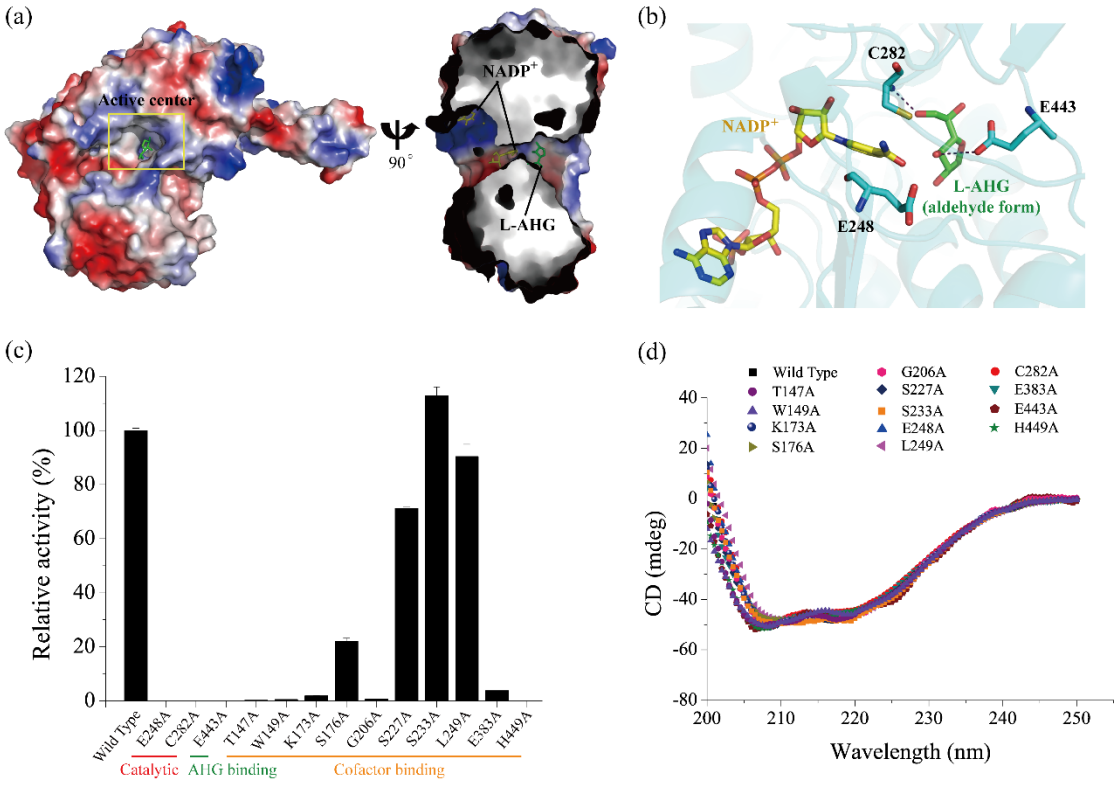


800

801

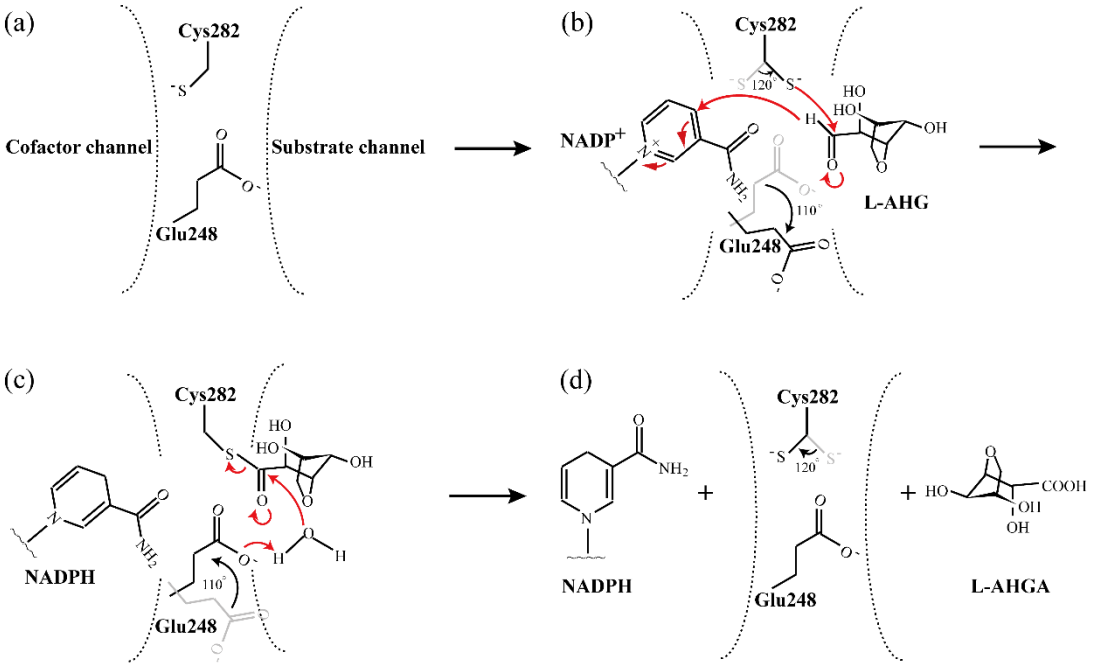
802 Fig. 6

803



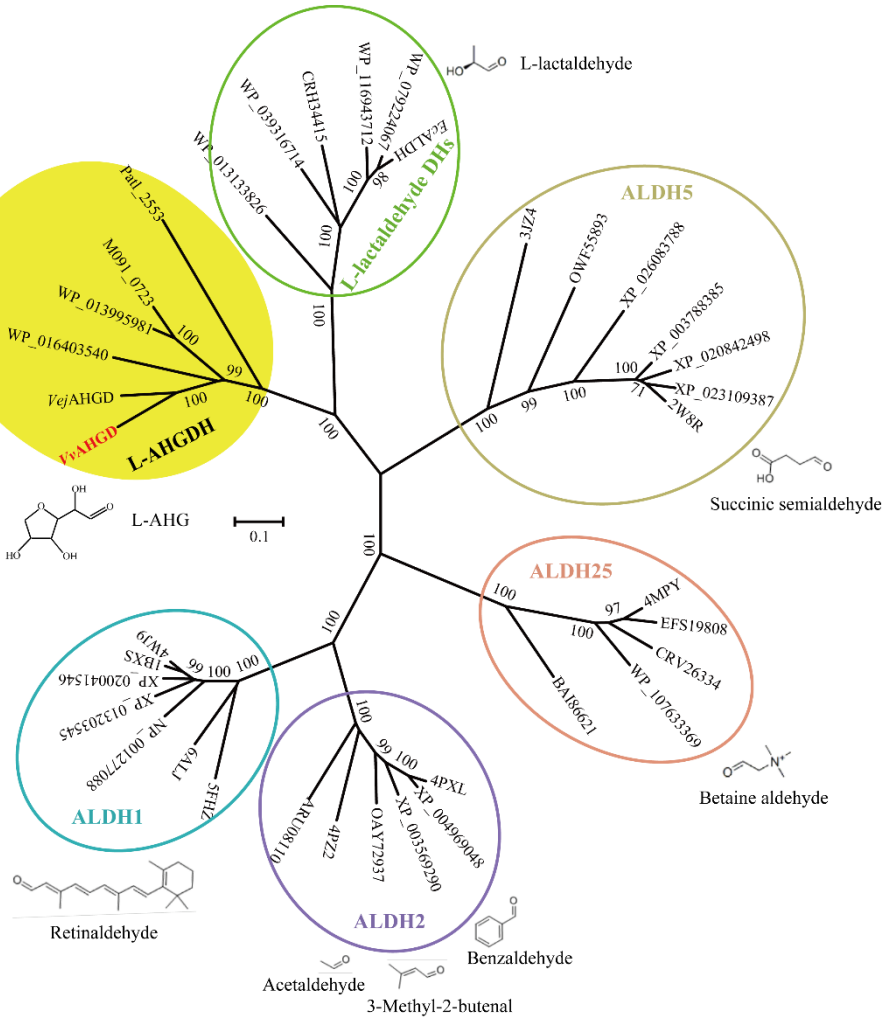
804

Fig. 7


(a)
Cys282
Cofactor channel
Substrate channel
Glu248
(b)
Cys282
NADP+
L-AHG
Glu248
(c)
Cys282
NADPH
Glu248
(d)
NADPH
Cys282
Glu248
L-AHGA

808

Fig. 8

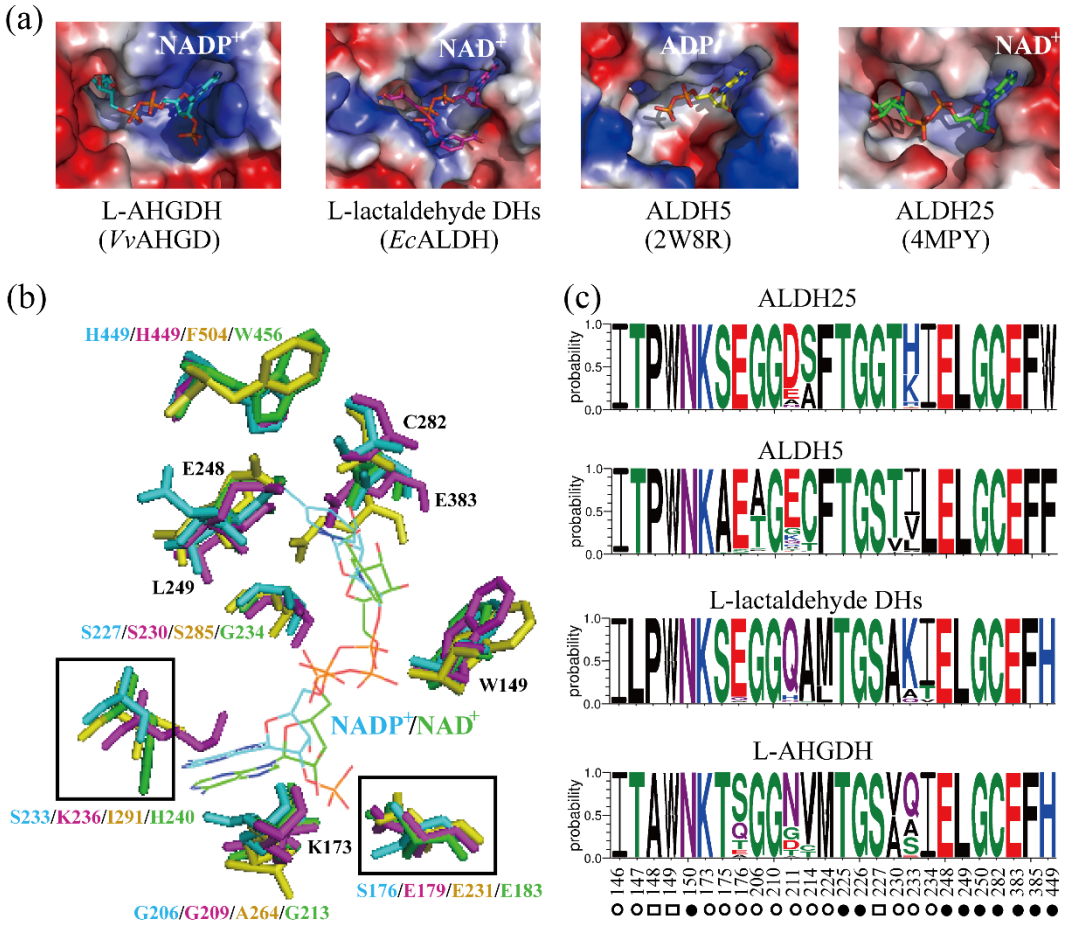


809

810

811

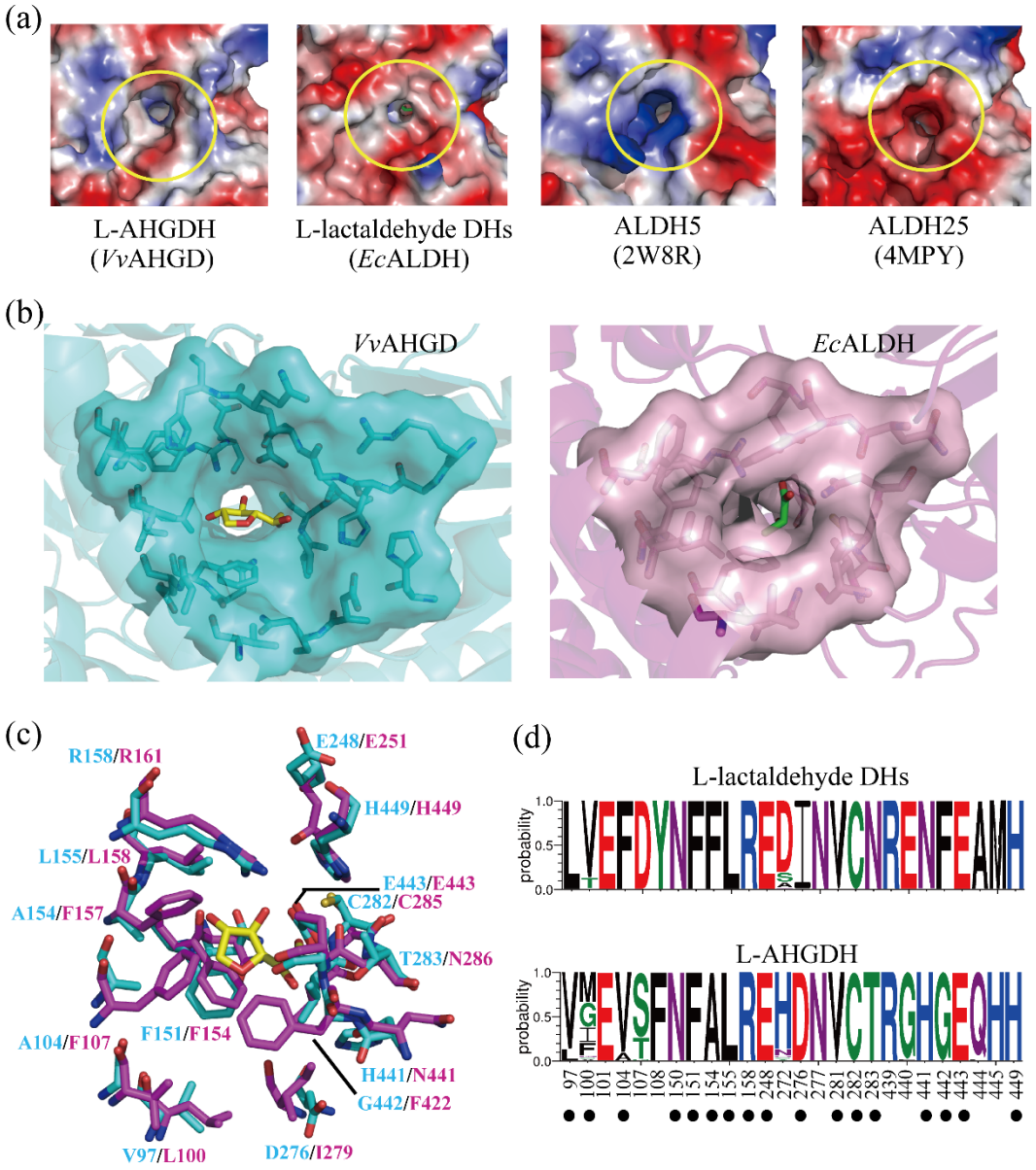
Fig. 9



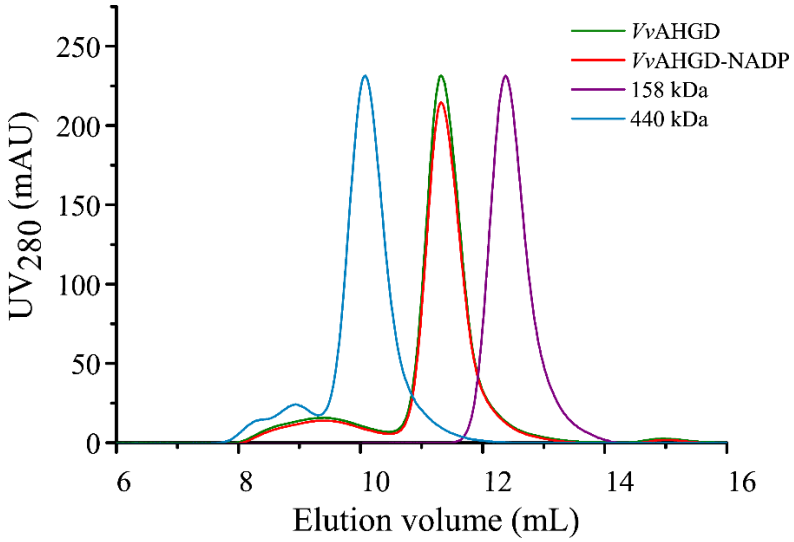
812

813

Fig. 10



SUPPLEMENTAL MATERIAL



818

819

820 **Fig. S1.** Gel filtration analysis of the apo *Vv*AHGD and the *Vv*AHGD-NADP complex.
821 *Vv*AHGD monomer has a calculated molecular mass of 52.6 kDa. Two protein size
822 markers are aldolase (158 kDa) and ferritin (440 kDa).

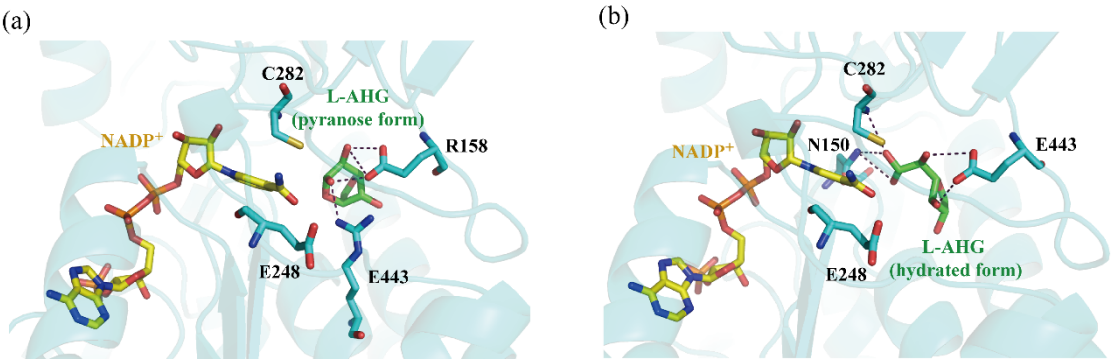


Fig. S2. Detailed structure of the NADP$^+$-bound *Vv*AHGD docked with L-AHG in pyranose (a) or hydrated form (b). Residues involved in the binding and catalysis of L-AHG are shown as cyan sticks. The substrate is shown as green sticks, and the cofactor as yellow sticks.