



HOKKAIDO UNIVERSITY

| | |
|------------------|--|
| Title | Structural insights into unique substrate selectivity of <i>Thermoplasma acidophilum</i> D-aldohexose dehydrogenase |
| Author(s) | Yasutake, Yoshiaki; Nishiya, Yoshiaki; Tamura, Noriko et al. |
| Citation | Journal of Molecular Biology, 367(4), 1034-1046 https://doi.org/10.1016/j.jmb.2007.01.029 |
| Issue Date | 2007-04-06 |
| Doc URL | https://hdl.handle.net/2115/22096 |
| Type | journal article |
| File Information | JMB367-4.pdf |



Structural Insights into Unique Substrate Selectivity of *Thermoplasma acidophilum* D-Aldohexose Dehydrogenase

Yoshiaki Yasutake¹, Yoshiaki Nishiya², Noriko Tamura¹ and Tomohiro Tamura^{1,3*}

¹Research Institute of Genome-based Biofactory, National Institute of Advanced Industrial Science and Technology (AIST), 2-17-2-1 Tsukisamu-Higashi, Toyohira-ku, Sapporo 062-8517, Japan

²Tsuruga Institute of Biotechnology, Toyobo Co., Ltd., 10-24 Toyo-cho, Tsuruga 914-0047, Japan

³Laboratory of Molecular Environmental Microbiology, Graduate School of Agriculture, Hokkaido University, Kita-9, Nishi-9, Kita-ku, Sapporo 060-8589, Japan

***Corresponding author:** T. Tamura; Tel, 81-11-857-8938; Fax, 81-11-857-8980; E-mail, t-tamura@aist.go.jp

Running title: Crystal structure of D-aldohexose dehydrogenase

Abbreviations: AldT, *Thermoplasma acidophilum* D-aldohexose dehydrogenase; BmGlcDH, *Bacillus megaterium* D-glucose dehydrogenase; MDR, medium-chain dehydrogenase/reductase; NCS, noncrystallographic symmetry; SDR, short-chain dehydrogenase/reductase; PDB, Protein Data Bank; PEG, polyethylene glycol; r.m.s.d., root-mean-squared deviation.

Summary

The D-aldohexose dehydrogenase from the thermoacidophilic archaea *Thermoplasma acidophilum* (AldT) belongs to the short-chain dehydrogenase/reductase (SDR) superfamily and catalyzes the oxidation of several monosaccharides with a preference for NAD⁺ rather than NADP⁺ as a cofactor. It has been found that AldT is a unique enzyme that exhibits the highest dehydrogenase activity against D-mannose. Here, we describe the crystal structures of AldT in ligand-free form, in complex with NADH, and in complex with the substrate D-mannose, at 2.1 Å, 1.65 Å, and 1.6 Å resolution, respectively. The AldT subunit forms a typical SDR fold with an unexpectedly long C-terminal tail and assembles into an intertwined tetramer. The D-mannose complex structure reveals that Glu84 interacts with the axial C2 hydroxyl group of the bound D-mannose. Structural comparison with *Bacillus megaterium* glucose dehydrogenase (BmGlcDH) suggests that the conformation of the glutamate side chain is crucial for discrimination between D-mannose and its C2 epimer D-glucose, and the conformation of the glutamate side chain depends on the spatial arrangement of nearby hydrophobic residues that do not directly interact with the substrate. Elucidation of the D-mannose recognition mechanism of AldT further provides structural insights into the unique substrate selectivity of AldT. Finally, we show that the extended C-terminal tail completely shuts the substrate-binding pocket of the neighboring subunit both in the presence and absence of substrate. The elaborate intersubunit interactions between the C-terminal tail and the entrance of the substrate-binding pocket imply that the tail may play a pivotal role in the enzyme activity.

Key words: D-aldohexose dehydrogenase, D-glucose dehydrogenase, D-mannose, short-chain dehydrogenase/reductase (SDR), *Thermoplasma acidophilum*

Introduction

NAD(P)⁺-dependent glucose dehydrogenase (EC 1.1.1.47; GlcDH) is an enzyme distributed in a wide variety of prokaryotic organisms and catalyzes the conversion of β-D-glucose to D-glucono-1,5-lactone, employing NAD(P)⁺ as a cofactor. Based on sequence and structural similarities, two distinct GlcDH classes are defined. Archaeal GlcDH is one of the members of the medium-chain dehydrogenase/reductase (MDR) superfamily, with a subunit weight of approximately 40 kDa, and contains structural and catalytic zinc ions.^{1,2} The structures of MDR-type GlcDH are reported for the *Thermoplasma acidophilum* GlcDH in a ligand-free form,² the *Haloferax mediterranei* enzyme in a ligand-free form,³ and for the *Sulfolobus solfataricus* enzyme in complex with a substrate and the cofactor NADP⁺.⁴ These archaeal GlcDHs generally display dual substrate specificity for both D-glucose and D-galactose and are found to participate in the first step of the nonphosphorylative Entner-Doudoroff pathway, a promiscuous sugar metabolic pathway for pyruvate synthesis.⁴⁻⁶ Eubacteria, such as *Bacillus* species, also contain GlcDHs that belong to the short-chain dehydrogenase/reductase (SDR) superfamily,^{7,8} with a typical molecular weight of 25 to 30 kDa. As against the MDR-type GlcDHs, the *Bacillus* GlcDH is highly expressed in germinating spores and plays a crucial role in the production of large amounts of NADH that serves as an ATP source for spores.^{9,10} The MDR and SDR enzymes are structurally related and have a common dinucleotide-binding motif, i.e., the Rossmann fold,¹¹ however the active site structure and reaction mechanism are dissimilar, indicating a functional convergence.

Although most archaea possess only MDR-type GlcDHs, genome analysis of the thermoacidophilic archaea *T. acidophilum*¹² identified three SDR-type GlcDH

homologous genes (Ta0754, Ta0191, and Ta0747). We previously demonstrated that Ta0754 was able to oxidize several monosaccharides, including D-mannose, 2-deoxy-D-glucose, D-xylose, D-glucose, and D-fucose, with a preference for NAD⁺ rather than NADP⁺ as a cofactor.¹³ Surprisingly, Ta0754 acts the most effectively on D-mannose, and it should be noted that Ta0754 is thus far the only enzyme that exhibits reasonable dehydrogenase activity toward D-mannose, employing NAD⁺ as a cofactor. Its activities against a wide range of monosaccharides suggest that Ta0754 might be involved in the metabolism of these sugars to utilize them as carbon or energy sources. The substrate selectivity of Ta0754 is correlated with that of the *Gluconobacter cerinus* D-aldohehexose dehydrogenase (EC 1.1.1.119)^{14,15} and the *Pseudomonas* sp. D-aldohehexose dehydrogenase (EC 1.1.1.118)¹⁶ rather than various *Bacillus* GlcDH isozymes. Ta0754 was thus considered to be a novel archaeal D-aldohehexose dehydrogenase and was designated AldT.

In contrast to ubiquitous glucose oxidizing enzymes distributed both in prokaryotes and eukaryotes, such as flavoprotein (FAD-containing) glucose oxidase (EC 1.1.3.4),^{17,18} quinoprotein (pyrroloquinoline-quinone (PQQ)-containing) glucose dehydrogenase (EC 1.1.5.2),^{19,20} and GlcDHs, AldT is a notable enzyme that exhibits high dehydrogenase activity for D-mannose and relatively low activity for D-glucose. The unique substrate selectivity of AldT may be useful for the quantitative determination of D-mannose in clinical tests. Furthermore, it is of interest to investigate how AldT structurally discriminates between various monosaccharides, particularly D-mannose and its C2 epimer D-glucose—the most chemically stable and abundant D-aldohehexose in nature. To date, the structure of the *Bacillus megaterium* glucose dehydrogenase (BmGlcDH) has been reported only for the complex with the cofactor

NAD⁺,²¹ and no structural information on its monosaccharide-binding mechanism is yet available. Here, we describe for the first time the crystal structures of AldT in ligand-free, NADH-bound, and substrate D-mannose-bound forms at 2.1 Å, 1.65 Å, and 1.6 Å resolution, respectively. Structural comparison between AldT and BmGlcDH provides structural insights into the substrate recognition mechanism by these enzymes. The structures also revealed an unusually extended C-terminal tail that might be involved in the enzyme activity.

Results and Discussion

Subunit structure of AldT

The structure of AldT was solved by the molecular replacement method using BmGlcDH fragments as a search model and was refined with reasonable stereochemical quality to a resolution of 2.1 Å for the ligand-free form ($R_{\text{work}} = 18.9\%$, $R_{\text{free}} = 21.9\%$), 1.65 Å for the NADH complex ($R_{\text{work}} = 19.6\%$, $R_{\text{free}} = 21.8\%$), and 1.6 Å for the D-mannose complex ($R_{\text{work}} = 19.8\%$, $R_{\text{free}} = 22.1\%$). Refinement statistics and the model quality are summarized in Table 1. All the crystals used for structure analyses are isomorphous to each other, and the asymmetric unit contains two AldT subunits with the Matthews' coefficient V_M of $2.2 \text{ \AA}^3 \text{Da}^{-1}$.²² Convincing electron density was observed continuously for main-chain atoms, and the atomic model for the residues from 2 to 255 was established. The first methionine and the C-terminal Leu-Glu-His₆ sequence, for which there was no electron density due to conformational flexibility, were omitted from the model. The ligand-free structure can be superimposed on the cofactor and substrate complex structures with a root-mean-square deviation (r.m.s.d.) of less than 0.3 Å for the main-chain atoms. These low r.m.s.d. values indicate that the binding of

the substrate or cofactor does not give rise to any conformational changes in the enzyme.

The subunit of AldT forms a single globular structure with approximate dimensions of 35 Å × 40 Å × 50 Å and consists of 7 α -helices, 7 β -strands, and four 3_{10} -helices (Figures 1(a) and 2). The structure of AldT contains a slightly modified Rossmann fold motif,¹¹ with a twisted parallel seven-stranded β -sheet in the order β_3 - β_2 - β_1 - β_4 - β_5 - β_6 - β_7 , flanked on both sides by a total of five α -helices (α_1 - α_4 , α_7). The other two α -helices, i.e., α_5 and α_6 , lie between β_6 and α_7 by forming a helix-turn-helix motif and are located adjacent to the core framework of an α/β structure. The helix-turn-helix motif protrudes from the globular body of the subunit structure and creates the entrance of the substrate-binding pocket. One intramolecular disulfide linkage is found between Cys173 and Cys238 that belong to β_6 and β_7 , respectively. An additional intriguing feature of the subunit structure is the existence of an extended C-terminal tail consisting of nine residues, i.e., 247–255 (Figures 1 and 5(a)). Clear electron density reveals that the C-terminal tail is not flexible but is highly ordered along the neighboring subunit and is likely to be involved in tetramer assembly and substrate binding, as discussed later.

Tetrameric assembly

Intermolecular interactions in the crystal lattice clearly indicate that AldT assembles into a stable tetramer (Figure 1(c)), which is consistent with the previous result of gel filtration chromatography. The tetramer is described as a dimer of dimers with a 222 point-group symmetry mediated by three perpendicular 2-fold axes that are conventionally termed the P -, Q -, and R -axis.²³ The asymmetric unit contains two

subunits (molecules A and B) that are related by noncrystallographic 2-fold symmetry, and these two subunits create the most extensive intersubunit interface (*Q*-interface). Two α -helices ($\alpha 3$ and $\alpha 4$) predominantly participate in the *Q*-interface interactions and form a four-helix bundle structure. The *Q*-interface is formed by a number of hydrophobic interactions that are formed with hydrophobic and aromatic residues such as Ile89, Trp97, Ile101, Leu105, Phe106, and Tyr109. The tetramer can be generated from the dimer in the asymmetric unit with crystallographic 2-fold symmetry, and additional two intersubunit interfaces AB' (or BA') and AA' (or BB') consequently arise (prime signs represent identical subunits related by crystallographic 2-fold symmetry). The AB' (or BA') interface (*P*-interface) is created by hydrophobic residues, such as Phe225, Ile234 and Leu239, located around the $\alpha 7$ - $\beta 7$ region and by one ion pair between Glu230 and Arg26. In contrast to these intersubunit interfaces, the AA' (or BB') interface (*R*-interface) is rather small-scale but exhibits a unique characteristic, i.e., the *R*-interface is created by the interactions between the entrance of the substrate-binding pocket and the C-terminal tail. Seven hydrogen bonds and a number of van der Waals contacts are formed for the creation of the *R*-interface. These elaborate interactions suggest that the *R*-interface is not an artifact generated by crystallization, as discussed in the last section.

Structure comparison with BmGlcDH

The structural similarity search utilizing the DALI server²⁴ demonstrated that several enzymes belonging to the SDR family fold into a structure similar to that of AldT, and it is found that the most structurally homologous enzyme is BmGlcDH (PDB code, 1GCO; Z score, 34.8; r.m.s.d., 1.9 Å for 244 C α atoms). Figure 1(b) shows the

superimposition of the AldT and BmGlcDH structures. Although the overall fold of the AldT and BmGlcDH structures is primarily identical, there are also three apparent differences. The first difference is that one α -helix is missing in AldT; this is observed between $\beta 2$ and $\beta 3$ as a part of the $\beta\alpha\beta\alpha\beta$ motif that is observed not only in BmGlcDH but also in all Rossmann fold enzymes. In AldT, 12 residues are deleted in comparison with the corresponding region of BmGlcDH, resulting in the disruption of the α -helix formation (Figure 2). This structural part is in proximity to the adenine ribose moiety of the bound cofactor and comprises a key structural determinant for cofactor specificity as described later. The second and third differences are found in the helix-turn-helix ($\alpha 5$ - $\alpha 6$) region and in the C-terminal tail. AldT has nine additional residues in the $\alpha 5$ - $\alpha 6$ region and creates a deep groove comprising the substrate-binding pocket (Figures 1 and 2). It is of interest that the extended C-terminal tail donated from one of the adjacent subunits is elaborately fitted into this groove with forming the most part of the *R*-interface, and completely shuts the substrate-binding pocket. In contrast to AldT, the substrate-binding site of BmGlcDH is opened up to the solvent due to the presence of additional 3_{10} -helix at the C-terminal region (Figures 1(b) and 2).

Cofactor binding site

Thanks to the high-resolution structure analysis, we established the complete atomic model of the bound NADH (Figure 3(a)), and all the residues and solvent molecules that directly interact with NADH were identified (Table 2). The bound NADH was located at the C-terminal edge of the seven-stranded parallel β -sheet in an extended conformation. The distance between C6 of the adenine ring and C2 of the nicotinamide ring is 14.2 Å; this is very similar to that of BmGlcDH²⁰ and is close to

those of various other SDR enzymes. The adenine ring is positioned in an *anti* conformation, while the nicotinamide ring moiety is in a *syn* conformation with respect to each corresponding ribose. Both the ribose rings of the bound NADH adopt the C2'-endo puckering mode. These conformational properties of the bound NADH are also common among most SDR enzymes.

SDR enzymes share a common structural framework for dinucleotide binding, i.e., the so-called Rossmann fold, including the extended consensus sequence GXXXGXG^{8,25} Among the SDR enzymes, BmGlcDH has the highest structural similarity to AldT, as described above, and the conformation of the cofactor and the nearby residues of BmGlcDH are similar to that of AldT. However, a striking difference is also found in the loop structure near the adenine ribose moiety of the bound NADH. Although the structure of BmGlcDH is solved in a complex with NAD⁺, biochemical studies revealed that the preference of BmGlcDH for NADP⁺ exceeds that for NAD⁺ by 10-fold.²⁶ The preference for NADP⁺ is consistent with the presence of the basic residues arginine (Arg39) and lysine (Lys41) that lie at a reasonable position for binding to the negatively charged 2'-phosphate of NADP⁺ (Figures 2 and 3(b)). These residues are derived from the loop structure linked between β 2 and β 3. In contrast to BmGlcDH, the corresponding loop in AldT is located even more proximal to the adenine ribose moiety owing to the deletion of 12 residues as compared to BmGlcDH (Figure 3(b)). There are no basic residues, such as arginine and lysine, in this loop structure, and moreover, a steric clash is expected to occur between the 2'-phosphate and the loop region when NADP⁺ is bound to AldT. These structural properties clearly account for the difference in cofactor specificity between AldT and BmGlcDH. Several NAD⁺-dependent enzymes in the SDR family also exhibit a strategy similar to that

adopted by AldT for the preclusion of NADP binding, that is, they provide neither basic residues nor the space for the 2'-phosphate to bind to the enzyme.²⁷

Structural discrimination between D-mannose and D-glucose

We first attempted to solve the structures of the AldT:monosaccharide binary complex by utilizing crystals grown in the presence of an excess molar concentration of various monosaccharides. However, no electron density for sugar substrates was identified. Crystallization of AldT:monosaccharide:NAD⁺ ternary complex also resulted in failure. We successfully crystallized AldT in the presence of both NADH and monosaccharides and achieved the capture of D-mannose at a substrate-binding pocket. No electron density was observed for other monosaccharides such as D-glucose and D-xylose. These results indicate a high affinity of D-mannose for AldT and are consistent with the previous results of the enzymatic assay.¹³ An unexpected finding is that the D-mannose complex structure completely lacks the electron density for NADH. One can assume that NADH would gradually depart from the enzyme in the crystallization droplets, but the reason is not known.

Clear electron density for the pyranose ring of the bound D-mannose is found; it shows that the bound D-mannose is in the C1 (⁴C₁) chair conformation (Figure 4(a)). The occupancies of the bound D-mannose were estimated to be 0.89 for chain A and 0.91 for chain B through model refinement. There is no clear electron density map for the C1 and C3 hydroxyl moieties of the bound D-mannose, and thus these two oxygen atoms are excluded from the model and are not used for refinement. Weak electron density for these hydroxyls may arise from the presence of the highly ordered water molecules that lie close to the C1 and C3 atoms of D-mannose. Another possibility is

that a certain side reaction occurred in the crystallization droplet to remove these hydroxyls, and some chemically modified molecules might be present at the active site with partial occupancy. In contrast, the electron density map for the C2 hydroxyl (in axial configuration), C4 hydroxyl (in equatorial configuration), C6, and C6 hydroxyl is very clear, and therefore all the residues that interact with these three hydroxyls of D-mannose are identified (Table 2 and Figure 4(b)). The side chains of Glu84 and Gln134 hydrogen bond to the C2 hydroxyl and pyranose ring oxygen O5, respectively. The side-chain hydroxyl of Thr176 interacts with both the C4 and C6 hydroxyls of D-mannose. Several other residues, such as Tyr86 and Thr139, indirectly interact with D-mannose via the hydrogen bonding network of water molecules.

In contrast to the *Bacillus* GlcDHs that act the most effectively on D-glucose, AldT displays the highest dehydrogenase activity toward D-mannose. To elucidate the structural mechanism for discrimination between D-glucose and its C2 epimer D-mannose, the structure of AldT is compared to that of BmGlcDH at the substrate-binding pocket. Superimposition of AldT and BmGlcDH at the substrate-binding pocket (Figure 4(b)) shows that Thr176 and Gln134 are substituted with alanine (Ala190) and histidine (His147) in the BmGlcDH structure, respectively. The BLAST search showed that these alanine and histidine residues are strictly conserved in most BmGlcDH homologs in the SDR family, suggesting that these residues are essential for GlcDH activity. The His147 imidazole group of BmGlcDH occupies the same position as the side chain of Gln134 of AldT (Figure 4(b)), and both residues must play an identical role in sugar recognition, that is, binding to the pyranose ring oxygen O5. Thr176 of AldT interacts with both the C4 and C6 hydroxyls, and the substitution of threonine with alanine in BmGlcDH would disable the recognition of

these two hydroxyls. Alternatively, the side chain of the asparagine residue (Asn196 in BmGlcDH) that is also conserved in most GlcDHs could hydrogen bond to the C3 hydroxyl. In the AldT structure, there are no residues that can interact with the C3 hydroxyl of D-mannose. Thus, the recognition mechanisms of the C3, C4, and C6 hydroxyls would be considerably different between AldT and GlcDHs.

The most significant finding at the substrate-binding pocket is the conformation of the Glu84 side chain that interacts with the characteristic C2 hydroxyl of the bound D-mannose in an axial configuration. The glutamate residue is conserved in BmGlcDH as Glu96, but the chi-1 angle of the Glu96 side chain is rotated by approximately 120° as compared to that of Glu84 of AldT. Consequently, the carboxyl group of the Glu96 side chain lies within a reasonable distance from the putative C2 hydroxyl position of the D-glucose (Figure 4(b)). The chi-1 rotation of the glutamate side chain does not depend on whether the substrate sugar is bound or not. Trp152 and Phe155 are present in the vicinity of the Glu96 side-chain in BmGlcDH, and these aromatic residues deprive the Glu96 side chain of its mobility. In contrast, AldT does not have such aromatic residues, and alternatively, Leu181 most probably restricts the conformation of Glu84 from the opposite side. In fact, the current structure analyses reveal that the side-chain conformation of Glu84 is identical with a low *B*-factor both in the absence and presence of D-mannose. In short, both the C2 hydroxyl of D-mannose (in axial configuration) and D-glucose (in equatorial configuration) can be recognized by the conserved glutamate residue both in AldT and BmGlcDH, but the spatial arrangement of the nearby hydrophobic residues prevents the flexibility of the glutamate side chain, thus enhancing the affinity for either D-mannose or D-glucose.

Structural insights into substrate selectivity of AldT

AldT is able to oxidize several five- and six-carbon sugars such as D-mannose (relative activity, 100%), 2-deoxy-D-glucose (78.9%), D-xylose (49.9%), D-glucose (14.8%), and D-fucose (8.2%), while no activities were detected for D-galactose (C4 epimer of D-glucose), D-arabinose (C4 epimer of D-xylose), D-allose (C3 epimer of D-glucose), and D-altrose (C3 epimer of D-mannose), as previously described.¹³ As far as we know, there are no reports that GlcDHs as well as AldT recognize the furanose form or aldehyde (linear) form of sugars. In addition, the aldohexoses and aldopentoses described in this paper are more preferably present in a pyranose form than in a furanose form in the solution state.²⁸ Thus, the following discussion is limited to those sugars in a pyranose form with stable C1 chair conformation.

Based on the current D-mannose complex structure, the possible or hypothetical models for each sugar complex could be constructed by the superimposition of the pyranose ring moieties of each sugar and D-mannose. Figure 4(c) shows the schematic representations of the sugar-bound condition at the active site of AldT, explaining the substrate selectivity of AldT. The monosaccharide 2-deoxy-D-glucose lacks the C2 hydroxyl of D-mannose (or D-glucose), and the lower activity against 2-deoxy-D-glucose presumably originates from the impossibility of hydrogen-bond formation between the Glu84 side chain and the C2 hydroxyl. On the other hand, D-glucose, D-xylose, and D-fucose have the same C2 hydroxyl in equatorial configuration. The current models indicate that hydrogen bonds also cannot be formed between these equatorial hydroxyls and Glu84, and that the equatorial C2 hydroxyl would be located close to the nicotinamide ring plane at a distance of 3.4 Å. These observations suggest that the equatorial C2 hydroxyl would be incompatible with the

active-site conformation of AldT and may account for why AldT is more active on 2-deoxy-D-glucose than on these three sugars. AldT is far more active on D-xylose than on D-glucose, although we cannot provide a satisfactory explanation based on the current models. Because the only difference between D-xylose and D-glucose is the presence or absence of the hydroxymethyl group at the C5 position, it is suggested that the interaction between the C6 hydroxyl and Thr176 might result in the lowering of the activity if the sugars have the equatorial C2 hydroxyl. It should be noted that no activity was detected against the D-galactose and D-arabinose that are C4 epimers of D-glucose and D-xylose, respectively.¹³ The equatorial C4 hydroxyl of D-mannose interacts with the side chain of Thr176, while the axial C4 hydroxyls of these sugars are located more than 4.0 Å away from any atoms of the enzyme and probably do not form any hydrogen bonds. D-Arabinose does not possess the C6 hydroxyl, and the lack of two hydrogen bonds formed by Thr176 might seriously decrease its affinity for the enzyme, presumably resulting in loss of activity. Considering the low level of activity against D-glucose, the lack of only one hydrogen bond between Thr176 and the C4 hydroxyl could also lead to the enzymatic inertness to D-galactose. D-Fucose can be described as a “6-deoxy-D-galactose” or as “5-methyl-D-arabinose.” Intriguingly, AldT exhibits slight activity against D-fucose, even though it is active on neither D-arabinose nor D-galactose. In the current model, the C6 of the bound D-mannose lies at a distance of approximately 3.9 Å from the Val133 side chain, suggesting that the hydrophobic interaction might partly contribute to the increase in D-fucose affinity for the enzyme. A previous enzyme assay demonstrated that no activities were detected for D-allose (C3 epimer of D-glucose) and D-altrose (C3 epimer of D-mannose), which are not common in nature. It is known that they also form pyranose rings substantially in solution with

C1 chair conformation.²⁸ The current models clearly show that the axial C3 hydroxyls of these sugars would be located near the Val182 side chain at a distance of approximately 2.8 Å and thus a repulsive force must preclude the binding of these C3 epimers to the enzyme.

Catalytic mechanism of AldT

The catalytic mechanism of the dehydrogenase activity of the SDR family enzymes has been well examined.^{7,29,30} There is a conserved catalytic triad, namely, Ser-Tyr-Lys at the catalytic site, and the tyrosine residue functions as a catalytic base and abstracts the proton from the substrate. The serine plays a role in the stabilization of the bound substrate, and the lysine indirectly interacts with hydroxyl of the tyrosine residue via the 2'-hydroxyl of the nicotinamide ribose of the cofactor and lowers its pK_a .^{7,29,30} The catalytic triad is also structurally conserved in AldT, suggesting that AldT employs the same catalytic mechanism. The current model of the bound D-mannose lacks the C1 hydroxyl due to poor electron density, and we could not experimentally determine whether the bound D-mannose is α -form, β -form, or a mixture of α - and β -forms. However, considering the structural environment at the substrate-binding site, it is suggested that the α -D-mannose cannot bind to the enzyme because the C1 hydroxyl of the α -form sterically clashes with the nicotinamide ring moiety of the bound cofactor. The C1 hydroxyl of β -D-mannose, in contrast, is able to reasonably form hydrogen bonds with the side-chain hydroxyls of Tyr145 and Ser132 that correspond to two residues of the catalytic triad described above (Figure 4(b) and Table 2). As a result, a true substrate for AldT is probably not α - but β -D-mannose as in the case of *Bacillus* GlcDHs that act on only β -D-glucose.

Since no structural changes are observed between the structures of NADH and the D-mannose complex, we could construct the putative model of the AldT:D-mannose:NADH ternary complex by superimposing two structures and then adding the NADH molecule to the D-mannose complex structure (Figure 4(b)). The initiation of the catalytic reaction is a proton abstraction at the C1 hydroxyl of the bound D-mannose by Tyr145 in combination with Ser132 and Lys149. The hydride transfer subsequently occurs from the C1 of D-mannose to the C4 of nicotinamide ring of the cofactor NAD⁺ with direct trajectory overlap. In the putative model of the ternary complex, the distance between the C1 of D-mannose and the C4 of nicotinamide ring is approximately 3.0 Å and the angle of N-C4-C1 is 115°. These values are consistent with those obtained from the theoretical calculations of hydride transfer³¹ and from a number of structures of NAD(P)-dependent enzymes.

C-terminal tail shuts the substrate-binding pocket

In all the three structures described in this study, the highly ordered C-terminal tail (residues 247–255) is found to be located along the groove of the neighboring subunit related by crystallographic 2-fold symmetry (Figure 5). The groove comprises the substrate-binding pocket, and thus the active site is inaccessible to the solvent owing to the C-terminal tail. The tail is sandwiched by the loop between $\beta 5$ and $\alpha 4$ and the helix-turn-helix motif $\alpha 5$ - $\alpha 6$, and the elaborate intersubunit interactions are formed (Figure 5(c)). The main-chain nitrogen and oxygen of Ile138 hydrogen bond to the main-chain oxygen of Ile247' and nitrogen of Ala249', respectively. The main-chain oxygen of Lys140 also interacts with the side-chain hydroxyl of Thr253', and the side-chain carboxyl group of Glu189 and the amino group of Lys200 lie at 3.1 Å from

the side-chain hydroxyl of Ser252'. Additionally, the side-chain carboxyl of Glu255' forms a hydrogen bond with the side-chain amide group of Asn141 with a distance of 2.6 Å. Other interatomic contacts between the C-terminal tail and the groove are formed at a distance of around 3.4–4.0 Å without any interpositions of solvent atoms, indicating that these interactions are made with van der Waals forces.

The elaborate intersubunit interactions at the entrance of the substrate-binding pocket imply that the C-terminal tail might be involved in the enzymatic function of AldT. In fact, C-terminal deletion mutants show no activity against any monosaccharides (unpublished data). One possibility is that the C-terminal tail plays a role in stabilization of the active-site conformation during the catalytic reaction. Although the tail does not directly interact with the substrate or the substrate-binding residues, the current structures unambiguously show that it interacts with the hydrophobic residues Tyr86, Ile137, Trp204, and Leu245, proximal to the substrate-binding site at distances of van der Waals contacts, and prevents them from being exposed to the solvent (Figures 5(b) and 5(c)). It is thus suggested that these interactions may stabilize the active-site conformation and promote the catalytic reaction. It is also noteworthy that the C-terminal tail is highly ordered along the active-site groove even in the substrate-free enzyme. Closed structures both in the substrate-free and substrate-bound forms indicate that AldT does not employ the so-called “induced-fit” mechanism for the capture of the substrate, and it is likely that the closed structure entails the limit of substrate accessibility to the active site. The previous results of the enzyme assay actually demonstrated that the activity at 20 °C is less than 10% of the maximum activity detected at 70 °C, and the closure of the active site may partially account for this relatively low activity of AldT at room temperature.

One can easily assume that the substrate-binding site would gradually be opened up and exposed to the solvent as the thermal vibration increases, and the resulting high mobility of the C-terminal tail would facilitate substrate binding and product release. There is another possibility that the ordering of the tail is a prerequisite for this crystal form to grow and that the tail orders on substrate binding in solution. The current structure analyses, however, provide no conclusive evidence for the relationship between the C-terminal tail and the substrate accessibility or enzyme activity. Various C-terminal deletion variants and more directed mutagenesis will allow us in future studies to determine its actual role in enzymatic function.

Materials and Methods

Protein expression and purification

Recombinant AldT was overexpressed and purified as described earlier¹³ but with a slight modification. Briefly, the gene encoding AldT was cloned into a pET28a expression vector (Novagen) with a hexahistidine tag at the C-terminus, and the plasmid was transformed into *Escherichia coli* BL21(DE3)-RIL cells. The overexpressed enzyme was purified from the cell lysate in one step using nickel nitrilotriacetic acid (Ni-NTA) resin (QIAGEN), according to the manufacturer's instructions. The quality and quantity of the purified enzyme were checked by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and the Bradford protein assay (Bio-Rad). The activity of His-tag-fused AldT was confirmed based on the method previously described.¹³ The high purity enzyme was dialyzed against 25 mM Tris-HCl buffer (pH 7.5) containing 20% (v/v) glycerol and was frozen at 193 K until the crystallization experiment.

Crystallization

All crystallization experiments were carried out by the hanging-drop or sitting-drop vapor-diffusion method at 293 K. Each drop was formed by mixing equal volumes (1.0–2.5 μl) of a reservoir solution and sample solution and was equilibrated against the reservoir solution (up to 500 μl). The initial crystallization screenings and the optimization of the hit conditions for ligand-free AldT were previously described.³² To obtain better crystals, we applied the streak-seeding method.³³ The crystals used as seeds were obtained with the solution containing 0.1 M sodium acetate pH 5.0, 0.2 M ammonium sulfate, 16% (w/v) polyethylene glycol (PEG) 3350, and 15% (v/v) glycerol. The micro seed crystals were prepared by crushing imperfect crystals and were transferred to a new drop using cat body hair; the new drops were pre-equilibrated overnight against a reservoir solution consisting of 0.1 M sodium acetate pH 5.0–5.4, 0.2 M ammonium sulfate, 14–18% (w/v) PEG 3350, and 15–20% (v/v) glycerol. To prevent overcrowding of crystals, the drops were carefully mixed by pipetting immediately after the transfer of the seed crystals. The high-quality single crystals appeared in a few hours, and they reached a typical size of approximately $0.3 \times 0.3 \times 0.2 \text{ mm}^3$ in one week.

The crystals of AldT in complex with cofactor and those in complex with substrate were successfully obtained without the seeding technique. The sample was adjusted to a concentration of 8.0 mg/ml and was incubated overnight at 293 K with 4 mM β -NADH (SIGMA). Crystals of the AldT:NADH complex were grown with a reservoir solution containing 0.1 M sodium acetate, pH 4.2–4.4, 0.2 M ammonium sulfate, 14–18% (w/v) PEG 3350, and 20% (v/v) glycerol. The single crystals of NADH

complex appeared within one week, and grew to a typical size of $0.5 \times 0.5 \times 0.3 \text{ mm}^3$. Crystals of AldT:D-mannose complex were grown with reservoir solution containing 0.1 M sodium acetate pH 4.2–4.4, 0.2 M ammonium sulfate, 14–16% (w/v) PEG 3350, 20% (v/v) glycerol, and 100 mM D-mannose. Crystals of the D-mannose complex appeared within two months and grew to a maximum size of $0.5 \times 0.5 \times 0.4 \text{ mm}^3$.

X-ray diffraction studies

All X-ray diffraction data sets were collected under cryogenic conditions at 100 K using synchrotron radiation. Since all the crystals used for X-ray diffraction studies could be prepared in the mother liquor containing 15–20% (v/v) glycerol, they were directly flash-cooled under a nitrogen gas stream at 100 K without soaking in an artificial cryoprotectant. Almost complete X-ray diffraction data sets for the ligand-free form and the NADH complex were collected at beamline BL5A of Photon Factory (PF; Tsukuba, Japan), using a Quantum-Q315 charge-coupled-device (CCD) detector (ADSC) up to a resolution of 2.2 Å and 1.65 Å, respectively. The diffraction data for the D-mannose-bound form were collected at PF BL6A using a Quantum-4R CCD detector up to a resolution of 1.6 Å. All the measured diffraction spots were indexed, integrated, and scaled using the HKL2000 program package.³⁴ These crystals were found to be isomorphous to each other: all the crystals were of space group $P3_221$ with unit-cell dimensions of approximately $a = b = 83 \text{ Å}$ and $c = 138 \text{ Å}$. Data collection statistics are summarized in Table 1.

Phasing and model refinement

The structure of AldT in the ligand-free form was solved by the molecular

replacement method using the programs AmoRe³⁵ and MOLREP,³⁶ with one of the dimers of BmGlcDH in the asymmetric unit as a search model (PDB code, 1GCO). To minimize the bias caused by the search model, we modified the BmGlcDH model based on sequence alignment using CLUSTALW.³⁷ Low sequence similarity and sequence gap areas are eliminated from the model. The possible solutions with reasonable molecular packing in the crystal lattice were obtained using BmGlcDH fragments consisting of residues 1–34, 62–151, 157–195, and 204–251 as a search model. The initial phases were calculated from the model using the program SFALL³⁸ within the CCP4 suite,³⁹ and phase improvement techniques (by solvent flattening, histogram matching, and noncrystallographic symmetry (NCS) averaging) were applied using the program DM.⁴⁰ The resultant electron density map was of high quality and was clearly interpretable.

Model building of the ligand-free structure was achieved using the graphic program XtalView/Xfit⁴¹ by altering the BmGlcDH fragments. Model refinement was performed using the program CNS (ver. 1.1).⁴² A randomly chosen 5% of all the measured reflections were set aside for cross-validation analysis. NCS restraints with a weight of 300 kcal/Å² were introduced in the initial refinement, and the restraints were gradually loosened as refinement progressed. Model fitting was performed manually based on the SIGMAA-weighted $2mF_o - DF_c$ and $mF_o - DF_c$ electron density maps at the end of each round of positional and individual *B*-factor refinement. After iterative rounds of model refinement and fitting, solvent atoms were automatically located using the program CNS,⁴² and the irrelevant solvent atoms were deleted manually using XtalView/Xfit.⁴¹ Model refinement of the NADH and D-mannose complex structures was also carried out using the program CNS⁴² without NCS restraints. The free *R*-factor was calculated by employing the same reflections used for the free *R*-factor calculations

in the model refinement of the ligand-free form. Clear electron density indicated that the NADH or D-mannose was bound to the enzyme. After the corrections of the protein model were almost completed, NADH or β -D-mannose molecules were added to each model. Occupancies of β -D-mannose were roughly estimated based on the SIGMAA-weighted $mF_o - DF_c$ map, and were refined by the program CNS.⁴² For the ligand-free and D-mannose complex structures, some sulfate ions were also added to the model based on the strong peak of the SIGMAA-weighted $mF_o - DF_c$ map and its consistent chemical environment. The topology and parameter files of the NADH and β -D-mannose molecules were obtained from the Hetero-compound Information Center of Uppsala on the Uppsala website.⁴³ The programs PROCHECK⁴⁴ and WHATIF⁴⁵ were used to assess the quality of the final refined model. The refinement statistics are summarized in Table 1.

Model analyses

DALI server²⁴ was utilized for structural similarity search against all known structures deposited in PDB. Superimpositions of protein models were performed using program LSQKAB.⁴⁶ Secondary structure elements were determined by program WHATIF.⁴⁵ The program SURFACE⁴⁷ within CCP4 suite³⁹ was used for calculation of solvent-accessible surface area. All figures for molecular drawings are prepared using PyMol (DeLano Scientific; <http://pymol.sourceforge.net/>).

Protein Data Bank accession code

The atomic coordinates and structure factor amplitudes for ligand-free, NADH-bound, and D-mannose-bound forms of AldT have been deposited in the RCSB

Protein Data Bank (<http://www.rcsb.org/pdb>) with the accession codes 2DTD, 2DTE, and 2DTX, respectively.

Acknowledgements

The authors would like to thank the beamline scientists at beamlines BL5A and BL6A (PF, Tsukuba, Japan) for their kind help during X-ray diffraction data collection. We are also grateful to Dr. I. Tanaka at Hokkaido University for providing generous access to the X-ray diffraction equipment. This work was supported by a research grant from the National Project on Protein Structural and Functional analyses from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. Synchrotron radiation experiments were conducted under the approval of 2006G165 at PF.

References

1. Jörnvall, H., Persson, B. & Jeffery, J. (1987). Characteristics of alcohol/polyol dehydrogenases. The zinc-containing long-chain alcohol dehydrogenases. *Eur. J. Biochem.* **167**, 195-201.
2. John, J., Crennell, S. J., Hough, D. W., Danson, M. J. & Taylor, G. L. (1994). The crystal structure of glucose dehydrogenase from *Thermoplasma acidophilum*. *Structure*, **2**, 358-393.
3. Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M.-J., Ferrer, J., Pire, C., Esclapez, J. & Rice, D. W. (2006). Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*. *Proc. Natl. Acad. Sci. USA*, **103**, 4846-4851.
4. Milburn, C. C., Lamble, H. J., Theodossis, A., Bull, S. D., Hough, D. W., Danson, M. J. & Taylor, G. L. (2006). The structural basis of substrate promiscuity in glucose dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Biol. Chem.* **281**, 14796-14804.
5. De Rosa, M., Gambacorta, A., Nicolaus, B., Giardina, P., Poerio, E. & Buonocore, V. (1984). Glucose metabolism in the extreme thermoacidophilic archaeobacterium *Sulfolobus solfataricus*. *Biochem. J.* **224**, 407-414.
6. Lamble, H. J., Heyer, N. I., Bull, S. D., Hough, D. W. & Danson, M. J. (2003). Metabolic pathway promiscuity in the archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. *J. Biol. Chem.* **278**, 34066-34072.
7. Jörnvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J. & Ghosh, D. (1995). Short-chain dehydrogenases/reductases (SDR). *Biochemistry*, **34**, 6003-6013.
8. Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Person, B. & Jörnvall, H. (2003). Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem. Biol. Interact.* **143-144**,

- 247-253.
9. Otani, M., Ihara, N., Umezawa, C. & Sano, K. (1986). Predominance of gluconate formation from glucose during germination of *Bacillus megaterium* QM B1551 spores. *J. Bacteriol.* **167**, 148-152.
 10. Otani, M., Umezawa, C. & Sano, K. (1988). Gluconate metabolism in germinated spores of *Bacillus megaterium* QM B1551: primary roles of gluconokinase and the pentose cycle. *Microbiol. Immunol.* **32**, 877-885.
 11. Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). Chemical and biological evolution of nucleotide-binding protein. *Nature*, **250**, 194-199.
 12. Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. & Baumeister, W. (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature*, **407**, 508-513.
 13. Nishiya, Y., Tamura, N. & Tamura T. (2004). Analysis of bacterial glucose dehydrogenase homologs from Thermoacidophilic archaeon *Thermoplasma acidophilum*: finding and characterization of aldohexose dehydrogenase. *Biosci. Biotechnol. Biochem.* **68**, 2451-2456.
 14. Avigad, G., Alroy, Y. & England, S. (1968). Purification and properties of a nicotinamide adenine dinucleotide phosphate-linked aldohexose dehydrogenase from *Gluconobacter cerinus*. *J. Biol. Chem.* **243**, 1936-1941.
 15. Avigad, G. & England, S. (1975). Aldohexose dehydrogenase from *Gluconobacter cerinus*-1. *Methods Enzymol.* **42**, 142-147.
 16. Dahms, A. S. & Anderson, R. L. (1971). D-Fucose metabolism in a Pseudomonad I: oxidation of D-fucose to D-fucono- δ -lactone by a D-aldohexose dehydrogenase. *J. Biol. Chem.* **247**, 2222-2227.
 17. Hecht, H. J., Kalisz, H. M., Hendle, J., Schmid, R. D. & Schomburg, D. (1993). Crystal structure of glucose oxidase from *Aspergillus niger* refined at 2.3 Å resolution. *J. Mol. Biol.* **229**, 153-172.

18. Wohlfahrt, G., Witt, S., Hendle, J., Schomburg, D., Kalisz, H. M. & Hecht, H. -J. (1999). 1.8 and 1.9 Å resolution structures of *Penicillium amagasakiense* and *Aspergillus niger* glucose oxidases as a basis for modeling substrate complexes. *Acta Crystallog. sect. D*, **55**, 969-977.
19. Oubrie, A., Rozeboom, H. J. & Dijkstra, B. W. (1999). Active-site structure of the soluble quinoprotein glucose dehydrogenase complexed with methylhydrazine: a covalent cofactor-inhibitor complex. *Proc. Natl. Acad. Sci. USA*, **96**, 11787-11791.
20. Oubrie, A., Rozeboom, H. J., Kalk, K. H., Duine, J. A. & Dijkstra, B. W. (1999). The 1.7 Å crystal structure of the apo form of the soluble quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus* reveals a novel internal conserved sequence repeat. *J. Mol. Biol.* **289**, 319-333.
21. Yamamoto, K., Kurisu, G., Kusunoki, M., Tabata, S., Urabe, I. & Osaki, S. (2001). Crystal structure of glucose dehydrogenase from *Bacillus megaterium* IWG3 at 1.7 Å resolution. *J. Biochem. (Tokyo)*, **129**, 303-312.
22. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
23. Rossmann, M. G., Adams, M. J., Buehner, M., Ford, G. C., Hackert, M. L., Liljas, A., Rao, S. T., Banaszak, L. J., Hill, E., Tsernoglou, D. & Webb, L. (1973). Molecular symmetry axes and subunit interfaces in certain dehydrogenases. *J. Mol. Biol.* **76**, 533-537.
24. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
25. Bottoms, C. A., Smith, P. E. & Tannar, J. J. (2002). A structurally conserved water molecule in Rossmann dinucleotide-binding domains. *Protein Sci.* **11**, 2125-2137.
26. Nagao, T., Makino, Y., Yamamoto, K., Urabe, I. & Okada, H. (1989). Stability-increasing mutants of glucose dehydrogenase. *FEBS Lett.* **253**, 113-116.
27. Didierjean, C., Rahuel-Clermont, S., Vitoux, B., Dideberg, O., Branlant, G. & Aubry, A. (1997). A crystallographic comparison between mutated glyceraldehyde-3-phosphate dehydrogenases from *Bacillus stearothermophilus* complexed with either NAD⁺ or

- NADP⁺. *J. Mol. Biol.* **268**, 739-759.
28. Hirabayashi, J. (2004). On the origin of glycome and saccharide recognition. *Trends Glycosci. Glycotechnol.* **16**, 63-85.
29. Filling, C., Berndt, K. D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jörnvall, H. & Oppermann, U. (2002). Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J. Biol. Chem.* **277**, 25677-25684.
30. Ghosh, D., Sawicki, M., Pletnev, V., Erman, M., Ohno, S., Nakajin, S. & Duax, W. L. (2001). Porcine carbonyl reductase: structural basis for a functional monomer in short chain dehydrogenases/reductases. *J. Biol. Chem.* **276**, 18457-18463.
31. Wu, Y-D. & Houk, K. N. (1991). Theoretical evaluation of conformational preferences of NAD⁺ and NADH: an approach to understanding the stereospecificity of NAD⁺/NADH-dependent dehydrogenases. *J. Am. Chem. Soc.* **113**, 2353-2358.
32. Yasutake, Y., Nishiya, Y., Tamura, N. & Tamura, T. (2006). Crystallization and preliminary crystallographic analysis of NAD⁺-preferring aldohexose dehydrogenase from the thermoacidophilic archaeon *Thermoplasma acidophilum*. *Acta Crystallog. sect. F*, **62**, 586-589.
33. Stura, E. A. & Wilson, I. A. (1991). Applications of the streak seeding technique in protein crystallization. *J. Cryst. Growth*, **110**, 270-282.
34. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
35. Navaza, J. (1994). *AMoRe*: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157-163.
36. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022-1025.
37. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680.

38. Agarwal, R. C. (1978). A new least-squares refinement technique based on the fast Fourier transform algorithm. *Acta Crystallog. sect. A*, **34**, 791-809.
39. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.
40. Cowtan, K. (1994). DM: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESFEACBM Newsletter on Protein Crystallography*, **31**, 34-38.
41. McRee, D. E. (1999). XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156-165.
42. Brünger, A. T., Adams P. D., Clore, G. M., DeLano W. L., Gros P., Grosse-Kunstleve R. W., Jiang J. -S., Kuszewski J., Nilges M., Pannu N. S., Read R. J., Rice L. M., Simonson, T. & Warren G. L. (1998). Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta crystallog. sect. D*, **54**, 905-921.
43. Kleywegt, G. J. & Jones, T. A. (1998). Databases in protein crystallography. *Acta Crystallog. sect. D*, **54**, 1119-1131.
44. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
45. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52-56.
46. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallog. sect. A*, **32**, 922-923.
47. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

Figure Legends

Figure 1. Structure of AldT. (a) Ribbon representation of subunit structure. The α -helices are colored in green, 3_{10} -helices blue, β -strands orange, and coils gray. The bound NADH is shown as a red stick model, and the secondary structure elements are also labeled. (b) Stereoview superimposition of the AldT and BmGlcDH subunit structures. AldT is colored in green and BmGlcDH in gray. Striking differences at the main-chain level between these structures are circled with red broken lines. (c) Ribbon representation of the tetrameric assembly of AldT. Two subunits A and B in the asymmetric unit are colored in green and blue, and the other two subunits A' and B' related by crystallographic 2-fold symmetry are shown in gray. The bound NADH is colored in red. Three perpendicular 2-fold axes (P -, Q -, and R -axes) that generate 222 point-group symmetry are also indicated.

Figure 2. Structure-based sequence alignment between AldT and BmGlcDH. The D-mannose-binding residues, cofactor-binding residues, and cofactor specificity determinant region are highlighted in black, gray, and light gray, respectively. The extended consensus sequence GXXXGXG for dinucleotide binding around the $\alpha 1$ - $\beta 1$ region is boxed, and the Ser-Tyr-Lys catalytic triad in the SDR family is boxed in a bold line. Nine residues at the C-terminus forming an extended C-terminal tail structure are underlined. The secondary structure elements of both AldT (determined in this study) and BmGlcDH (given by 1GCO structure) are also shown. Identical, strongly similar, and weakly similar residues are denoted by asterisks, semicolons, and colons, respectively.

Figure 3. The bound cofactor NADH. (a) SIGMAA-weighted $mF_o - DF_c$ difference Fourier map contoured at the 2.0σ level, and the final refined model of the bound NADH. The atoms are shown in standard colors: carbon, yellow; oxygen, red; nitrogen, blue; phosphorus, orange. (b) Stereoview superimposition of the cofactor-binding site structures of AldT and BmGlcDH. AldT is colored in green and BmGlcDH in gray. Cofactor specificity determinant

residues in the vicinity of the adenine ribose moiety are represented as a stick model. The pink sphere represents the possible position of the 2'-phosphate group of NADP.

Figure 4. The D-mannose binding site. (a) SIGMAA-weighted $mF_o - DF_c$ difference Fourier map contoured at the 1.8σ level, and the model of β -D-mannose. Two oxygen atoms—O1 and O3—labeled in red are not included in the refined model due to the unclear electron density. (b) Stereoview superimposition of the substrate-binding site structures of AldT and BmGlcDH. AldT is colored in green and BmGlcDH in blue. The bound D-mannose (in standard colors) and the nicotinamide ribose moiety of NADH (in gray) are depicted as stick models. Hydrogen bonds formed between AldT and the bound D-mannose are shown as broken gray lines. The C2 hydroxyl group of D-glucose (in equatorial configuration) is also depicted as a thin transparent model, and the putative hydrogen bond between Glu96 of BmGlcDH and this equatorial C2 hydroxyl is shown as a broken red line. (c) Schematic representation of various sugar-binding conditions at the substrate-binding site of AldT. The nicotinamide ring moiety of NAD is shown, and the pathway of hydride transfer (from sugar C1 to the nicotinamide ring C4) is also indicated by red broken lines. Hydrogen bonds are shown as black broken lines, and the impossible interactions are marked by x in red. Val133 and Val182 are colored in blue and purple, respectively (see text).

Figure 5. Interactions between the C-terminal tail and the entrance of the substrate-binding pocket. (a) Overall view. Subunit A is represented as a surface model colored in brown, and the other subunits as ribbon models in green. The bound NADH (in red) and D-mannose (in standard colors) are depicted as stick models. (b) Cross section of subunit A (surface representation), the bound cofactor, D-mannose, and the neighboring subunit A' (stick representation) with the same color scheme as in (a). This picture shows that the shortest distance between substrate and C-terminal tail is roughly 4.0 Å and that no specific interatomic interactions are found between them. (c) Stereoview representation of the detailed

interactions between the C-terminal tail and the entrance of the substrate-binding pocket. The carbon atoms of the C-terminal tail are colored in green and those at the entrance of the substrate-binding pocket in ochre. Hydrogen bonds are depicted as broken gray lines. The hydrophobic residues near the substrate-binding pocket covered with the C-terminal tail are also labeled in blue.

Table 1. Data-collection and refinement statistics

| | Ligand-free form | NADH complex | D-Mannose complex |
|--|---------------------------|---------------------------|---------------------------|
| Data collection statistics | | | |
| Beamline | PF BL5A | PF BL5A | PF BL6A |
| Wavelength (Å) | 0.90000 | 0.90000 | 1.0000 |
| Resolution (Å) ^a | 50–2.10 | 40–1.65 | 50–1.60 |
| | (2.18–2.10) | (1.71–1.65) | (1.66–1.60) |
| Unit-cell parameters (Å) | a = b = 81.8 c = 138.5 | a = b = 81.5 c = 138.4 | a = b = 81.9 c = 138.3 |
| Space group | P3 ₂ 21 | P3 ₂ 21 | P3 ₂ 21 |
| No. of observed reflections | 342,233 | 565,979 | 597,391 |
| No. of unique reflections | 32,006 | 64,639 | 71,448 |
| $R_{\text{sym}}^{\text{a,b}}$ | 0.073 (0.283) | 0.049 (0.371) | 0.070 (0.275) |
| Completeness (%) ^a | 100.0 (99.9) | 99.9 (99.9) | 99.8 (98.5) |
| Multiplicity ^a | 10.7 (9.3) | 8.8 (8.8) | 8.4 (5.1) |
| Mean $I/\sigma(I)$ ^a | 24.2 (6.7) | 29.3 (6.4) | 34.6 (6.7) |
| Wilson plot B -factor (Å ²) | 32.8 | 19.9 | 18.0 |
| Refinement statistics | | | |
| Resolution range (Å) | 40–2.10 | 40–1.65 | 40–1.60 |
| No. of reflections (working set /test set) | 31,942 (30,322/1,620) | 64,586 (61,335/3,251) | 71,296 (67,692/3,604) |
| $R_{\text{work}}^{\text{c}}$ | 0.189 | 0.196 | 0.198 |
| $R_{\text{free}}^{\text{d}}$ | 0.219 | 0.218 | 0.221 |
| No. of protein atoms | 3,904 | 3,904 | 3,904 |
| No. of solvent atoms | 246 | 334 | 381 |
| No. of SO ₄ ²⁻ atoms | 25 | – | 10 |
| No. of NADH atoms | – | 88 | – |
| No. of D-mannose atoms | – | – | 20 |
| Mean B -factor values (Å ²) | 29.1 | 23.3 | 21.5 |
| Protein | 27.8 | 22.2 | 20.1 |
| Solvent | 36.3 | 34.0 | 33.4 |
| Sulfate ion | 65.1 | – | 29.7 |
| NADH | – | 23.2 | – |
| D-Mannose | – | – | 33.5 |
| Model quality | | | |
| r.m.s.d. bond length (Å) | 0.007 | 0.007 | 0.010 |
| r.m.s.d. bond angle (°) | 1.31 | 1.39 | 1.43 |
| Ramachandran plot | | | |
| Most favored regions (%) | 90.6 | 90.6 | 91.1 |
| Additional allowed regions (%) | 8.7 | 8.7 | 8.0 |
| Generously allowed regions (%) | 0.7 | 0.7 | 0.9 |
| Disallowed regions (%) | 0.0 | 0.0 | 0.0 |

^aFigures in parentheses refer to data in the highest resolution shell.

^b $R_{\text{sym}} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$, where $\langle I_h \rangle$ is the mean intensity of a set of equivalent reflections.

^c $R_{\text{work}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ for the 95% of the reflection data used in the refinement. F_{obs} and F_{calc} are observed and calculated structure factor amplitudes, respectively.

^d R_{free} is the equivalent of R_{work} , except that it was calculated for a randomly chosen 5% test set excluded from the refinement.

Table 2. Hydrogen-bond distances between AldT and NADH or D-mannose

| NADH/D-mannose atoms | AldT atoms | Distances (Å) | |
|----------------------------|------------------------------|------------------|------------------|
| | | Subunit A | Subunit B |
| NADH | | | |
| Adenine ribose moiety | | | |
| AN1 | Val54 N | 2.9 | 3.0 |
| AN6 | Asp53 OD1 | 3.0 | 3.1 |
| AN7 | Wat582 (Wat440) ^b | 2.8 | 2.8 |
| AO2* | Ser16 OG | 3.3 | 3.2 |
| AO3* | Ser16 OG | 2.7 | 2.7 |
| Pyrophosphate moiety | | | |
| AO1 | Wat490 (Wat595) ^b | 2.7 | 2.8 |
| | Wat615 | 2.8 | - |
| NO1 | Thr179 OG | 2.7 | 2.7 |
| NO2 | Ile19 N | 2.9 | 2.9 |
| Nicotinamide ribose moiety | | | |
| NO2* | Tyr145 OH | 2.9 | 2.9 |
| | Lys149 NZ | 3.0 | 2.9 |
| NO3* | Lys149 NZ | 2.9 | 2.9 |
| | Asn80 O | 2.7 | 2.7 |
| NO7 | Ala175 O | 3.3 | 3.2 |
| | Ile177 N | 2.9 | 2.8 |
| NN7 | Thr179 OG | 3.0 | 3.1 |
| D-Mannose | | | |
| O2 | Glu84 OE2 | 3.1 | 3.0 |
| O4 | Thr176 OG1 | 2.5 | 2.9 |
| O5 | Gln134 NE2 | 2.8 | 2.8 |
| O6 | Thr176 OG1 | 2.9 | 2.8 |
| O1 (β-anomer) | Tyr145 OH | 2.2 ^a | 2.0 ^a |
| O1 (β-anomer) | Ser132 OG | 2.5 ^a | 2.6 ^a |

Atom names (AN1, AO2*, OD1 etc.) are those used in the PDB file.

^aFinal refined model of D-mannose does not contain O1 atom. The distances are estimated based on the theoretical position of O1 of β-D-mannose.

^bParentheses refer to water molecules at subunit B.

Figure 1

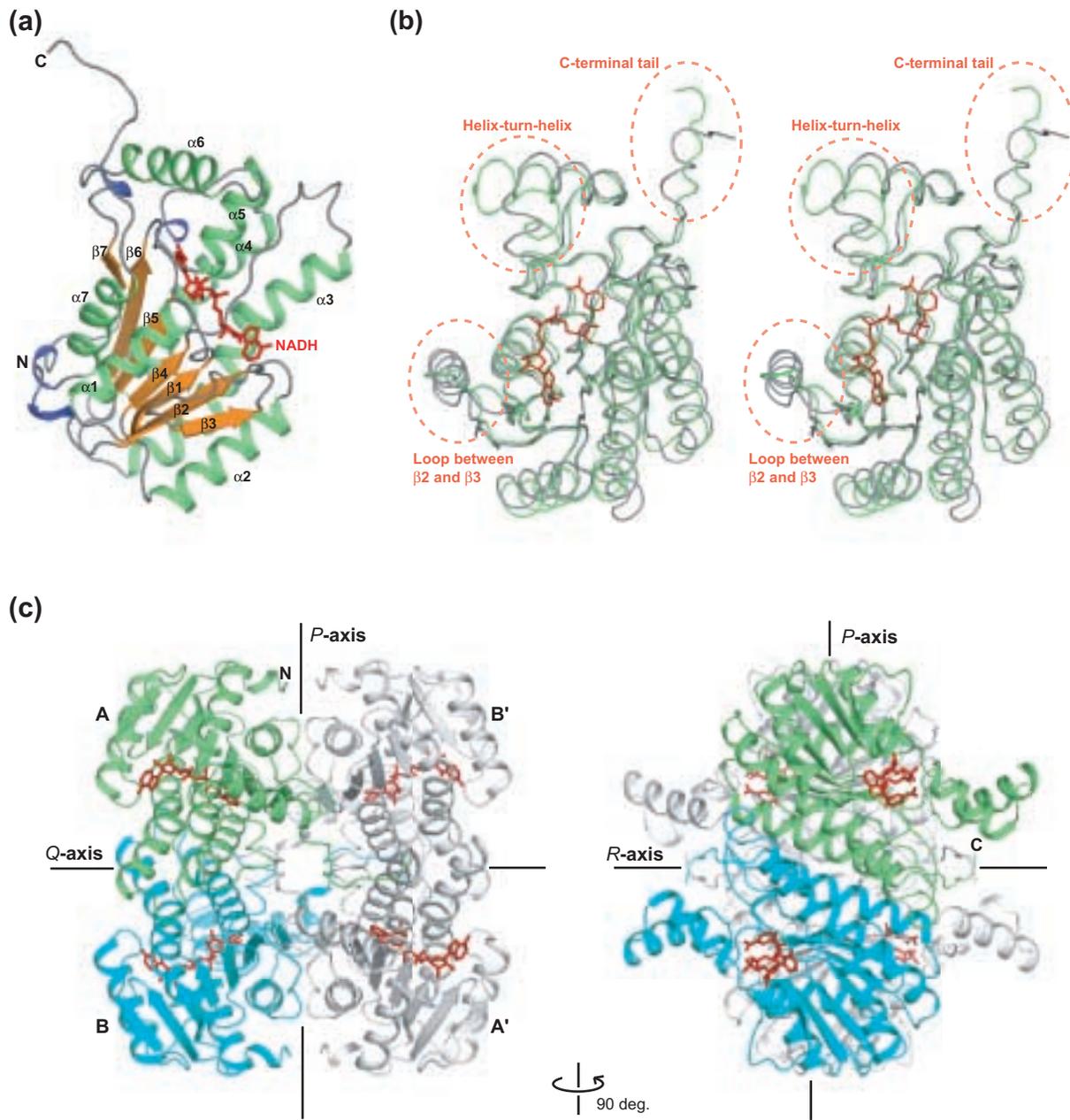


Figure3

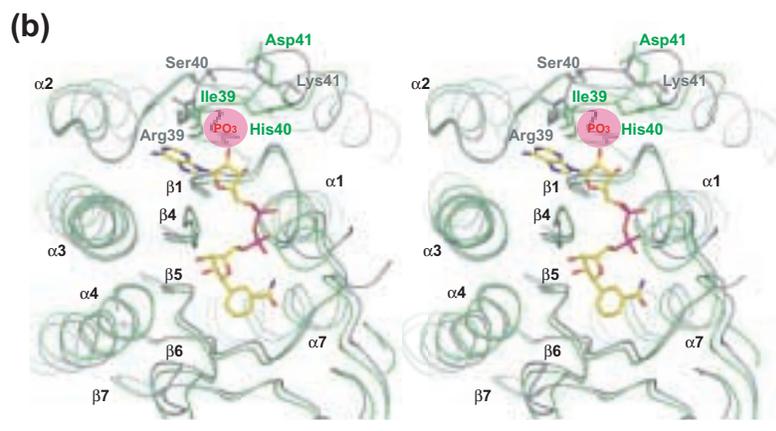
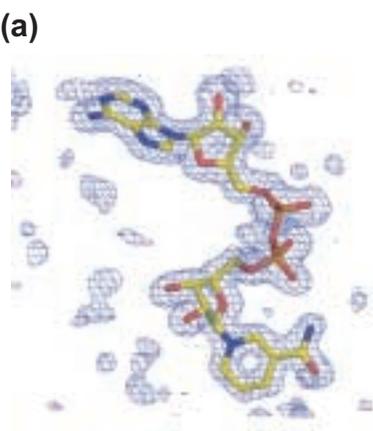
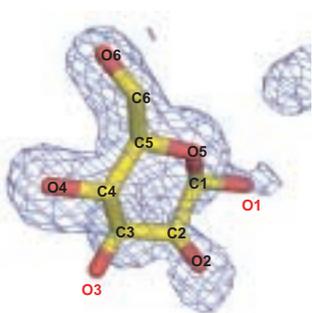
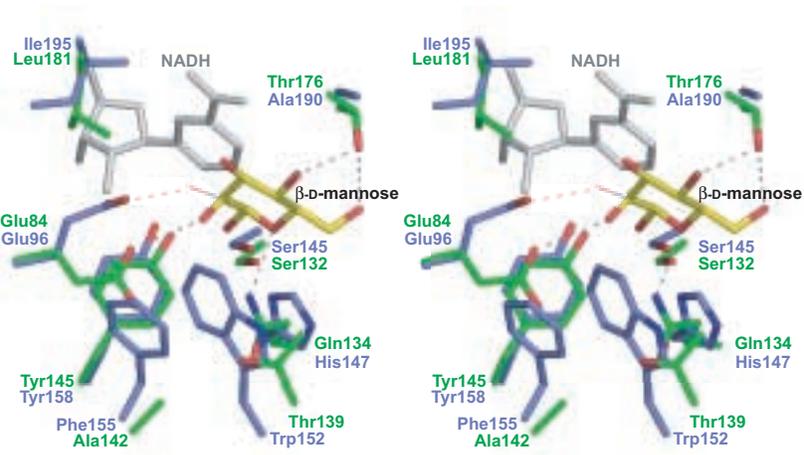


Figure4

(a)



(b)



(c)

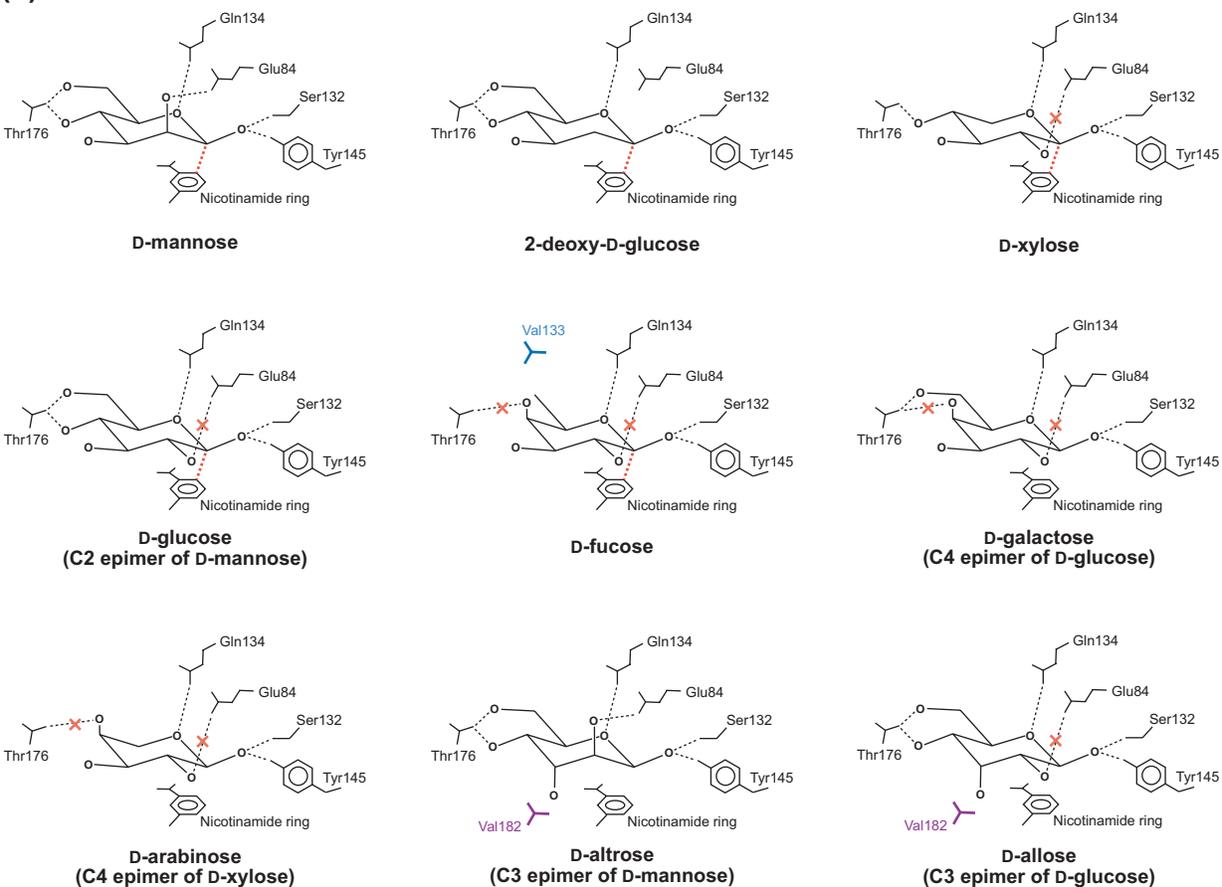


Figure 5

