PCP
PLANT & CELL PHYSIOLOGY

Regular Paper

Structural Characterization of L-Galactose Dehydrogenase: An Essential Enzyme for Vitamin C Biosynthesis

Jhon A. Vargas[1,†], Diego A. Leonardo[1,†], Humberto D'Muniz Pereira[1], Adriana R. Lopes[2], Hicler N. Rodriguez[3], Marianela Cobos[3,4,5], Jorge L. Marapara[3,5], Juan C. Castro[3,5] and Richard C. Garratt[1,*]

[1]São Carlos Institute of Physics, University of São Paulo, Avenida João Dagnone 1100, São Carlos, SP 13563-120, Brazil
[2]Laboratory of Biochemistry, Instituto Butantan, Av. Vital Brasil, São Paulo 1500, Brazil
[3]Unidad Especializada del Laboratorio de Investigación en Biotecnología (UELIB), Centro de Investigaciones de Recursos Naturales de la UNAP (CIRNA), Universidad Nacional de la Amazonia Peruana (UNAP), Psje. Los Paujiles S/N, Iquitos 1600, Peru
[4]Laboratorio de Biotecnología y Bioenergética, Universidad Científica del Perú, Av. Abelardo Quiñones km 2.5, Iquitos 16006, Peru
[5]Departamento Académico de Ciencias Biomédicas y Biotecnología (DACBB), Facultad de Ciencias Biológicas (FCB), Universidad Nacional de la Amazonia Peruana (UNAP), Ciudad Universitaria - Zungarococha, San Juan Bautista 16000, Peru
[†]These authors contributed equally.

*Corresponding author: E-mail, richard@ifsc.usp.br

(Received 15 April 2022; Accepted 27 June 2022)

In plants, it is well-known that ascorbic acid (vitamin C) can be synthesized via multiple metabolic pathways but there is still much to be learned concerning their integration and control mechanisms. Furthermore, the structural biology of the component enzymes has been poorly exploited. Here we describe the first crystal structure for an L-galactose dehydrogenase [*Spinacia oleracea* GDH (*So*GDH) from spinach], from the D-mannose/L-galactose (Smirnoff–Wheeler) pathway which converts L-galactose into L-galactono-1,4-lactone. The kinetic parameters for the enzyme are similar to those from its homolog from camu camu, a super-accumulator of vitamin C found in the Peruvian Amazon. Both enzymes are monomers in solution and have a pH optimum of 7, and their activity is largely unaffected by high concentrations of ascorbic acid, suggesting the absence of a feedback mechanism acting via GDH. Previous reports may have been influenced by changes of the pH of the reaction medium as a function of ascorbic acid concentration. The structure of *So*GDH is dominated by a $(\beta/\alpha)_8$ barrel closely related to aldehyde-keto reductases (AKRs). The structure bound to NAD^+ shows that the lack of Arg279 justifies its preference for NAD^+ over $NADP^+$, as employed by many AKRs. This favors the oxidation reaction that ultimately leads to ascorbic acid accumulation. When compared with other AKRs, residue substitutions at the C-terminal end of the barrel (Tyr185, Tyr61, Ser59 and Asp128) can be identified to be likely determinants of substrate specificity. The present work contributes toward a more comprehensive understanding of structure–function relationships in the enzymes involved in vitamin C synthesis.

Keywords: Crystal structure • Enzyme kinetics • L-galactose dehydrogenase • *Myrciaria dubia* 'camu-camu' • Spinach • Vitamin C biosynthesis

Introduction

L-Ascorbic Acid (ascorbate or AsA), commonly called vitamin C, is an essential biomolecule with key pleiotropic functions in both animals and plants. In animals, in addition to its free radical-scavenging activity, AsA functions as a cofactor for hydroxylases, mono-oxygenases and dioxygenases (Englard and Seifter 1986, Fenech et al. 2019). Furthermore, emerging evidence suggests that AsA participates in the regulation of relevant epigenomic processes such as the modulation of both DNA methylation and histone chemical modification (Young et al. 2015, Nur et al. 2021). Similarly, in plants AsA has a strong antioxidant activity, is a cofactor for mono- and dioxygenases (including enzymes for the biosynthesis of phytohormones and anthocyanins), acts as a photoprotectant and controls plant growth through cell division and expansion (Smirnoff 2000, Smirnoff and Wheeler 2000).

In plants, potentially there are multiple biosynthetic routes for the production of AsA (Valpuesta and Botella 2004, Castro et al. 2015, Caruso et al. 2021, Liao et al. 2021), although it is still unclear how many of these are of physiological relevance. Of these, the best characterized is the D-mannose/L-galactose (or Smirnoff–Wheeler) pathway (Wheeler et al. 1998, **Supplementary Fig. S1**). However, despite considerable effort, structural information concerning its component enzymes is still very limited. Structures currently available include

Plant Cell Physiol. 63(8): 1140–1155 (2022) doi:https://doi.org/10.1093/pcp/pcac090, Advance Access publication on 28 June 2022, available online at https://academic.oup.com/pcp

GDP-mannose 3,5-epimerase (GME) from *Arabidopsis thaliana* (Major et al. 2005) and L-galactono-1,4-lactone dehydrogenase (GalDH) as part of a macromolecular complex from *Brassica oleracea* (Soufari et al. 2020). One of the enzymes for which a structure has yet to be reported is L-galactose dehydrogenase (L-GDH, EC 1.1.1.316), which catalyzes the oxidation of L-galactose (L-gal) at position C1 to produce L-galactono-1,4-lactone in the penultimate step of the synthesis of AsA. L-GDH was one of the first enzymes of the pathway to be properly studied, being recognized as one of the most important in the regulation of AsA biosynthesis (Wheeler et al. 1998, Gatzek et al. 2002, Laing et al. 2004, Mieda et al. 2004).

Biochemical studies of L-GDH, either purified from leaves or via heterologous expression, have been carried out on different species including *Pisum sativum* (*Ps*GDH), *A. thaliana* (*At*GDH), *Spinacia oleracea* (*So*GDH) and *Actinidia deliciosa* (*Ad*GDH) (Gatzek et al. 2002, Laing et al. 2004, Mieda et al. 2004). L-GDH shows a high affinity and selectivity for L-gal with K_m ranging from 0.08 to 0.43 mM, as well as a preference for NAD^+ rather than $NADP^+$ as its cofactor (Gatzek et al. 2002, Laing et al. 2004, Mieda et al. 2004). Furthermore, AsA has been shown to slowly and irreversibly inactivate the enzyme or act as a feedback inhibitor through competitive inhibition with the substrate. This suggests that L-GDH may play a significant role in the regulation of AsA biosynthesis in plants (Laing et al. 2004, Mieda et al. 2004). However, this potential regulatory mechanism for the biosynthesis of AsA needs to be further clarified at the molecular level. Despite the extensive knowledge about the biochemical characteristics of L-GDH, together with two entries for the enzyme from rice (*Oryza sativa*, *Os*GDH) in the Protein Data Bank (PDB; 7EZL and 7EZI), its structural characteristics have yet to be described in the literature.

Myrciaria dubia (Kunth) McVaugh, also known as camu camu, is a fruit, native to the Amazon rainforest. Camu camu is characterized by its high concentration of vitamin C in fruits, which varies between 0.96 and 2.99 g/100 g of pulp (Castro et al. 2018). By comparison, species which are considered to be references for vitamin C content, such as oranges or kiwis, have concentrations of only 0.05 g and 0.085 g/100 g of pulp, respectively (Locato et al. 2013). This characteristic allows us to speculate that the enzymes from camu camu may have potential biotechnological applications in the artificial production of vitamin C.

Camu camu potentially has five metabolic pathways for vitamin C biosynthesis, including the D-mannose/L-galactose pathway (Castro et al. 2015). However, many of these have yet to be demonstrated to be effective in vivo. Here, we purify the recombinant L-GDH from *M. dubia* (*Md*GDH) and compare its biochemical properties with that from spinach (*So*GDH). In so doing we bring into question the previously reported role of AsA in feedback inhibition. Finally, for the first time, we describe the crystal structure of the spinach enzyme in both its apo and NAD^+ bound forms (*So*GDH), allowing for the elucidation of the cofactor-induced conformational change and the identification of residues potentially important for substrate specificity.

Results and Discussion

The oligomeric state of L-GDH

Heterologously expressed camu camu *Md*GDH and spinach *So*GDH showed a single peak on size exclusion chromatography (SEC) and a predominant band on SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) (**Supplementary Fig. S2A**). Consistent with this result, the SEC coupled with multi-angle light scattering (SEC-MALS) profiles showed a predominant monodisperse peak, corresponding to the expected molecular mass of the monomeric state for both recombinant enzymes (**Supplementary Fig. S2B**). In contrast, in some previous investigations, purified L-GDH exhibited variation in the oligomeric state, depending on the plant species and source employed. For example, *Ps*GDH purified from *P. sativum* (pea) embryonic axis was reported to be a tetrameric protein (Gatzek et al. 2002). On the other hand, *So*GDH and *Ad*GDH, purified from the leaves of *S. oleracea* (spinach) and *A. deliciosa* (kiwifruit), were described as being homodimers and monomers, respectively (Laing et al. 2004, Mieda et al. 2004). Similarly, the recombinant enzymes from *A. thaliana* (*At*GDH) and *O. sativa* (*Os*GDH) were reported to be homodimeric and monomeric, respectively (Gatzek et al. 2002, Momma and Fujimoto 2013). Although there are differences in the techniques used for protein purification from natural or heterologous sources, this would not normally be expected to affect the oligomeric state of the enzyme. Moreover, the investigations that reported the enzyme to form dimers or tetramers provided relatively limited data to substantiate these claims (Gatzek et al. 2002, Mieda et al. 2004). By contrast, the studies that reported the monomeric state for this enzyme used size exclusion chromatographic analysis (Laing et al. 2004, Momma and Fujimoto 2013).

The monomeric state of both L-GDHs studied here, which appears to be unambiguous from our SEC-MALS results, is also consistent with the inclusion of L-GDH as part of the aldehyde-keto reductase (AKR) protein superfamily (Gatzek et al. 2002). This is composed of 16 different families that are distributed across all phyla. These enzymes catalyze oxidation–reduction reactions on carbonyl substrates (e.g. sugar aldehydes, ketosteroids, quinones, etc.) and usually are monomers whose fold is dominated by a triosephosphate isomerase (TIM)-barrel motif $(\beta/\alpha)_8$. They are further characterized by a conserved cofactor-binding region for NAD(P)H, variable loop structures that define substrate specificity and a catalytic tetrad (Jez et al. 1997, Hyndman et al. 2003, Penning 2015). Of the 16 families of AKRs (Penning 2015), only three dimeric AKRs have been reported: a rat liver aflatoxin dialdehyde reductase (AKR7A1), the *Candida tenuis* NAD(P)H-dependent xylose reductase (AKR2B5) and the *Escherichia coli* tyrosine auxotrophy suppressor protein (Kozma et al. 2002, Kavanagh et al. 2003, Obmolova et al. 2003). The crystal structures of these proteins show a conserved contact interface for dimer formation, which is absent from the enzymes described here. This will be discussed later when describing the crystal structure of *So*GDH and adds weight to the expectation that L-GDH is, in general, a monomeric enzyme.

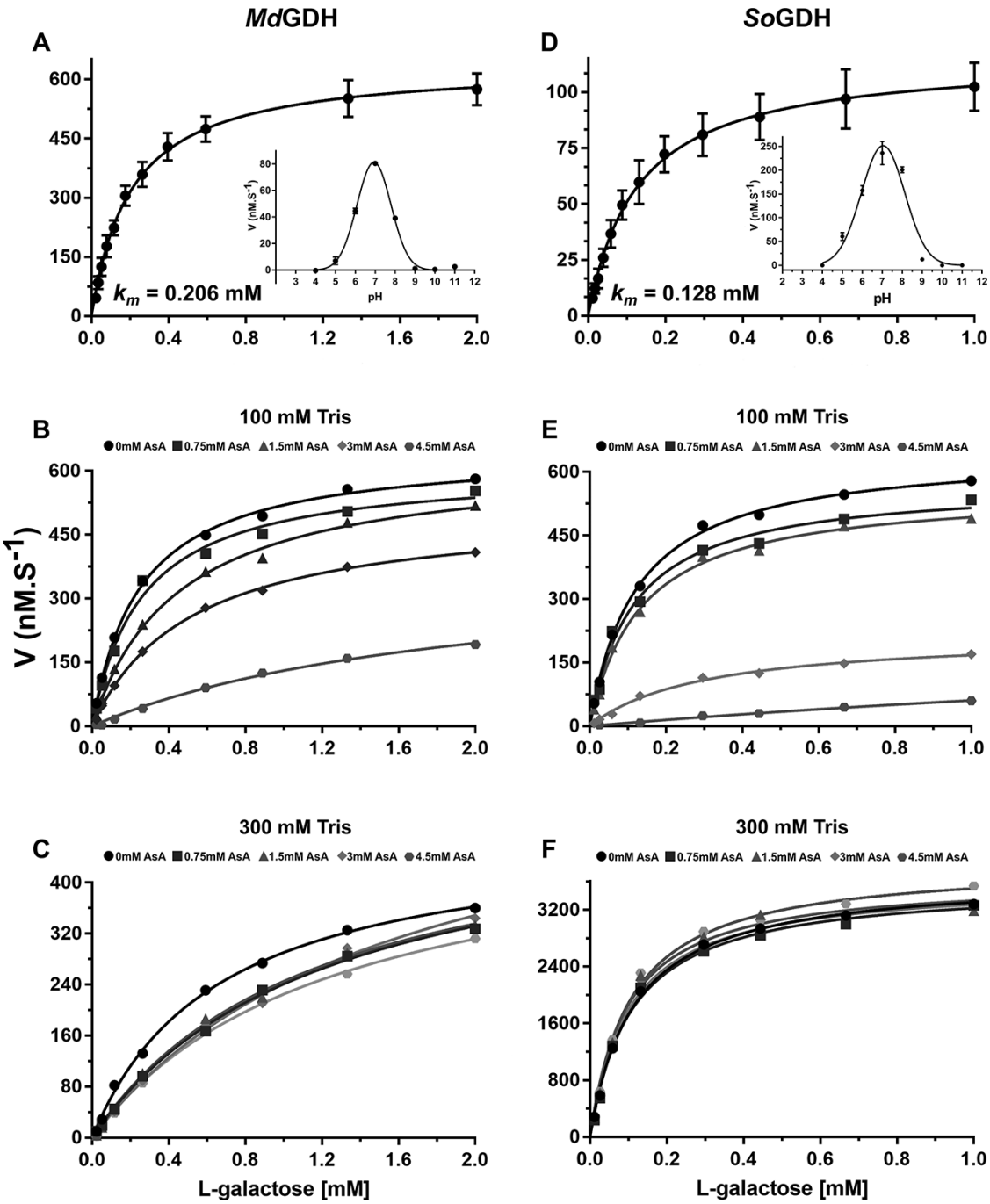

Fig. 1 Kinetic properties of camu camu *Md*GDH and spinach *So*GDH in the presence of excess NAD^+ (200 μM). Michaelis–Menten-like type kinetics were observed for *Md*GDH (A) and *So*GDH (D) with K_m values of 0.206 and 0.128 mM, respectively (mean values of five independent measurements with corresponding error bars). Optimum pH curves are shown as insets. The inhibition of *Md*GDH and *So*GDH by AsA in either 100 mM Tris–HCl buffer (B and E) or 300 mM Tris–HCl (C and F) are shown. The inhibitory effect of AsA drops off drastically in the higher concentration buffer.

High concentrations of AsA induce inhibition of L-GDH by pH changes

Camu camu *Md*GDH shows Michaelis–Menten-like kinetics using L-gal as substrate, with an optimal pH value of 7.0 and K_m of 0.21 mM (**Fig. 1A**). According to recent reports (Gatzek et al. 2002, Laing et al. 2004, Mieda et al. 2004), the K_m value for L-GDH from different plant species varies from 0.08 to 0.43 mM, and consequently, *Md*GDH can be considered to be an enzyme with a medium affinity for the L-gal substrate (**Table 1**). Furthermore, *Md*GDH has a turnover number (k_{cat}) of 4.26 s^{-1}

Table 1 Kinetics parameters for plant L-GDHs.

Enzymes	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}.s^{-1}$)	K_i (mM)	Optimal pH
Camu camu *Md*GDH[a]	0.206	4.261	20.67	1.2	7.0
Spinach *So*GDH[a]	0.128	1.158	9.06	0.5	7.0
Arabidopsis *At*GDH[b]	0.085	N/D	N/D	N/D	8.5–9
Pea *Ps*GDH[b]	0.430	N/D	N/D	N/D	8.5–9
Kiwi *Ad*GDH[c]	0.300	N/D	N/D	0.5	9.0
Spinach *So*GDH[d]	0.116	N/D	N/D	0.1	9.25

[a]L-GDH parameters according to this study.
[b]L-GDH parameters according to Gatzek et al. (2002).
[c]L-GDH parameters according to Laing et al. (2004).
[d]L-GDH parameters according to Mieda et al. (2004).

and a catalytic efficiency (k_{cat}/K_m) of 20.67 mM^{-1}/s^{-1}. These kinetic parameters are being reported here for the first time for any species so far studied. The spinach recombinant enzyme (*So*GDH) presented a K_m of 0.13 mM (**Fig. 1D**), close to that reported for *So*GDH purified from spinach leaves (0.12 mM, Mieda et al. 2004). By contrast, however, we observed an optimal pH of 7.0 (similar to *Md*GDH) while Mieda et al. reported a value of 9.25. For *So*GDH we observed a k_{cat} of 1.2 s^{-1} and a catalytic efficiency of 9.1 mM^{-1}/s^{-1} (**Table 1**). Overall, the two enzymes show kinetic parameters which are broadly similar and well within an order of magnitude of one another for all values. The discrepancy in the reported values for the optimal pH of *So*GDH is difficult to explain at this moment but it is of note that the report by Mieda et al. (2004) provides no experimental detail on how the value of 9.25 was determined. Furthermore, the authors only cite the use of Tris buffer, which would be inappropriate for investigating a wide pH range. Finally, despite quoting pH 9.25 as the optimal pH for spinach GDH, all of their kinetic studies were performed in 100 mM Tris buffer at pH 7.5.

The study of Mieda et al. (2004) also suggested feedback regulation of L-GDH by AsA, the end product of the D-mannose/L-galactose biosynthetic pathway, by competitive inhibition. According to this report, AsA at 1 mM inhibits up to 41% of *So*GDH activity with a K_i of 0.13 mM and L-GDH purified from *A. thaliana* was similarly inhibited by AsA (Mieda et al. 2004). However, a more detailed analysis of the inhibitory effect of AsA on the enzyme from kiwi fruit (*Ad*GDH) showed that the inactivation is essentially slow and irreversible and can be attributed to oxidative damage of a key amino acid residue in the active site (Laing et al. 2004).

Camu camu is a plant that stores high concentrations of AsA (2 g/100 g pulp) in its fruits (Castro et al. 2018) and consequently *Md*GDH could potentially be refractory to feedback regulation by AsA. In order to test this hypothesis, we conducted inhibition assays of *Md*GDH and *So*GDH with AsA concentrations up to 4.5 mM (**Fig. 1B**). AsA at 1.5 mM, 3 mM and 4.5 mM inhibited the activity of *Md*GDH by 53%, 71% and 79%, respectively ($K_i = 1.2$ mM). At the same AsA concentrations, the activity of *So*GDH was inhibited by 82%, 89% and 92%, respectively ($Ki = 0.5$ mM) (**Fig. 1E**), indicating that camu camu *Md*GDH has a slightly higher tolerance to inhibition by AsA than does spinach *So*GDH and kiwi *Ad*GDH.

Although AsA is only a weak acid, the concentrations used in the inhibition assays could lead to significant alterations to the pH of the reaction buffer (100 mM Tris–HCl, pH 7.0). Considering this effect and knowing that changes in pH alter the activity of L-GDH (**Fig. 1**), we performed pH measurements of the reaction buffer with each of the AsA concentrations used in the assay. We observed that on increasing the AsA concentration, the pH of the reaction solution decreased. For example, using a reaction solution with Tris–HCl pH 7.0 at 100 mM, the pH dropped to 5.9 in the presence of 4.5 mM AsA (**Supplementary Table S1**). The classical bell-shaped pH-dependence curves shown in the insets to **Fig. 1A** and **D** indicate that this pH drop of >1 pH unit would have a significant impact on enzyme activity, of the order we observed experimentally. This suggests that the inhibition observed in both camu camu and spinach L-GDH could be the effect of pH and not a genuine feedback inhibition by AsA, the pathway's final product. In order to investigate this further, inhibition assays were performed in the presence of a stronger buffer (300 mM Tris–HCl) where the pH of the reaction buffer is not drastically affected by high concentrations of AsA (**Supplementary Table S1**). Under these conditions, AsA showed no inhibitory effect on the activity of *So*GDH (**Fig. 1F**). However, in the case of *Md*GDH, mild inhibition persisted with an estimated K_i of 2.1 mM. This effect may be related to the reaction buffer itself, since when comparing the K_m values of the enzyme in the absence of AsA in the 100- and 300 mM buffers, it varies from 0.206 mM to 0.650 mM, respectively. Nevertheless, the effect produced by AsA on the enzyme activity appears to be largely due to a drop in pH in the case of both enzymes. It is worth noting that Mieda et al. (2004) also performed their study using Tris buffer at only 100 mM concentration. Our data call into question the conclusions of the study of Mieda et al. (2004) who suggest negative feedback inhibition by the final product of the pathway (AsA) as a potential regulatory mechanism of physiological relevance to plants. We find no evidence for this in the present study.

Overall description of the structure of L-GDH

To date, no crystal structures for L-GDHs have been described in the literature. In order to fill this gap in our current knowledge, we embarked upon a crystallization campaign for both of the enzymes of interest to the current study. From crystallization assays performed on camu camu *Md*GDH and spinach *So*GDH, X-ray diffraction quality crystals were obtained for *So*GDH in both its apo and holo (NAD^+ bound) forms. Diffraction data for *So*GDH and *So*GDH–NAD^+ were collected, and the structures were refined to 1.40 and 1.75 Å resolution, respectively. The good quality of the models is reflected in the values of R_{work} and R_{free}, as well as in their stereochemistry. *So*GDH has a final $R_{work} = 19.42\%$ and $R_{free} = 20.95\%$, with 97.76% of amino acids in the most favored regions of Ramachandran space. In the case

of *So*GDH–NAD^+ the equivalent values are 18.37%, 21.48% and 98.08%, respectively (**Supplementary Table S2** summarizes the full data collection and refinement statistics).

A preliminary X-ray analysis of *Os*GDH from rice (*O. sativa*) was reported in 2013 but the structure remained unsolved due to difficulties with molecular replacements (Momma and Fujimoto 2013). Nevertheless, in 2021 the same authors deposited two structures for the enzyme, one in its apo form at 1.2 Å (*Os*GDH; PDB accession: 7EZI) and one in the form of a complex with the cofactor NAD^+ at 1.8 Å (*Os*GDH–NAD^+; PDB accession: 7EZL). However, thus far, no publication describing the details of these first structures has appeared in the literature. At the time when the structures reported here were solved, the homologous structures from rice were not yet available and initially we had similar difficulties in encountering a correct molecular replacement solution. However, the structure of the apo enzyme (*So*GDH) was finally solved using *MoRDa*, a pipeline for automated molecular replacement based on a protein domain database derived from the PDB (Vagin and Lebedev 2015). The structure of the cofactor-bound complex (*So*GDH–NAD^+) was solved by molecular replacement using the apo structure as the search model with the program 'Phaser' in the 'Phenix' package (Adams et al. 2010).

The crystal structure of *So*GDH has one molecule in the asymmetric unit with a $(\beta/\alpha)_8$-barrel fold, as observed in the AKR superfamily (**Fig. 2A**) (Jez et al. 1997). It has eight parallel β-strands which are alternated with eight α-helices which run antiparallel relative to the strands, forming the classical barrel-type fold. At the N-terminus is an additional β-hairpin ($\beta1$ and $\beta2$) that forms the bottom of the barrel. The $(\beta/\alpha)_8$-barrel is conserved in the AKR superfamily; however the loops connecting the strands and helices, as well as the presence of external helices to the barrel, can vary widely and are important for the classification of AKRs into the 16 different families (Penning 2015). The loops A ($\beta6$–$\alpha4$), B ($\beta9$–$\alpha7$) and C (the C-terminal region) are the most important because they are related to cofactor binding and/or substrate specificity (Jez et al. 1997). **Fig. 3** shows a sequence alignment including five representatives of L-GDH from different species of plant together with seven AKRs from different families, which have known 3D structures available. L-GDHs have a short loop-A, similar to that seen in the AKR3, AKR5, AKR7 and AKR11 families, but different when compared to AKR1, AKR2 and AKR4. AKRs1–5 have a short loop-B, L-GDHs a medium-sized loop-B and AKR7/AKR11 a long loop-B. Loop-C is very variable among all the sequences. Like most AKRs, L-GDHs have two α-helices external to the barrel, the H1 helix within loop-B and the H2 helix between $\alpha8$ and loop-C (both helices are conserved in most AKRs). The lack of a description of the 3D structure for L-GDH up until now has precluded its adequate classification within the AKR superfamily.

NAD^+ binding

AKRs are characterized by binding pyridine nucleotide coenzymes, commonly NAD(P). Likewise, they present a conserved oxidation–reduction mechanism using a catalytic tetrad composed of highly conserved amino acids (Asp, Lys, Tyr and His), having specificity for sugar or steroid substrates (Jez et al. 1997). It is well established that L-GDH has specificity for NAD^+ as its cofactor (rather than the more common $NADP^+$) and that the sugar L-gal is its preferred substrate (Gatzek et al. 2002). In order to fully explore the differences between L-GDH and other members of the AKR protein superfamily, it was also necessary to solve the structure of *So*GDH in the form of its complex with NAD^+. This structure will be referred to here as *So*GDH–NAD^+.

The crystal form of *So*GDH–NAD^+ had two molecules in the asymmetric unit. The polder map (Liebschner et al. 2017) for one of the NAD^+ cofactors is given in **Supplementary Fig. S3A**, where the quality of the electron density is clearly visible for the entire molecule. Protein Interfaces, Surfaces, and Assemblies (PISA) analyses (Krissinel and Henrick 2007) indicated that the apparent homodimer is not expected to exist under physiological conditions and is merely the result of crystal packing. This is consistent with the SEC-MALS results presented previously. Furthermore, SEC analysis of the enzyme under either acidic conditions (as used for crystallization) or in the presence of NAD^+ indicated it to be monomeric under all conditions investigated (**Supplementary Fig. S2C**).

To the present date, the only members of the AKR superfamily which are known to be homodimers are aflatoxin dialdehyde reductase (AKR7A1), NAD(P)H-dependent xylose reductase (AKR2B5) and the tyrosine auxotrophy suppressor protein (Kozma et al. 2002, Kavanagh et al. 2003, Obmolova et al. 2003). A similar dimerization interface using helices $\alpha5$, $\alpha6$, H2 and loop-C is observed in all of these structures (**Supplementary Fig. S3B**). On the other hand, in *So*GDH–NAD^+ the contact interface between the two monomers of the asymmetric unit is completely different and involves the loop $\beta5$–$\alpha3$, loop-B and loop-C of chain A and the helices $\alpha5$, $\alpha6$ and loop $\alpha5$–$\beta8$ of chain B (**Supplementary Fig. S3B**). Consequently, the two molecules in the asymmetric unit of *So*GDH–NAD^+ interact merely as the product of crystal contacts and have no physiological relevance. This observation is reinforced by the apo structure of *So*GDH, which crystallizes with a monomer in the asymmetric unit in a space group that lacks 2-fold rotation axes and therefore dimers are impossible. The sum total of the evidence presented here indicates that spinach *So*GDH is a monomeric protein (at least under the conditions used in the present work) and not a homodimer as reported previously (Mieda et al. 2004).

*So*GDH has the conserved catalytic tetrad observed in AKRs: Asp57, Tyr62, Lys90 and His127 (**Fig. 4**). The tetrad is also part of a group of residues that interact directly with NAD^+ (**Fig. 2B**). This strongly suggests that L-GDH follows the same basic catalytic mechanism described for other AKRs but with a preference for favoring the oxidation (dehydrogenation) reaction rather than reduction (see below). This leads to the production of L-galactono-1,4-lactone from L-galactose with the concomitant reduction of NAD^+ to NADH.

AKRs in general have a group of conserved residues responsible for NAD(P) binding, the most conserved being Asp57,

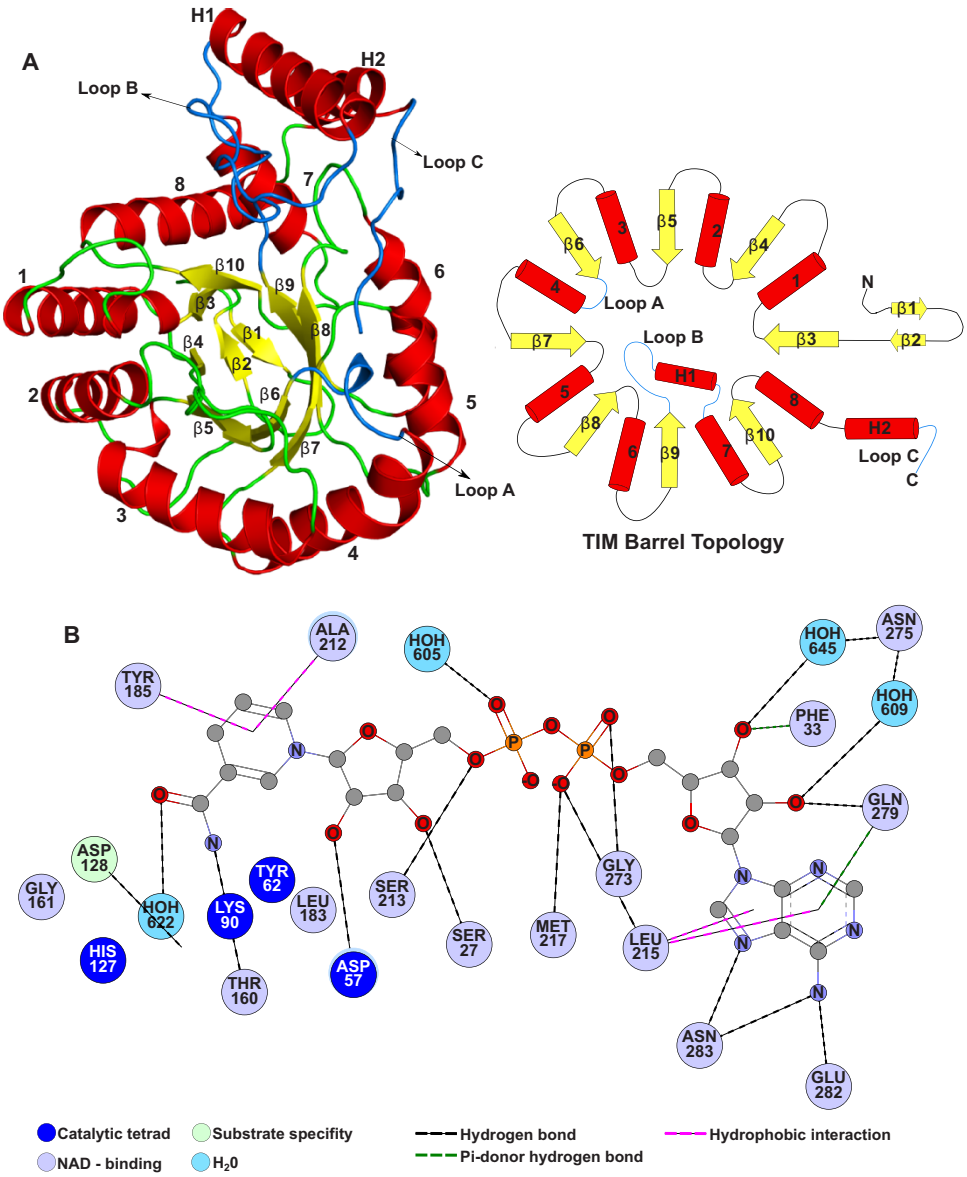

Fig. 2 *So*GDH fold and NAD^+-binding interactions. (A) Cartoon representation (left) and topology diagram (right) for the structure of *So*GDH, which shows a $(\beta/\alpha)_8$-barrel (TIM-barrel) fold, characteristic of the AKR superfamily. This fold has eight β-strands interspersed by eight α-helices. At the bottom of the barrel, at the N-terminus of the polypeptide chain, there is a β-hairpin composed of strands β1 and β2. In addition, *So*GDH presents two helices (H1 and H2) external to the barrel, which together with the loops loop-A, loop-B and loop-C are important for the classification of the members of AKRs. (B) The network of interactions with the NAD^+ cofactor. Figures were generated with PyMol v2.05 (Schrödinger, LLC, San Diego, CA, USA) and Discovery Studio Visualizer V21.1.0 (BIOVIA, Dassault Systèmes, Waltham, MA, USA).

Ser160, Asn161 and Gln183 (**Fig. 3**) (Bennett et al. 1997, Jez et al. 1997). Of these residues, *So*GDH retains only Asp57, while at position 160 it has a Thr, at 161 a Gly and at 183 a Leu. These changes lead to the loss of two hydrogen bonds with the nicotinamide ring, forming new van der Waals interactions with Gly161 and Leu183 (**Fig. 5A** e **B**). The substitution of the aromatic residue at position 212 by Ala in *So*GDH also stands out (**Fig. 3**). These aromatic amino acids in AKRs occupy a position on the underside of the nicotinamide ring as seen in **Fig. 5A** and are a notable feature of the cofactor-binding site (Bennett et al. 1997, Jez et al. 1997). Given its structural and functional importance, the lack of this aromatic residue in L-GDH is, at first glance, surprising. However, in L-GDHs, the pi-stack interaction is recovered with Tyr185, a conserved amino acid in these enzymes (**Figs. 5B** and **3**). This appears to be a classical example of a compensating substitution. A hydrogen bond between Thr27 and the ribose ring on the nicotinamide side of the cofactor in AKRs is maintained via Ser27. Additionally, in

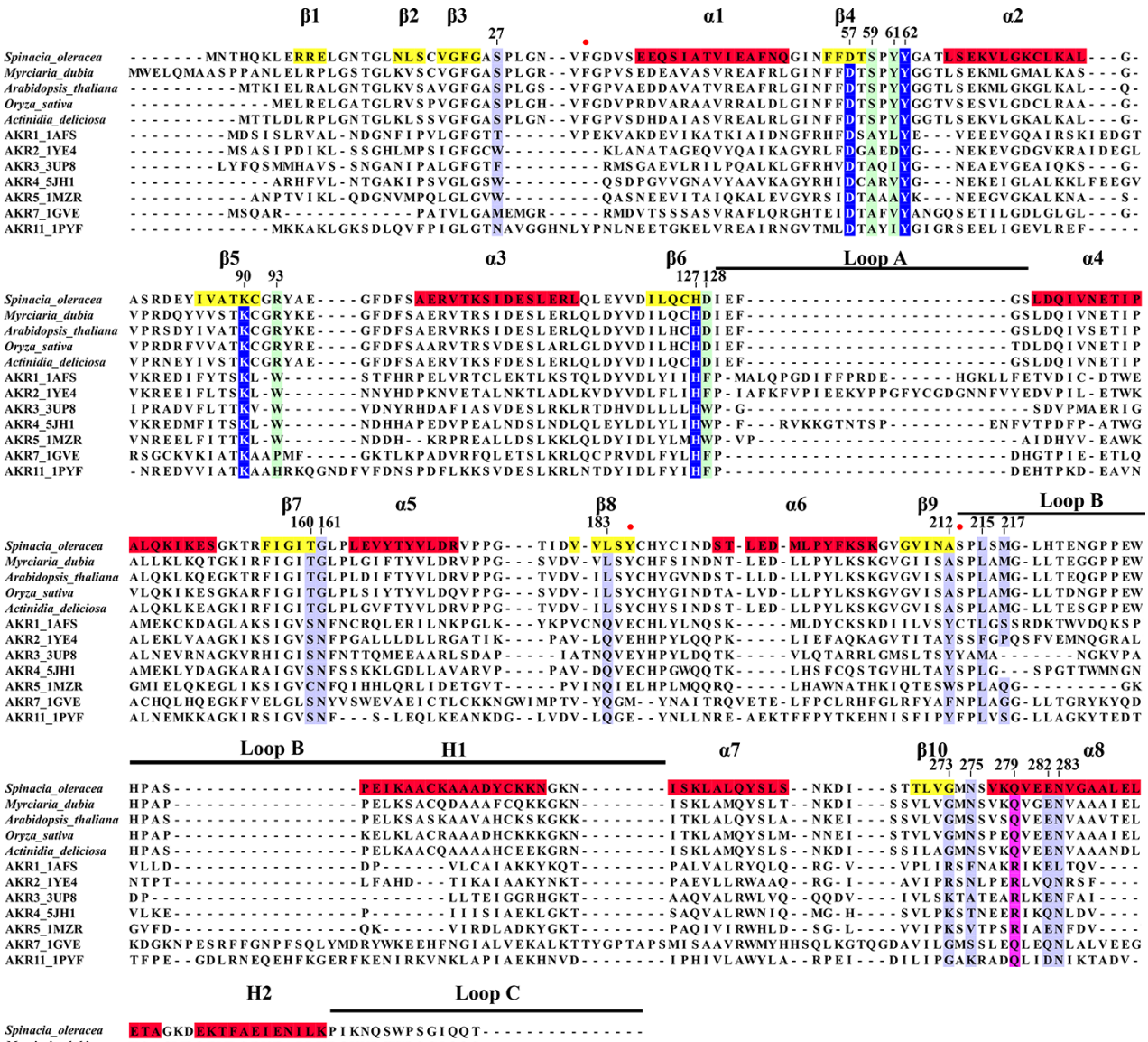

Fig. 3 Multiple sequence alignment between L-GDHs and AKRs. Five L-GDH sequences from different species were aligned with seven representatives of different families of AKRs. Using the spinach SoGDH sequence, the position of secondary structure elements are indicated. Likewise, the numbering adopted for the description of the amino acid positions was based on *So*GDH. Highlights: in dark blue the amino acids of the catalytic tetrad, in light blue those responsible for binding to NAD^+, in light green the amino acids important in determining substrate specificity and in purple a conserved position in ARKs that confers a preference for the cofactor (NAD^+ or $NADP^+$). Positions with a red ball represent residues that interact with NAD^+, which are only observed in the L-GDH structures presented here.

*So*GDH, a new hydrogen bond is observed between Ser213 and the phosphate group on the nicotinamide side of the cofactor (P_N) (**Fig. 5B**).

Changes in the interaction between the enzyme and the cofactor are also evident on the adenine side of the NAD^+. Lys/Arg273 and Arg279 confer $NADP^+$ specificity to the majority of AKRs (**Fig. 5C**) (Jez et al. 1997). In *So*GDH, position 273 is replaced by a Gly and 279 by a Gln (**Fig. 5D**). In AKRs, Lys/Arg273 forms a hydrogen bond with P_N (the phosphate on the nicotinamide side of the cofactor). However, in *So*GDH,

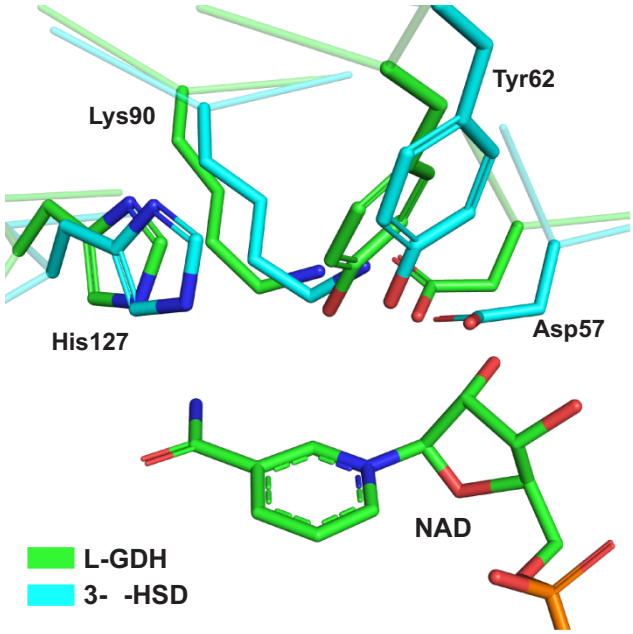

Fig. 4 The active site in *So*GDH. A superposition between the active site of *So*GDH and 3-α-hydroxysteroid dehydrogenase (a representative of the AKR1 family) shows the catalytic tetrad (Asp27, Tyr52, Lys90 and His127) conserved in L-GDH. 3-α-HSD. PDB code: 1AFS.

Gly273 makes a hydrogen bond, using its main chain to the phosphate on the adenine side (P_A). However, the most important appears to be the loss of Arg279 in GDHs (**Fig. 3**). Arg279 forms hydrogen bonds with the P_{2B} phosphate moiety on the adenine-side ribose, a specific interaction for enzymes which employ $NADP^+$ as the cofactor (**Fig. 5C**) (Jez et al. 1997). In *So*GDH, the lack of Arg279 means that the enzyme is unable to compensate the formal charge on P_{2B} rendering L-GDHs specific for NAD^+ rather than $NADP^+$. Nevertheless, the substitution by Gln still allows for direct interaction with one of the ribose hydroxyl groups on the adenine side (**Fig. 5D**).

In AKRs, residue 275 is highly variable (**Fig. 3**) but this does not influence its interaction with P_{2B} because this occurs via its main chain amine (Jez et al. 1997) (**Fig. 5C**). However, in GDHs, the homologous residue, Asn275, is highly conserved, and due to the absence of P_{2B}, Asn275 can interact with the ribose moiety via two water molecules (**Fig. 5D**). In the same region of *So*GDH, a new interaction is observed. Phe33 makes a hydrophobic contact and a pi-donor hydrogen bond with the ribose on the adenine side (**Fig. 5D**). Furthermore, Leu283, which makes hydrophobic contact with the adenine in AKRs, is substituted by Asn in *So*GDH. This allows for the formation of two additional hydrogen bonds with the adenine ring, further stabilizing the cofactor (**Fig. 5D**).

Overall, as **Fig. 5** makes clear, the nicotinamide side of the cofactor-binding site is much more conserved than the adenine side. In the case of the former, this is not unexpected given the need, in both cases, for the binding-site cavity to maintain the orientation of the nicotinamide moiety oriented for productive catalysis. However, in the case of the latter, although some aspects of the binding site are preserved when comparing AKRs and L-GDHs, overall the cavity has been radically altered in order to accommodate NAD^+ rather than $NADP^+$. The preference for NAD^+ as the cofactor for L-GDH is probably related to the direction of the reaction being catalyzed which, in order to generate AsA as the final product, must be in the oxidative direction. As pointed out by Barski et al. (2008) this is different to most AKRs which generally favor reduction. These enzymes use NADPH as the cofactor that is more abundant in the cytoplasm than the oxidized form $NADP^+$, thereby favoring the reduction reaction (Barski et al. 2008). For $NADH/NAD^+$ the reverse is true with NAD^+ being the predominant species. By preferring NAD^+ as the cofactor, L-GDH favors the oxidative production of L-galactono-1,4-lactone from L-galactose and thereby the effective production of ascorbic acid.

Substrate specificity

Thus far, we have been unsuccessful in obtaining crystals of *So*GDH bound to the natural substrate (L-gal), a substrate analog or an inhibitor. Nevertheless, comparisons with crystal structures of AKRs complexed to inhibitors are still possible in order to investigate the structural basis for differences in substrate specificity. For this purpose, we will use as an example of AKRs 3-α-hydroxysteroid dehydrogenase (3-α-HSD) and human aldose reductase (ADR), both bound to inhibitors (**Fig. 6**). AKRs present amino acids within the substrate-binding pocket, which are responsible for differentiating between a wide range of steroids and sugar substrates (light green in **Fig. 3**) (Bennett et al. 1997, Jez et al. 1997). **Fig. 6** shows that in the active site of the three enzymes, Asp57, Lys90, Tyr62 and His127 (the catalytic tetrad) are fully conserved (**Fig. 6**). Trp228 from loop-B is essential for stability rather than substrate discrimination and is also conserved in all three proteins (Bennett et al. 1997, Jez et al. 1997). Furthermore, Ala59 and Trp93 play the same role in substrate stabilization in both 3-α-HSD and ADR (**Fig. 6A** and **B**). However, in *So*GDH these positions are occupied by Ser59 and Arg93, which could be important for substrate discrimination since they have the potential to form new hydrogen bonds with the hydroxyl groups on the L-gal substrate (**Fig. 6C**). Amino acids at positions 61 and 128 are involved in the discrimination between sugars and steroids (Jez et al. 1997). At position 61, hydroxysteroid dehydrogenases (HSDs) have Leu or Ile, while ADRs have Val. Differences in this position can alter the topology of the binding site to accommodate substrates of different sizes. Furthermore, at position 128, HSDs have Phe while ADRs have a conserved Trp. The presence of the indole nitrogen allows discrimination between the sugar and the steroid, controlling the accessibility of the substrate or its orientation in the active site (**Fig. 6A** and **B**) (Jez et al. 1997). Spinach *So*GDH exhibits what would appear to be highly significant differences at these two positions, which are occupied by Tyr61 and Asp128 (**Fig. 6C**). Tyr61 could potentially perform a stacking interaction with the sugar ring or hydrogen bonds with

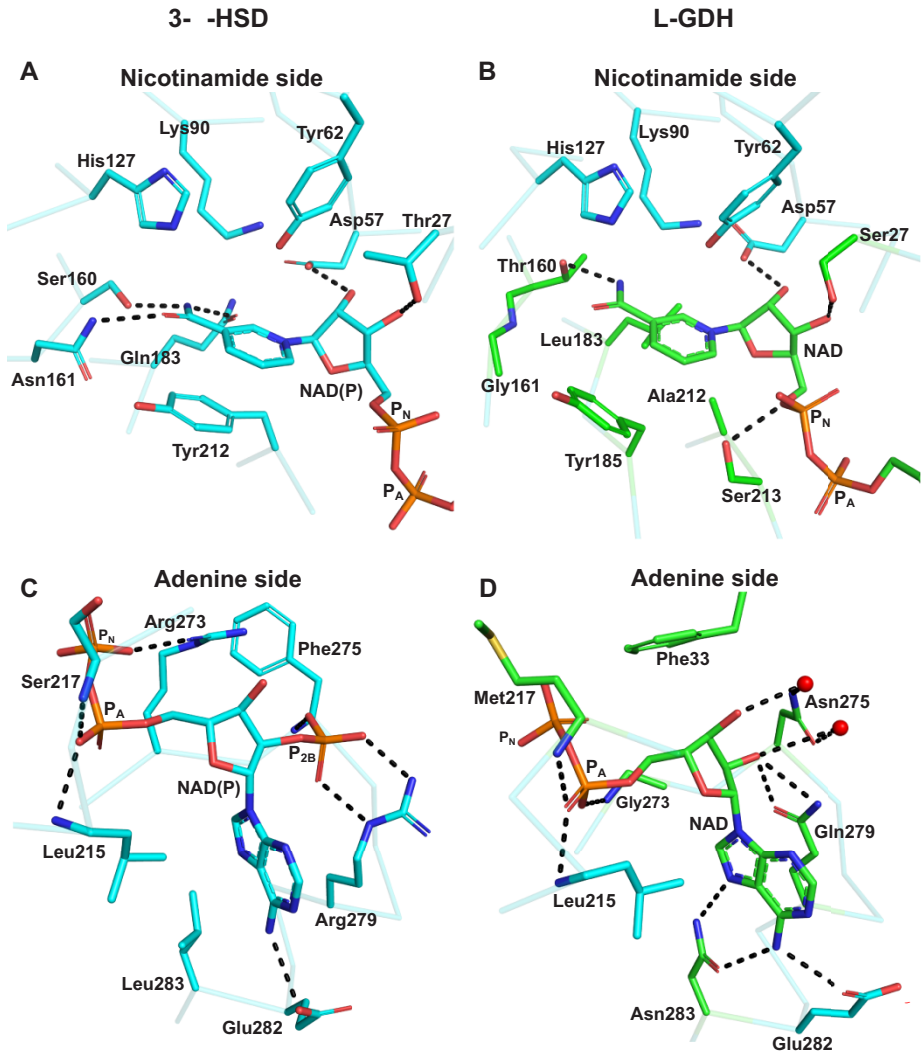

Fig. 5 Comparison between the $NADP^+$-binding interactions in AKRs and NAD^+ binding in *So*GDH. For a better understanding, the interactions have been divided into the nicotinamide and adenine sides which are treated separately. (A and C) The interactions made with $NADP^+$ in 3-α-HSD (PDB identifier: 1AFS). On the nicotinamide side (A), three hydrogen bonds are observed between Ser160, Asn161 and Gln183, which together with the pi-stacking formed by Tyr212, correctly orient the nicotinamide ring of $NADP^+$ for catalysis. On the adenine side (C), the interaction between Arg279 (conserved in AKRs) and the P_{2B} phosphate stands out and confers specificity for $NADP^+$. On the nicotinamide side of *So*GDH (B), the hydrogen bonds are lost, as well as the pi-stack at position 212. However, the orientation of the nicotinamide ring is maintained by the hydrophobic interaction of Leu183 and the pi-stacking of Tyr18. In addition, a hydrogen bond is maintained between Thr160 and the nicotinamide ring. On the adenine side (D), we observe that the absence of P_{2B} allows for interaction via water molecules between Asn275 and the ribose. Likewise, Gln279 interacts directly with the ribose. A hydrophobic interaction involving Phe33 and two new hydrogen bonds between Asn283 and the adenine appear in *So*GDH. P_N: phosphate on nicotinamide side of the cofactor; P_A: phosphate on adenine side.

the substrate via its hydroxyl group. Asp128, on the other hand, has the potential to form a pair of hydrogen bonds with adjacent hydroxyls on L-gal, aiding in establishing the correct orientation of the substrate while discriminating between different sugars. Furthermore, galactose-binding enzymes often use an aromatic amino acid (Trp, Phe or Tyr) to correctly orient the galactose substrate within the active site (Sujatha et al. 2004). GDHs, including *So*GDH, have a conserved Tyr185, which differentiates them from the remaining AKRs (**Fig. 3**). In addition to restoring the interaction with the NAD^+ nicotinamide ring, this tyrosine could also play a role in establishing the specificity of the enzyme for L-galactose and orienting it correctly in the binding pocket (**Fig. 6C**).

Loop-C is essential for determining substrate specificity in AKRs (Jez et al. 1997). However, in both forms of *So*GDH reported here, loop-C is not oriented toward the active site due to the absence of substrate and we are so far unable to shed light on its role in the case of GDHs. In summary, despite the absence of a complex with the substrate or a substrate analog, the structures presented here provide the basis for testing, by site-directed mutagenesis, those residues which are most likely to be involved in substrate recognition and selectivity.

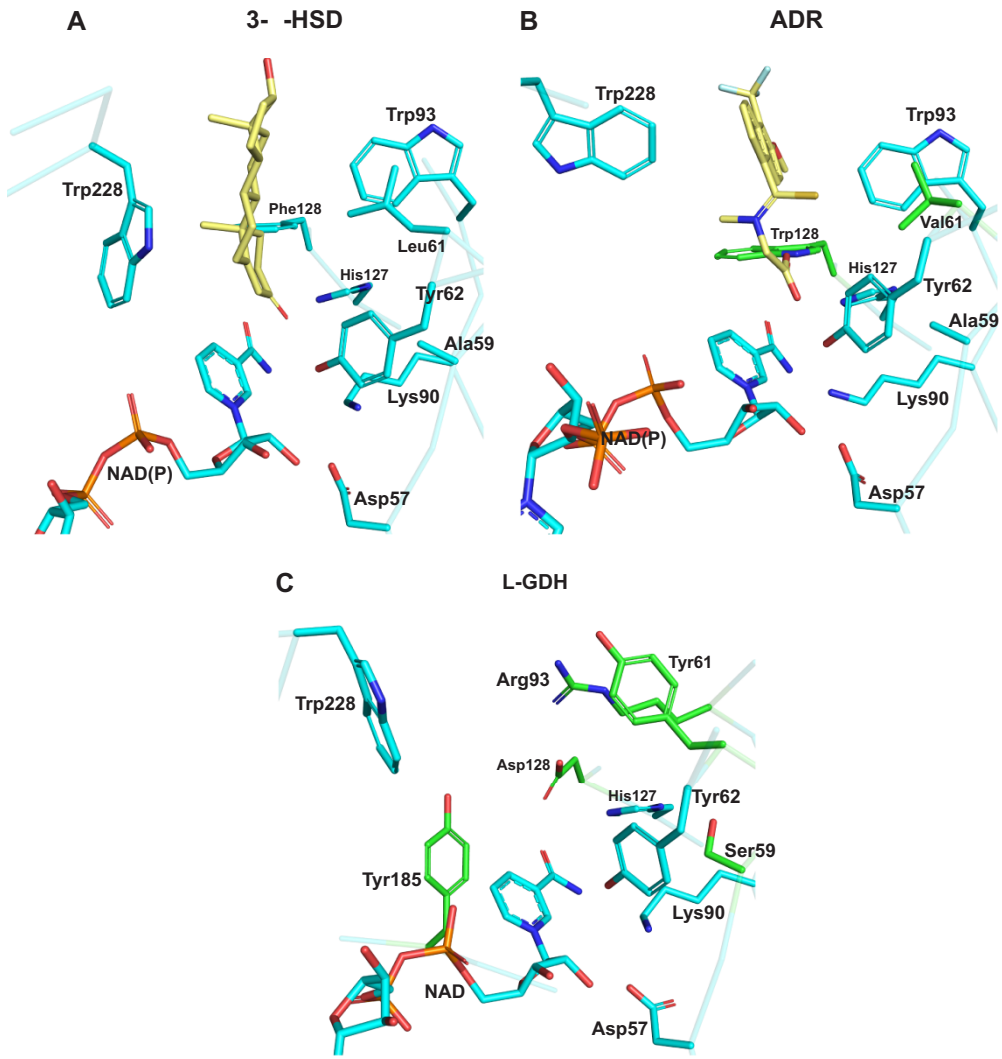

Fig. 6 Substrate specificity in AKRs and *So*GDH. (A) 3-α-HSD bound to testosterone, showing amino acids important for substrate binding. Leu61 and Phe128 are specific for AKRs that bind steroids. (B) Human ADR bound to the inhibitor tolrestat. Val61 and Trp128 are specific for AKRs that bind sugars. (C) The empty active site as observed in the *So*GDH–NAD^+ complex. Despite *So*GDH having a sugar as its natural substrate, it presents a Tyr at position 61 and an Asp at 128. It also has differences in positions 59 (Ser) and 93 (Arg). Tyr185 appears to be specific for L-GDH, since galactose-binding proteins use an aromatic amino acid to orient the ligand to the active site. 3-α-HSD PDB code: 1AFS; ADR PDB code: 2FZD.

Despite the conservation of the $(\beta/\alpha)_8$ fold, the significant differences observed at both the cofactor- and substrate-binding sites would appear to justify the classification of GDHs as a new family within the AKR superfamily. This is borne out by a phylogenetic analysis of representative sequences of the existing families taken from the University of Pennsylvania protein kinase (PKR) database (https://hosting.med.upenn.edu/akr/), which shows GDHs to cluster into a monophyletic group (**Supplementary Fig. S4**).

Open and closed state

AKRs have an NAD(P)-dependent opening and closing mechanism, which allows for the formation of a tunnel that maintains NAD(P) bound. This transitory structure is formed after cofactor binding by a conformational change in loop-B and loop β3–α1 (Bennett et al. 1997). The interactions that allow tunnel formation are different in HSDs and ADRs. ADRs differ by the presence of a salt bridge involving Asp227, Lys35 and Lys273 and hydrophobic interactions made by Trp27 (loop-β3–α1) and Pro226–Trp228 (loop-B). Due to the mutation of Asp227 to a Ser, HSDs lack the salt bridge that helps stabilize the closed state. Consequently, in HSDs, the tunnel is stabilized by different interactions formed between several residues from loop-β3–α1 (Pro33, Glu34 and Lys35) and loop-B (Asp225, Lys226, Thr227 and Trp228) (Bennett et al. 1997).

Spinach *So*GDH presents both the open and closed states as a function of NAD^+ binding. An overlay of the two structures yields an root mean square deviation (RMSD) of 0.66 Å for 318 aligned Cα atoms with the regions showing the greatest

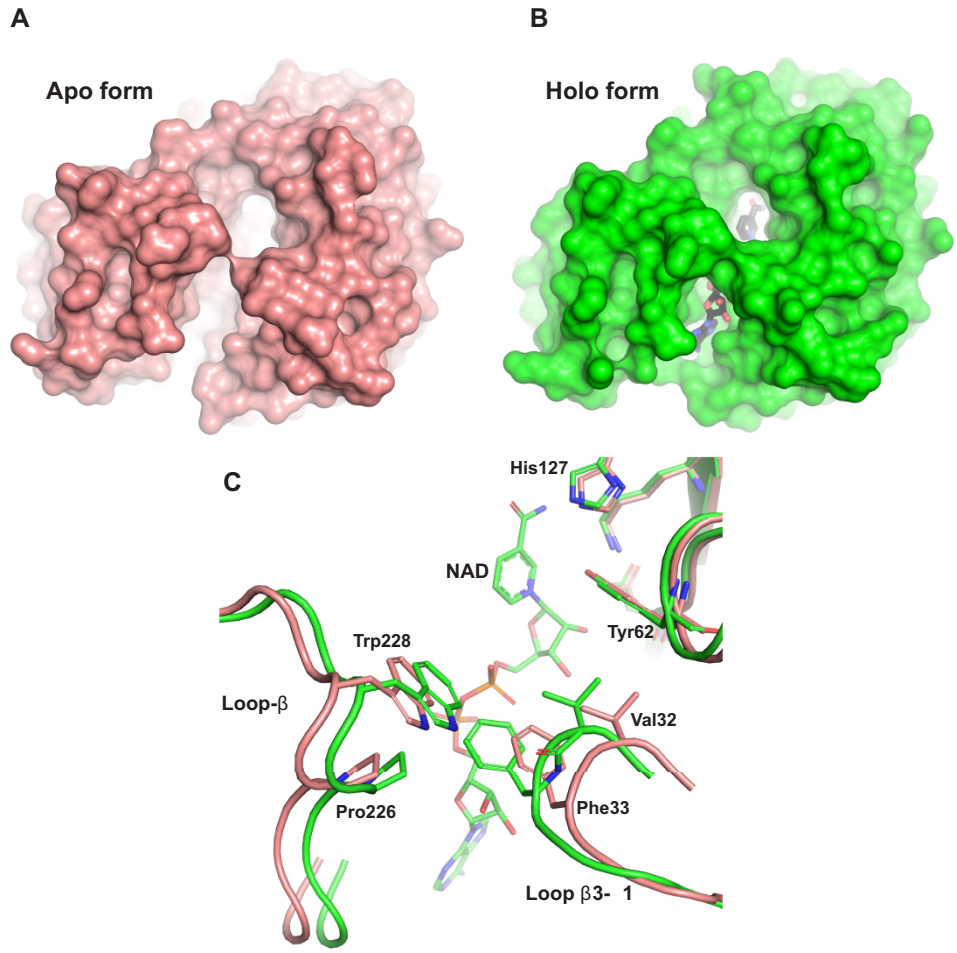

Fig. 7 The open and close states of *So*GDH. (A) The apo form of *So*GDH shows the open state, with little contact between loop-B and loop-β3–α1. (B) The holo form of *So*GDH–NAD^+ shows the closed state, forming a tunnel generated by the contact made between loop- B and loop-β3–α1 (NAD^+ is shown within the tunnel at the bottom). (C) A superposition between SoGDH apo and holo forms shows a close-up of the loop-B and loop-β3–α1 regions. NAD^+-binding results in hydrophobic contacts between Val32, Phe33, Pro226 and Trp228, as well as hydrogen bonding between the indole nitrogen of Trp228 and the Val32 main chain.

divergence being loop-β3–α1 and loop-B (important for tunnel formation) (**Fig. 7**). Despite *So*GDH having a sugar substrate, like ADRs, they do not have a salt bridge in the tunnel because at position 273, a Gly substitutes for the Lys/Arg conserved in ARKs. Tunnel formation and stabilization are due to hydrophobic interactions between Val32–Phe33 (loop-β3–α1) and Pro226–Trp228 (loop-B), as well as a hydrogen bond between the indole nitrogen of Trp228 and the Val32 main chain (**Fig. 7C**). Although there is a significant variation among AKRs and GDH in the details of the interactions responsible for tunnel formation, this appears to be a common mechanism for all such enzymes indicating that NAD^+ binding induces a conformational change to the enzyme which is presumably essential for competent catalysis.

L-GDH structure comparison

The primary structures of L-GDHs show high levels of sequence identity among different plant species (percentage identity of 76%), preserving key amino acids for NAD^+ binding and those deduced from the discussion above to be important for L-galactose substrate specificity (**Fig. 3**). Therefore, they would be expected to have a well-conserved 3D structure, independent of species.

The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) currently only has available the 3D structures of rice *Os*GDH in its apo (7EZI) and holo (7EZL) states. However, there is no published description of these structures. As anticipated, rice *Os*GDH has the highly conserved $(\beta/\alpha)_8$-barrel fold similar to spinach *So*GDH. The apo form of both proteins has an RMSD of 0.45 Å for Cα atoms and the holo-structures 0.46 Å. Significant differences only exist in the loop-β3–α1 and loop-β4–α2 regions (**Fig. 8A**). The importance of the loop-β3–α1 region for tunnel formation and the fact that the amino acid sequences in these regions are almost identical in both enzymes make these differences to be of interest and worthy of further investigation. At first glance, such similar sequences would not be expected to adopt such radically different conformations.

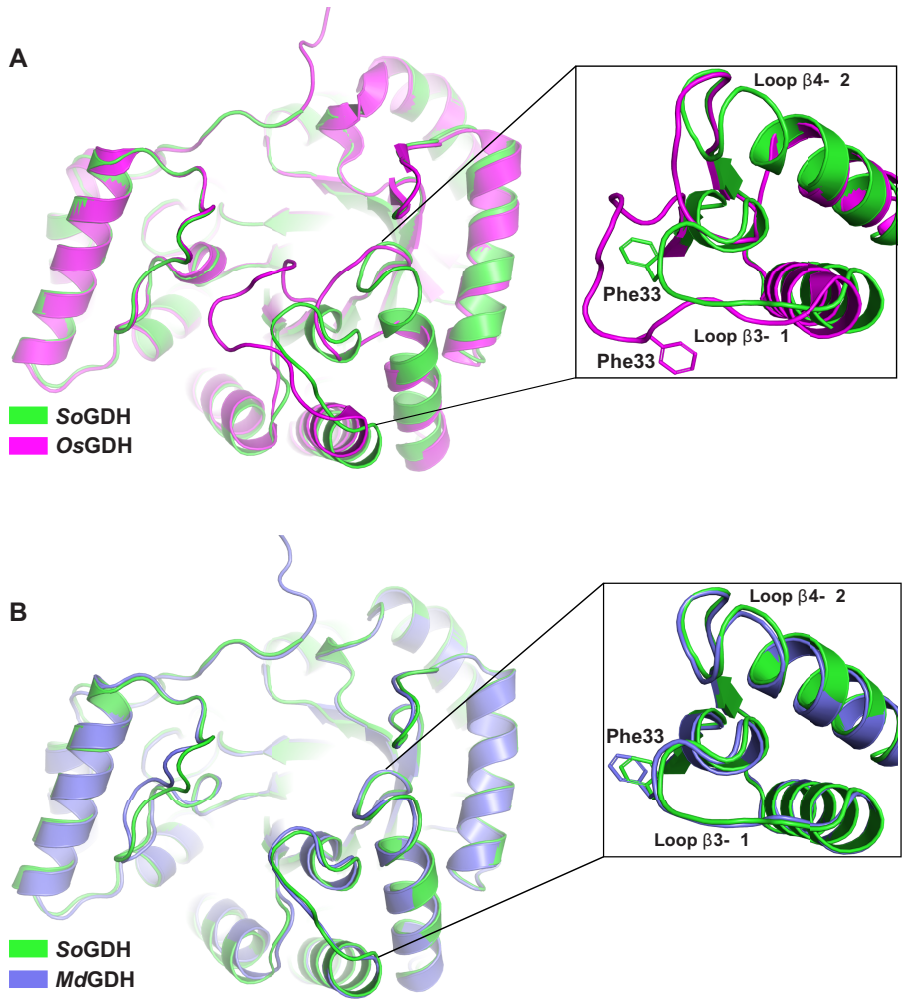

Fig. 8 L-GDH structure comparisons. (A) A superposition between spinach *So*GDH and rice *Os*GDH crystal structures in the apo form. Despite having an RMSD of 0.45 Å for Cα atoms, the orientations of loop-β3–α1 and loop-β4–α2 are very different (with a displacement of approximately 6 Å). (B) Superposition between the crystal structure of spinach *So*GDH and the Alphafold2 model for camu camu *Md*GDH. loop-β3–α1 and loop-β4–α2 show the same orientation. Rice *Os*GDH PDB code: 7EZI.

The different conformation of loop-β3–α1 in the rice enzyme leaves Phe33, necessary for tunnel formation, in an orientation opposite to that observed in spinach *So*GDH (**Fig. 8A**, inset). In addition, the shift in the position of loop-β3–α1 (∼6 Å) toward loop-B induces the formation of a tunnel even in the apo form, something not normally observed in AKRs or in *So*GDH (**Fig. 8A**). The quality of the electron density map in loop-β3–α1 and loop-β4–α2 of spinach *So*GDH in both the apo and holo forms leaves no room for ambiguity and shows that the conformation reported for the loop-β3–α1 region has been correctly interpreted (**Supplementary Fig. S5**). However, analyzing the electron density map of *Os*GDH it becomes evident that this region could have been misinterpreted. Furthermore, it appears to have a knock-on consequence for the neighboring loop β4–α2, which also adopts a conformation radically different to that seen in *So*GDH.

Since we have thus far been unsuccessful in obtaining diffraction quality crystals of GDH from camu camu (*Md*GDH) we performed its structure prediction using Alphafold2 (Jumper et al. 2021). The resulting model for camu camu *Md*GDH presents a very similar structure to that of *So*GDH with an RMSD of 0.56 Å for Cα atoms. In *Md*GDH, an open state is observed with loop-β3–α1 and loop-β4–α2 in the same positions as observed in spinach *So*GDH (**Fig. 8B**) and very different to that seen in the structure for rice *Os*GDH. Well there are no notable differences between *So*GDH and *Md*GDH in the vicinity of the substrate- and NAD^+-binding sites, implying that it would be expected that they would present similar kinetic properties as well observe experimentally (**Table 1**).

Notwithstanding the fact that AlphaFold2 predictions are still under scrutiny by the scientific community, our results strongly suggest that the conformation described here for the

loops in apo *So*GDH is correct and would be expected to be conserved in other GDHs. Since this region is involved in tunnel formation and therefore is of functional significance, it appears that the structural interpretation of this region in apo *Os*GDH is worthy of revisiting.

Conclusion

In this work, we contribute to closing the knowledge gap about functional and structural aspects of critical enzymes involved in vitamin C biosynthesis in plants, particularly of the D-mannose/L-galactose pathway. In this context, we report enzyme kinetic parameters (e.g. k_{cat}, catalytic efficiency, etc.) for L-GDH from two different plant species and describe its crystal structure for the first time in both its apo and holo forms. This reveals the structural basis for the preference for NAD^+ as the cofactor (important for driving the reaction in the direction of ascorbic acid synthesis) and that the enzyme undergoes a similar conformational change on cofactor binding to that observed in AKRs in general. This leads to the burial of NAD^+ inside a binding-site tunnel with the nicotinamide directed toward the region of the catalytic tetrad. Despite the absence of a crystal structure of the enzyme with L-gal, it has been possible to identify residues which are likely involved in imbuing selectivity for its natural substrate. These include Tyr61, Tyr185, Ser59 and Asp128, and their identification opens up new perspectives for future experiments.

The crystal structure of spinach *So*GDH shows its great similarity with aldo-keto reductases, which is based in the classical $(\beta/\alpha)_8$-barrel fold. This reveals a common evolutionary origin despite the preference for most such enzymes to favor reductive catalysis rather than oxidation, as is the case for GDHs. Oxidative catalysis justifies the use of NAD^+ as the cofactor rather than $NADP^+$, and the crystal structure presented here provides a rational explanation for such a preference. On the other hand, the origin of L-gal selectivity remains speculative at this stage due to the absence of a crystal structure with a bound substrate or analog.

The regulatory mechanism behind the control of AsA synthesis in plants is an important open question. According to our results, both *Md*GDH and *So*GDH are refractory to inhibition by AsA, indicating that in camu camu and spinach plants, the catalytic activity of L-GDH is not regulated by feedback inhibition. Previous reports to the contrary may have been artifacts introduced by a change in pH rather than competitive inhibition.

Materials and Methods

Plasmid construction

The protein-coding sequence for L-galactose dehydrogenase from *M. dubia* (*Md*GDH) (GenBank accession no. OK632632.1) was optimized using the GeneOptimizer software suite (Raab et al. 2010), then synthetized by GeneArt® (Invitrogen, Carlsbad, CA, USA) and ligated into the plasmid expression vector pET151/D-TOPO. The synthetic gene for L-GDH from spinach (*S. oleracea*, *So*GDH) (GenBank accession no. AB160990.1) was ligated into the plasmid expression vector pET-28a, which was purchased from FastBio (São Paulo, Brazil).

Protein expression and purification

Escherichia coli BL21 Rosetta™ (DE3) cells harboring the constructs were grown at 37°C in Luria-Bertani medium supplemented with ampicillin (50 $\mu g \cdot ml^{-1}$) and chloramphenicol (34 $\mu g \cdot ml^{-1}$) for *Md*GDH and kanamycin (30 $\mu g \cdot ml^{-1}$) and chloramphenicol (34 $\mu g \cdot ml^{-1}$) for *So*GDH. Once the $OD_{600\ nm}$ by optical density at 600 nm ($OD_{600\ nm}$) = 0.5–0.6 was reached, the culture was cooled to 20°C and the protein expression was induced by addition of 0.3 mM isopropyl 1-thio-β-D-galactopyranoside. After 16 h, cells were harvested by centrifugation at 10,000 g for 45 min at 4°C and suspended in lysis buffer (150 mM NaCl and 20 mM Tris–HCl pH 7.5). The soluble fraction, after cell lysis by sonication, was isolated by centrifugation at 16,000 g for 45 min at 4°C. This was then loaded onto a column with 5 ml His60 Ni-Superflow Resin (Takara Bio USA, San Jose, CA, USA) previously equilibrated in the lysis buffer. Subsequently, the resin was washed with five column volumes of lysis buffer and then again with five column volumes of lysis buffer containing 50 mM imidazole. Bound proteins were eluted from the resin using two column volumes of lysis buffer containing 250 mM imidazole. The next purification step, SEC, was performed using a Superdex 200 XK16 column (GE Healthcare, Madison, WI, USA) pre-equilibrated in lysis buffer. The purity of the eluted enzymes recovered from both purification steps was evaluated by SDS-PAGE. The desired concentration of proteins was achieved by centrifugation at 800 g, using an Amicon® Ultra (molecular weight cut-off 30 kDa) centrifugal filter device (Merck Millipore, Darmstadt, Germany). Samples were kept frozen at −80°C for future use.

SEC was also used to investigate possible changes to the oligomeric state of the enzyme as a function of pH and the presence of NAD^+. For this purpose, a Superdex 200 10/300 GL column was employed with a protein sample at 2 mg/ml. For these experiments, the running buffer was either 150 mM NaCl and 20 mM Tris–HCl pH 7.5 or 150 mM NaCl and 100 mM Bis–Tris–HCl pH 5.5 in the absence or presence of NAD^+ at a final concentration of 2 mM.

Size exclusion chromatography coupled with multi-angle light scattering

The oligomeric state of *Md*GDH and *So*GDH was evaluated by SEC-MALS using a three-angle light scattering detector miniDAWN® TREOS® (Wyatt Technology, Santa Barbara, CA, USA) and Optilab® T-rEX differential refractometer (Wyatt Technology). For the SEC, this system was coupled to an HPLC (Waters, Milford, MA, USA) consisting of a pump and controller (Waters 600). A total of 50 µl of each sample (at a concentration of 3 $mg \cdot ml^{-1}$) were loaded onto either Superdex 75 or 200HR 10/300 GL columns (GE Healthcare), equilibrated in 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.8), 300 mM NaCl and 5 mM $MgCl_2$. The data collection and analysis were performed via Wyatt ASTRA 7 software (Wyatt Technology).

Enzyme assays

Enzyme kinetic properties of camu camu *Md*GDH and spinach *So*GDH were performed in a SpectraMax Plus 384 Microplate Spectrophotometer (Molecular Devices, Sunnyvale, California, USA). The conversion of NAD^+ to NADH ($\varepsilon = 2{,}650 M^{-1}cm^{-1}$) was monitored spectrophotometrically, in a 96-well UV-transparent microplate, by recording the absorbance at 324 nm every 6 s for 5 min at 26°C. Initial enzyme velocities were estimated as the maximum linear rates of absorbance increase at 340 nm. Milliarbitrary units (mAU) values were converted to molar concentration using the Beer–Lambert equation [a], where *A* is the absorbance, ε is the molar extinction coefficient of NADH at 324 nm, [a] is the optical path length in cm and *c* is the molar concentration of NADH. Then, these values were converted to reaction rate units (concentration/time).

Enzyme optimal pH determination The optimal pH for *So*GDH and *Md*GDH was determined in a pH range from 4 to 11. Catalytic activity was measured as mentioned above, in 200 μl reaction mixture in a universal buffer (a mixture of three triprotic acids: citric acid, boric acid and phosphoric acid) (Ganesh et al. 2017) containing 200 μM NAD^+, L-galactose (L-gal, 1 mM for *So*GDH and 2 mM for *Md*GDH), and the purified recombinant enzyme was added to a final concentration of 100 nM.

Enzyme kinetics analysis Recombinant *So*GDH and *Md*GDH were assayed in a 200 μl reaction mixture containing 100 mM Tris–HCl pH 7.0, 200 μM NAD^+, L-gal (1 mM for *So*GDH and 2 mM for *Md*GDH) and purified enzyme at 100 nM. Enzyme kinetic parameters (K_m, V_{max}, V_{max}/K_m and k_{cat}) were determined by nonlinear regression fitting of experimental data using Prism—GraphPad v7.00 software.

To evaluate the inhibitory effect of L-ascorbic acid on L-GDH activity, a 200 μl reaction mixture containing 100 mM Tris–HCl pH 7.0, 200 μM NAD^+, L-gal (1 mM for *So*GDH and 2 mM for *Md*GDH) and 100 nM purified enzyme in the presence of varying AsA concentrations (0.75, 1.5, 3.0 and 4.5 mM) were monitored as described above. Likewise, the inhibition was also monitored using a more concentrated buffer (300 mM Tris–HCl pH 7.0). The percentage inhibition was calculated using the following equation: [a], where α is $(1+[I])/K_I$, β is $[S]/K_m$, [I] is inhibitor concentration and [S] is substrate concentration.

Protein crystallization

*So*GDH was crystallized by the sitting-drop vapor diffusion method using the Index™ HT screening kit (Hampton Research, Laguna Niguel, CA, USA). A drop of 0.2 μl of freshly purified protein (10 mg/ml) was mixed with 0.2 μl of the reservoir solution and suspended over the latter at 291K. After 48 h, crystals were identified in the drop suspended over the reservoir solution consisting of 0.2 M ammonium acetate, 0.1 M Bis–Tris pH 5.5 and 25% (w/v) Poly (ethylene glycol) 3350. Subsequently, crystals of *So*GDH complexed with the cofactor NAD^+ were obtained using co-crystallization. In this case, *So*GDH (10 mg/ml) was mixed with 5 mM NAD^+ and 1 mM L-gal in the same reservoir solution and crystallized in a similar fashion to that of the ligand-free enzyme. In all cases, the crystals were harvested and cryo-cooled in liquid nitrogen for data collection.

Data collection and structure determination

The X-ray diffraction data were collected at 100K on the 'Sirius' synchrotron (Laboratório Nacional de Luz Síncrotron - Centro Nacional de Pesquisa em Energia e Materiais - LNLS-CNPEM, Campinas - Brazil) using beamline Macaná housing a PILATUS 2M detector for cofactor-free *So*GDH and on the 'Diamond Light Source' synchrotron (Harwell Science & Innovation Campus, Didcot, Oxfordshire, UK) using beamline I03 with an Eiger2 XE detector for *So*GDH–NAD^+. The data were indexed, integrated and scaled using Xia2 (Winter 2010). The coordinates for *Os*GDH were still unavailable at the time of structure solution and so attempts at phasing, using conventional molecular replacement, were made employing pdb structure 3V0T as the search model. These proved unsuccessful and so instead we adopted the 'MoRDa' pipeline (Vagin and Lebedev 2015). 'MoRDa' uses a database of preprocessed chains from the PDB, which have been trimmed to remove flexible regions and therefore optimize the chances of molecular replacement success. The molecular replacement itself is then performed with the detected homologs using 'MolRep' (Vagin and Teplyakov 1997). In the case of *So*GDH, several structures were used in the process including 1YNP (29.3% identity), 3UYI (30.6% identity), 6HG6 (26.1% identity), etc. The best model (based on *Z*-score and final values of R_{work} and R_{free}) was subjected to automated model building with *ARP/wARP* (Langer et al. 2008) to obtain a final model with the correct sequence. Subsequently, *So*GDH–NAD^+ was solved by molecular replacement with the previously refined *So*GDH structure as the search model using Phaser (McCoy 2007). Alternate rounds of refinement and model rebuilding conducted using Phenix (Adams et al. 2010) and Coot (Emsley and Cowtan 2004) yielded the final models. The residues involved in the NAD^+ binding site were determined with the Discovery Studio Visualizer V21.1.0 (BIOVIA, Dassault Systèmes) and figures were generated with PyMol v2.05 (Schrödinger, LLC). The data collection, refinement statistics and PDB codes are summarized in **Supplementary Table S3**.

Protein structure prediction

*Md*GDH structure prediction was performed with ColabFold, an online extension of the AlphaFold2 software (Jumper et al. 2021, Mirdita et al. 2021). An MMseqs2 (Uniref+environmental) msa_mode, automatic model_type, unpaired+paired pair_mode and 3 rum_recycles were selected as the input parameters. One hundred and fifty sequences were selected as templates for model building, obtaining five final models. The quality of the modeled tridimensional protein structures was evaluated by the local distance difference test parameters (Jumper et al. 2021, Mirdita et al. 2021) and stereochemistry by Procheck (Laskowski et al. 1993) and Verify3D (Eisenberg et al. 1997).

Phylogenetic analysis

The evolutionary history of 13 subfamilies of AKRs annotated from the University of Pennsylvania PKR database and 5 representatives of plant L-GDHs was inferred using the maximum likelihood method (Goldman 1990). A bootstrap consensus tree was inferred from 1,000 replicates (Felsenstein 1985), with subfamily branches collapsed where possible. Branches corresponding to partitions reproduced in less than 30% bootstrap replicates were also collapsed. Phylogenetic analyses were conducted in MEGA11 (Tamura et al. 2021).

Supplementary Data

Supplementary data are available at *PCP* online.

Data Availability

The data underlying this article are available in the GenBank (accession number OK632632.1) and Protein Data Bank (PDB codes 7SMI and 7SVQ).

Funding

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (88887.505769/2020-00 to J.A.V.); *Programa Nacional de Innovación Agraria* (188-2018-INIA-PNIA-PASANTIA to J.A.V.); *Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica* (069-2019-FONDECYT to J.A.V.); *Universidad Nacional de la Amazonia Peruana* (UNAP) (R.R. No. 1152-2020-UNAP to J.C.C.).

Acknowledgements

We gratefully acknowledge to Macromolecular Crystallography School 'From data collection to structure refinement and beyond'—Montevideo (2021) for help in the collection and processing of the crystallographic data for *So*GDH–NAD^+. We are also grateful to Susana Sculaccio and Andressa Alves Pinto for excellent technical support. We recognize the essential role played by DIAMOND UK and CNPEM Brazil by providing access to beamlines and to their technical staff during synchrotron data collection. We are grateful to Zui Fujimoto for depositing and releasing the coordinates for the rice GDH structures.

Disclosures

All the authors declare that they have no conflicts of interest in relation to this work.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66: 213–221.

Barski, O.A., Tipparaju, S.M. and Bhatnagar, A. (2008) The aldo-keto reductase superfamily and its role in drug metabolism and detoxification. *Drug Metab. Rev.* 40: 553–624.

Bennett, M.J., Albert, R.H., Jez, J.M., Ma, H., Penning, T.M. and Lewis, M. (1997) Steroid recognition and regulation of hormone action: crystal structure of testosterone and $NADP^+$ bound to 3α-hydroxysteroid/dihydrodiol dehydrogenase. *Structure* 5: 799–812.

Caruso, P., Russo, M.P., Caruso, M., Guardo, M.D., Russo, G., Fabroni, S., et al.(2021) Plants.

Castro, J.C., Maddox, J.D., Cobos, M., Requena, D., Zimic, M., Bombarely, A., et al. (2015) De novo assembly and functional annotation of *Myrciaria dubia* fruit transcriptome reveals multiple metabolic pathways for L-ascorbic acid biosynthesis. *BMC Genomics* 16: 997.

Castro, J.C., Maddox, J.D. and Imán, S.A. (2018) Camu camu—Myrciaria dubia (Kunth) McVaugh. In Exotic Fruits Reference Guide. Edited by Rodrigues, S., de Oliveira Silva, E. and Sousa de Brito, E. pp. 97–105. Academic Press United Kingdom.

Eisenberg, D., Lüthy, R. and Bowie, J.U. (1997) [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 277: 396–404.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D* 60: 2126–2132.

Englard, S. and Seifter, S. (1986) The biochemical functions of ascorbic acid. *Annu. Rev. Nutr.* 6: 365–406.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N Y)* 39: 783–791.

Fenech, M., Amaya, I., Valpuesta, V. and Botella, M.A. (2019) Vitamin C content in fruits: biosynthesis and regulation. *Front. Plant Sci.* 9: 2006.

Ganesh, K., Soumen, R., Ravichandran, Y. and Janarthanan, K. (2017) Dynamic approach to predict pH profiles of biologically relevant buffers. *Biochem. Biophys. Rep.* 9: 121–127.

Gatzek, S., Wheeler, G.L. and Smirnoff, N. (2002) Antisense suppression of L-galactose dehydrogenase in *Arabidopsis thaliana* provides evidence for its role in ascorbate synthesis and reveals light modulated L-galactose synthesis. *Plant J.* 30: 541–553.

Goldman, N. (1990) Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analyses. *Syst. Biol.* 39: 345–361.

Hyndman, D., Bauman, D.R., Heredia, V.V. and Penning, T.M. (2003) The aldo-keto reductase superfamily homepage. *Chem. Biol. Interact.* 143–144: 621–631.

Jez, J.M., Bennett, M.J., Schlegel, B.P., Lewis, M. and Penning, T.M. (1997) Comparative anatomy of the aldo–keto reductase superfamily. *Biochem. J.* 326: 625–636.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583–589.

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

Kozma, E., Brown, E., Ellis, E.M. and Lapthorn, A.J. (2002) The crystal structure of rat liver AKR7A1: a dimeric member of the aldo-keto reductase superfamily. *J. Biol. Chem.* 277: 16285–16293.

Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372: 774–797.

Laing, W.A., Frearson, N., Bulley, S. and MacRae, E. (2004) Kiwifruit L-galactose dehydrogenase: molecular, biochemical and physiological aspects of the enzyme. *Funct. Plant Biol.* 31: 1015–1025.

Langer, G., Cohen, S.X., Lamzin, V.S. and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.* 3: 1171–1179.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26: 283–291.

Liao, G., Chen, L., He, Y., Li, X., Lv, Z., Yi, S., et al. (2021) Three metabolic pathways are responsible for the accumulation and maintenance of high AsA content in kiwifruit (*Actinidia eriantha*). *BMC Genomics* 22: 13.

Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., et al. (2017) Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr Sect D.* 73: 148–157.

Locato, V., Cimini, S. and De Gara, L. (2013) Strategies to increase vitamin C in plants: from plant defense perspective to food biofortification. *Front. Plant Sci.* 4: 152.

Major, L.L., Wolucka, B.A. and Naismith, J.H. (2005) Structure and function of GDP-Mannose-3′,5′-epimerase: an enzyme which performs three chemical reactions at the same active site. *J. Am. Chem. Soc.* 127: 18309–18320.

McCoy, A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 63: 32–41.

Mieda, T., Yabuta, Y., Rapolu, M., Motoki, T., Takeda, T., Yoshimura, K., et al. (2004) Feedback inhibition of spinach l-galactose dehydrogenase by l-ascorbate. *Plant Cell Physiol.* 45: 1271–1279.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2021) ColabFold: making protein folding accessible to all. *Nat Methods*. 19: 679–682.

Momma, M. and Fujimoto, Z. (2013) Expression, crystallization and preliminary X-ray analysis of rice {\sc l}-galactose dehydrogenase. *Acta Crystallogr. Sect. F* 69: 809–811.

Nur, S.M., Rath, S., Ahmad, V., Ahmad, A., Ateeq, B. and Khan, M.I. (2021) Nutritive vitamins as epidrugs. *Crit. Rev. Food Sci. Nutr.* 61: 1–13.

Obmolova, G., Teplyakov, A., Khil, P.P., Howard, A.J., Camerini-Otero, R.D. and Gilliland, G.L. (2003) Crystal structure of the *Escherichia coli* Tas protein, an NADP(H)-dependent aldo-keto reductase. *Proteins Struct. Funct. Bioinforma.* 53: 323–325.

Penning, T.M. (2015) The aldo-keto reductases (AKRs): overview. *Chem. Biol. Interact.* 234: 236–246.

Raab, D. Graf, M. Notka, F. Schödl, T. and Wagner, D. (2010) The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Syst Synth Biol.* 4: 215–225.

Smirnoff, N. (2000) Ascorbate biosynthesis and function in photoprotection. *Philos. Trans. R Soc. London Ser. B Biol. Sci.* 355: 1455–1464.

Smirnoff, N. and Wheeler, G.L. (2000) Ascorbic acid in plants: biosynthesis and function. *Crit. Rev. Biochem. Mol. Biol.* 35: 291–314.

Soufari, H., Parrot, C., Kuhn, L., Waltz, F. and Hashem, Y. (2020) Specific features and assembly of the plant mitochondrial complex I revealed by cryo-EM. *Nat Commun.* 11: 5195.

Sujatha, M.S., Sasidhar, Y.U. and Balaji, P.V. (2004) Energetics of galactose–and glucose–aromatic amino acid interactions: Implications for binding in galactose-specific proteins. *Protein Sci.* 13: 2502–2514.

Tamura, K., Stecher, G. and Kumar, S. (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38: 3022–3027.

Vagin, A. and Lebedev, A. (2015) MoRDa, an automatic molecular replacement pipeline. *Acta Crystallogr. Sect. A* 71: s19.

Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* 30: 1022–1025.

Valpuesta, V. and Botella, M.A. (2004) Biosynthesis of L-ascorbic acid in plants: new pathways for an old antioxidant. *Trends Plant Sci.* 9: 573–577.

Wheeler, G.L., Jones, M.A. and Smirnoff, N. (1998) The biosynthetic pathway of vitamin C in higher plants. *Nature* 393: 365–369.

Winter, G. (2010) XIA2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* 43: 186–190.

Young, J.I., Züchner, S. and Wang, G. (2015) Regulation of the epigenome by vitamin C. *Annu. Rev. Nutr.* 35: 545–564.