

# SCIENTIFIC REPORTS



OPEN

## The crystal structure of D-xylonate dehydratase reveals functional features of enzymes from the Ilv/ED dehydratase family

Mohammad Mubinur Rahman<sup>1</sup>, Martina Andberg<sup>2</sup>, Anu Koivula<sup>2</sup>, Juha Rouvinen<sup>1</sup> & Nina Hakulinen<sup>1</sup> 

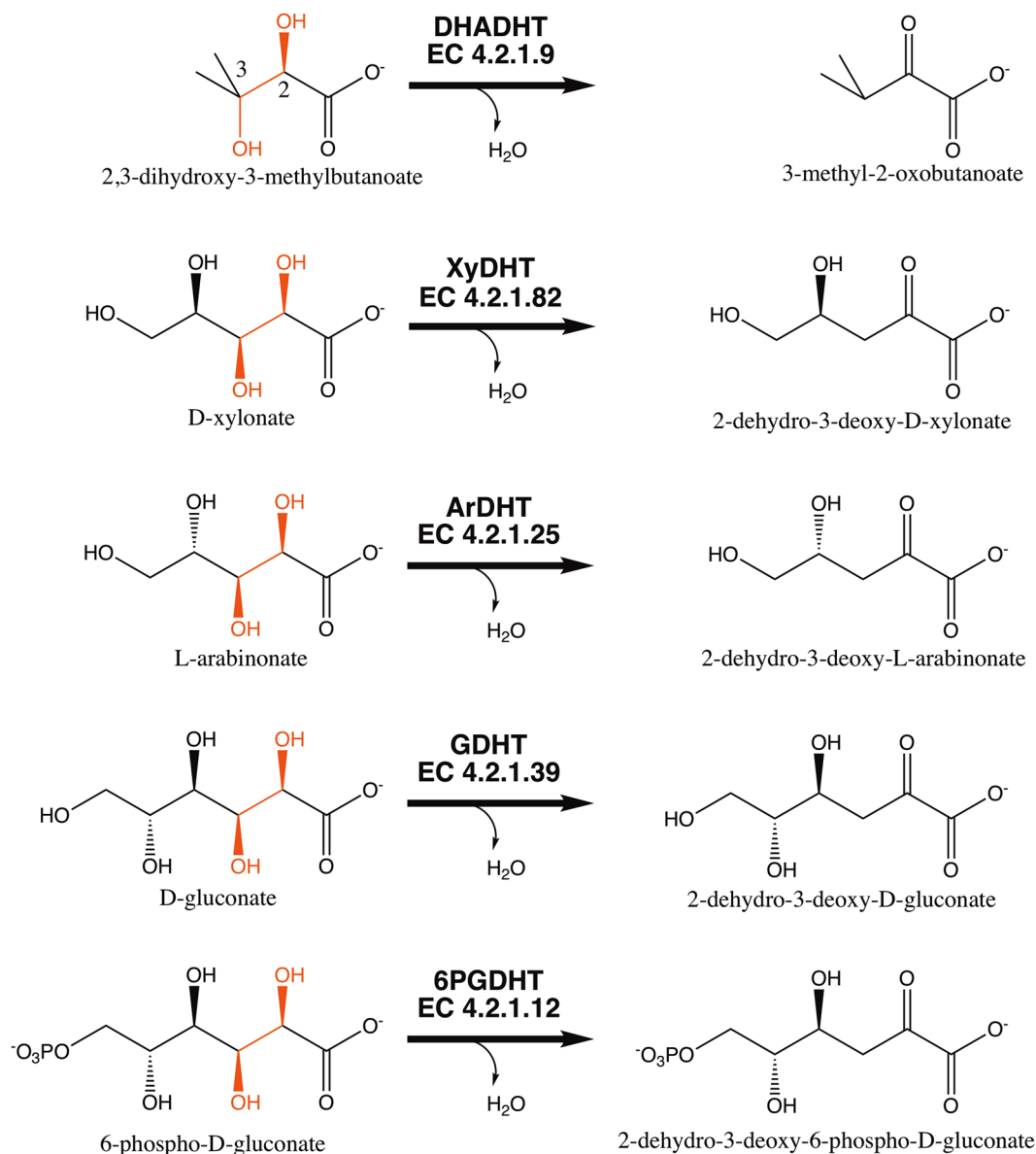
The Ilv/ED dehydratase protein family includes dihydroxy acid-, gluconate-, 6-phosphogluconate- and pentonate dehydratases. The members of this family are involved in various biosynthetic and carbohydrate metabolic pathways. Here, we describe the first crystal structure of D-xylonate dehydratase from *Caulobacter crescentus* (CcXyDHT) at 2.7 Å resolution and compare it with other available enzyme structures from the IlvD/EDD protein family. The quaternary structure of CcXyDHT is a tetramer, and each monomer is composed of two domains in which the N-terminal domain forms a binding site for a [2Fe-2S] cluster and a Mg<sup>2+</sup> ion. The active site is located at the monomer-monomer interface and contains residues from both the N-terminal recognition helix and the C-terminus of the dimeric counterpart. The active site also contains a conserved Ser490, which probably acts as a base in catalysis. Importantly, the cysteines that participate in the binding and formation of the [2Fe-2S] cluster are not all conserved within the Ilv/ED dehydratase family, which suggests that some members of the IlvD/EDD family may bind different types of [Fe-S] clusters.

There is increasing interest in developing economical and sustainable bioprocesses to convert biomass into valuable organic compounds. Metabolic engineering and synthetic biology have enabled the development of optimized microbial strains for the production of target compounds by taking advantage of known pathways and enzymes. To fully exploit the possibilities offered by nature, a more detailed molecular-level understanding of key metabolic enzymes is required. One important group of these enzymes belongs to the IlvD/EDD protein family, which is involved in the pathways used to produce biofuels such as ethanol, butanol, isobutanol, and 3-methyl-1-butanol as well as platform chemicals such as isobutyric acid, 1,2,4-butanetriol and D-pantothenic acid (vitamin B5)<sup>1–6</sup>.

Enzymes belonging to the IlvD/EDD protein family are widely distributed among bacteria, archaea, fungi, and algae and in higher plant species<sup>7–12</sup>. The family includes dihydroxy acid dehydratases (IlvD, DHADHT, DHAD; EC 4.2.1.9), which are involved in short-chain amino acid biosynthesis and catalyse the dehydration of 2,3-dihydroxy-3-methylbutanoate<sup>13</sup>, and aldonic acid dehydratases, which are involved in the metabolism of hexose and pentose sugars via the Entner-Doudoroff (ED) pathway or a modified ED pathway<sup>14–18</sup>. Pentonate dehydratases from the IlvD/EDD family participate in the non-phosphorylative oxidation pathways of pentose sugars in which D-xylonate dehydratase (XyDHT) and L-arabinonate dehydratase (ArDHT) catalyse the dehydration of D-xylonate and L-arabinonate, respectively<sup>17–21</sup>. D-gluconate dehydratases (GDHT, GnaD) and 6-phosphogluconate dehydratases (6PGDHT) are involved in the metabolism of glucose via the ED pathway. These two enzymes catalyse the dehydration of gluconate and 6-phosphogluconate, respectively<sup>16,22</sup>. These enzymes all function similarly, by catalysing an elimination reaction in which a hydroxyl group at the C3 position and a proton at the C2 position are cleaved, resulting in the loss of a water molecule (Fig. 1).

Enzymes from the IlvD/EDD family are suggested to contain a [Fe-S] cluster as a co-factor in the active site<sup>19</sup>. Some enzymes from this family have been characterized by using electron paramagnetic resonance (EPR) and UV-Vis spectroscopic methods to explore the structural features of the [Fe-S] cluster and its role in enzyme activity<sup>19,20,23–25</sup>. Based on these studies, two different catalytically active forms of [Fe-S] clusters have been found. A

<sup>1</sup>Department of Chemistry, University of Eastern Finland, PO Box 111, FIN-80101, Joensuu, Finland. <sup>2</sup>VTT Technical Research Centre of Finland Ltd, PO Box 1000, FIN-02044 VTT, Espoo, Finland. Correspondence and requests for materials should be addressed to N.H. (email: [nina.hakulinen@uef.fi](mailto:nina.hakulinen@uef.fi))



**Figure 1.** The reactions catalysed by IlvD/EDD enzymes. The identical reactive regions in different substrates are highlighted in red.

catalytically active [4Fe-4S] cluster that is unstable in aerobic environments has been reported in 6PGDHT from *Zymomonas mobilis*<sup>26</sup> and in DHADHT from *Escherichia coli*<sup>23</sup>. In contrast, a functional and stable [2Fe-2S] cluster in DHADHTs from several organisms, including bacteria, archaea, fungi and plants, has been observed<sup>24,25</sup>.

A detailed structural and functional characterization of these enzymes is important for understanding the catalytic mechanism and the substrate specificity, which could help improve the synthetic metabolic pathways in biorefinery industries. However, to date, only limited three-dimensional structural information is available for the IlvD/EDD enzymes. The coordinates of *Shewanella oneidensis* 6PGDHT in the apo form have been submitted to the Protein Data Bank (PDB code: 2GP4) but are not yet published. Moreover, this 6PGDHT structure is incomplete and partially incorrect since the original 2GP4 model contains a mixed  $\beta$ -sheet at N-terminal domain, but a corrected model is composed of parallel  $\beta$ -sheet. The model also lacks the cofactor and several surrounding loops in the active site. Recently, we solved the first representative of the IlvD/EDD family in its holo form, an L-arabinonate dehydratase from *Rhizobium leguminosarum* bv *trifolii* (RIArDHT, PDB code: 5J84). This structure revealed the presence of a [2Fe-2S] cluster and a  $\text{Mg}^{2+}$  ion in the active site<sup>27</sup>.

In this paper, we describe the first crystal structure of a D-xylonate dehydratase at 2.7 Å resolution and compare it with the two other known crystal structures (PDB codes: 2GP4 and 5J84) belonging to the IlvD/EDD family. D-xylonate dehydratase from *Caulobacter crescentus* (CcXyDHT) shares 42% sequence identity with RIArDHT, and the crystal structure of CcXyDHT shows that this enzyme has a [2Fe-2S] and a  $\text{Mg}^{2+}$  ion in its active site, similar to RIArDHT. The three different enzyme structures available from the IlvD/EDD enzyme family now allow the analysis of the structure and functional features within this protein family.

## Results and Discussion

**The re-refined structure of So6PGDHT.** The crystal structure of 6-phosphogluconate dehydratase from *Shewanella oneidensis* (So6PGDHT) at 2.5 Å resolution ( $R_{\text{work}} = 0.23$  and  $R_{\text{free}} = 0.29$ ) is available in the Protein Data Bank (PDB code: 2GP4), but the structure lacks a [Fe-S] cluster and a divalent metal ion. When inspecting the electron density map of So6PGDHT, we found several obscured regions in the crystal structure. Therefore, the 2GP4 model was re-refined, which resulted in improved R-values ( $R_{\text{work}} = 0.17$  and  $R_{\text{free}} = 0.23$ ). However, the improved model of 6PGDHT still had missing regions, Val35-Leu63, Lys182-Ile185 and Thr215-Thr226, and it was lacking density for the [Fe-S] cluster or  $\text{Mg}^{2+}$  ion in the active site. The re-refinement resulted in new locations for the active site residues Cys112 and Asp113. In the re-refined model, the putative iron-sulphur cluster binding residues Cys112 and Cys154 are sufficiently close to ligate a [Fe-S] cluster. The location of the third putative cluster-binding residue Cys220 is undefined because, it is located in the missing loop, the structure of which is unknown due to the absence of electron density.

**Overall structure of CcXyDHT.** The crystal structure of holo-CcXyDHT was determined at 2.7 Å resolution by molecular replacement, using L-arabinonate dehydratase from *Rhizobium leguminosarum* bv *trifolii* (5J84) as a template. The final structure had good R-factor and R-free values of 17.8% and 22.3%, respectively. The structure contained a tetramer in the asymmetric unit in the space group C2. Unambiguous electron density was seen for all of the amino acid residues of CcXyDHT, except for the disordered region at the N-terminus, which includes Asp1-Ser2 and the N-terminal Strep-tag II. The refinement resulted in a final model with good refinement statistics (Table 1).

The enzyme is an  $\alpha/\beta$  protein that consists of two domains: a N-terminal domain (residues Asn3-Leu358) and a C-terminal domain (residues Leu383-His591) (Fig. 2a,b). The two domains are connected by a long loop from Gln359 to Phe382. The N-terminal domain is composed of a  $\beta$ -sheet with four parallel  $\beta$ -strands surrounded by four  $\alpha$ -helices. In addition, the N-terminal domain contains an extension at the N-terminus (residues Arg10-Ser43) with three  $\alpha$ -helices and an insertion with  $\beta$ -hairpin (residues Gly157-Val164) and a helix-loop-helix (residues Thr168-Ser194). The C-terminal domain is composed of a  $\beta$ -sheet consisting of six parallel and two anti-parallel  $\beta$ -strands that are arranged like a  $\beta$ -barrel. Secondary structure analysis showed that the complete polypeptide chain consists of 27  $\alpha$ -helices and 16  $\beta$ -strands. Each monomer contains one [2Fe-2S] cluster and a  $\text{Mg}^{2+}$  ion. The substrate binding site is located in the cavity between the domains, but the [Fe-S] cluster and the  $\text{Mg}^{2+}$  binding site are located entirely in the N-terminal domain (Fig. 2c).

**Quaternary structure.** The quaternary structure of CcXyDHT is a homotetramer (Fig. 2d), similar to the homotetrameric structure of RlArDHT. D-gluconate dehydratase from *Achromobacter xylosoxidans* has also been reported to form a tetramer in solution<sup>22</sup>. A packing analysis of the crystal structure using the PISA server<sup>28</sup> suggested that the quaternary structure contains a stable dimer with an extensive monomer-monomer interface (4640 Å<sup>2</sup>). This interface is predominantly formed by residues from the larger-sized N-terminal domain. An  $\alpha$ -helix (residues 99–115) from this domain forms the central part of the interface. The tight dimer then packs against another similar dimer in such a way that one monomer has an interface with both polypeptide chains from the second dimer (887 and 607 Å<sup>2</sup>). The total monomer-monomer interface area in the tetramer is 12268 Å<sup>2</sup>.

In contrast, So6PGDHT forms a dimer in the crystal structure that is similar to the tight dimer of CcXyDHT (Fig. 2e,f). However, unlike CcXyDHT, So6PGDHT does not form a tetrameric structure. Many IlvD/EDD enzymes such as L-arabinonate dehydratase from *Azospirillum brasiliense*<sup>20</sup>, dihydroxy acid dehydratase from *Spinacia oleracea*<sup>24</sup>, and 6-phosphogluconate dehydratase from *Zymomonas mobilis* have been reported to form dimers in solution<sup>29</sup>.

The formation of the extensive dimer interface in CcXyDHT (and in RlArDHT) is essential for the formation of the active site. Although the residues that participate in the binding of  $\text{Mg}^{2+}$  and the [2Fe-2S] cofactor are located in the N-terminal domain of the first monomer, the polypeptide chain from the second monomer covers the active site by extending its N-terminal helix (residues 21–33) into the first monomer. Because the amino acid residues of this helix, which point towards the active site, are different in CcXyDHT and RlArDHT, this helix may be called a substrate recognition helix. In addition, the C-terminus of the dimeric counterpart protrudes into the active site of the first monomer by placing His591, which forms a salt bridge with Glu463, in the active site (Fig. 2e). The crystal structure of So6PGDHT shows similar dimer formation to that observed in CcXyDHT but does not show any evidence that the polypeptide chain from the dimeric counterpart participates in the formation of the active site. In contrast, the So6PGDHT structure represents the apo form, and some of the loop regions are not visible in the electron density map due to disorder in the absence of cofactor (Fig. 2f).

**Structural comparison of three IlvD/EDD enzymes.** The sequence of CcXyDHT is 42% identical to RlArDHT and 27% identical to So6PGDHT. A sequence alignment of these three enzymes and of representative members of the IlvD/EDD family is shown in Fig. 3. The higher sequence identity between CcXyDHT and RlArDHT, compared to So6PGDHT, is also reflected in a low, 0.70 Å RMS deviation between the C $\alpha$  atoms of monomer proteins. The overall structures are very similar except for a small insertion of a  $\beta$ -hairpin at Leu409-Val417 in CcXyDHT. The [2Fe-2S] cluster and  $\text{Mg}^{2+}$  ion coordinating residues are well conserved between these enzymes. The improved model of So6PGDHT can be superimposed with CcXyDHT with a RMSD of 1.95 Å. The N-terminal core domain consisting of a four-stranded parallel  $\beta$ -sheet surrounded by  $\alpha$ -helices and the C-terminal core containing a mixed eight-stranded  $\beta$ -sheet are similar. However, the large RMSD value between So6PGDHT and CcXyDHT is due to the following major differences between the structures: 1) the N-terminal regions, 2) the helix-loop-helix motifs and 3) the  $\beta$ -hairpin structures (Fig. 4).

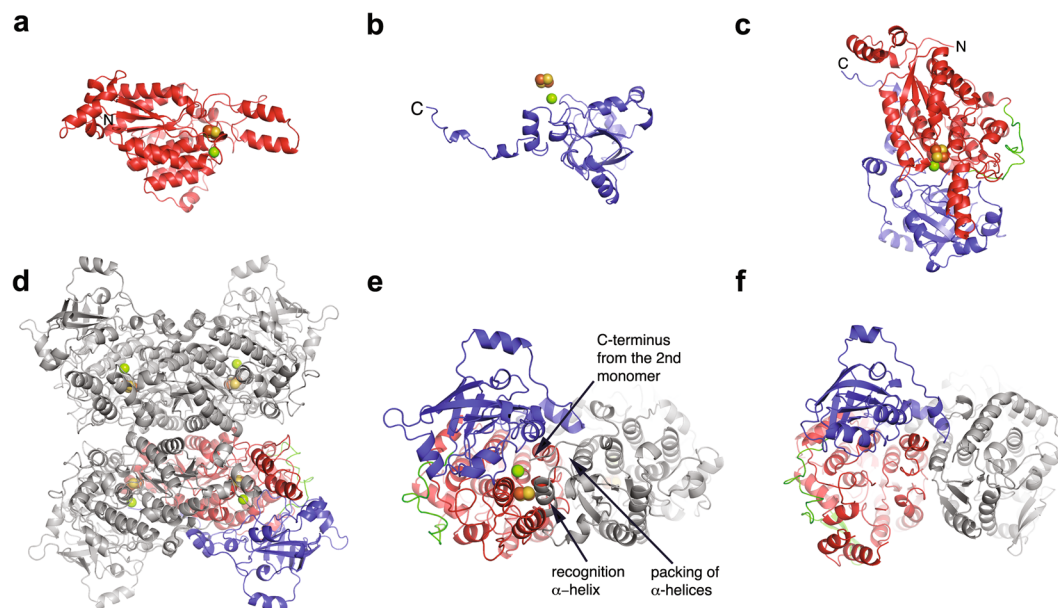
The N-termini of CcXyDHT and So6PGDHT are very different. In CcXyDHT, this region (residues 1–45) forms a bundle of three short helices, but in So6PGDHT, this region is longer (residues 1–65) and contains two

Data collection	CcXyDHT
Beam line	I04, DLS, England
Wavelength (Å)	0.979497
Resolution (Å)	59.12 – 2.66 (2.76 – 2.66)
Space group	C 1 2 1
Cell dimension	
a, b, c (Å)	270.41, 236.14, 65.17
$\alpha, \beta, \gamma$ (°)	90, 97.38, 90
Total reflections	393801 (60419)
Unique reflections	115354 (11560)
Multiplicity	3.4 (3.4)
Completeness (%)	99.54 (99.80)
Mean $I/\sigma$ (I)	17.2 (2.1)
Wilson B-factor (Å <sup>2</sup> )	59.91
R-meas (%)	7.0 (82.1)
$CC_{1/2}$	99.9 (95.8)
<b>Refinement statistics</b>	
Reflections used in refinement	110218 (10961)
Reflections used for R-free	5509 (546)
R-work	0.1769 (0.3295)
R-free	0.2174 (0.3814)
Number of non-hydrogen atoms	18336
Protein	17992
Solvent	324
Fe/S (cluster)	8/8
Mg	4
Protein residues	2354
<b>RMS deviations</b>	
Bond length (Å)	0.007
Bond angles (°)	0.95
<b>Ramachandran analysis (%)</b>	
Favoured	95.6
Allowed	3.9
Outliers	0.5
Rotamer outliers	0.0
Clashscore	8.33
<b>Average B-factor (Å<sup>2</sup>)</b>	
Protein	68
Solvent	58
Fe/S (cluster)	79/71
Mg	83
<b>PDB ID</b>	<b>5OYN</b>

**Table 1.** Data collection and refinement statistics. The structure was determined from a single crystal. The value in parentheses are for the highest-resolution shell.

long  $\alpha$ -helices, which have a very different orientation and pack against the core protein in an extended fashion. A part of the N-terminal region (35–63) is undefined in the structure of *So6PGDHT*. Perhaps the most remarkable structural difference between the proteins is the location of the Pro186-Ser214 region that contains a helix-loop-helix motif structure at the N-terminal domain of *So6PGDHT*. The helix-loop-helix in *So6PGDHT* has been flipped and moves by approximately 30 Å from the position observed in *CcXyDHT* (and in *RIArDHT*), where it partially covers the active site. Interestingly, in an open form of an *RIArDHT* mutant co-crystallized with L-arabinonate, this helix-loop-helix structure is slightly displaced by 7 Å, which suggests that it has a functional role in the catalytic cycle<sup>27</sup>.

In *So6PGDHT*, the helix-loop-helix region is twisted away from the active site. Unfortunately, the preceding loop (residues 182–185) and the following loop (residues 215–226) regions are disordered in the *So6PGDHT* structure, so how accessible the active site of *So6PGDHT* is compared to those of *CcXyDHT* and *RIArDHT* remains an open question. The third difference between *So6PGDHT* and *CcXyDHT* is an additional  $\beta$ -hairpin structure (residues 383–396) in the connecting loop of *So6PGDHT*. This packs against the abovementioned helix-loop-helix structure (Fig. 4a–c).



**Figure 2.** Structure of CcXyDHT and comparison of the dimers; the N- and C-terminal domains are shown in red and in blue, respectively. A long loop connecting the domains is shown in green. The [2Fe-2S] cluster is shown with yellow and orange spheres, and  $Mg^{2+}$  is shown as a green sphere. (a) N-terminal domain, (b) C-terminal domain, (c) monomer, (d) tetramer, (e) the dimer of CcXyDHT in which the monomeric unit is shown in colours and the dimeric counterpart is in grey, and (f) the dimer of So6PGDHT. The key features on the monomer-monomer interface are marked with arrows.

**Sequence motifs to identify the IlvD/EDD family.** We have analysed the multiple amino acid sequence alignment of some representative members of the IlvD/EDD family. The sequence alignment is presented in Fig. 3, including the sequences of the three enzymes (CcXyDHT, RlArDHT, So6PGDHT) for which three-dimensional structure information is available. The description of the IlvD/EDD protein family in the PROSITE database<sup>30</sup> reports that this family has two common sequence motifs. The first sequence motif consists of eleven residues with a peptide pattern Cys-Asp-Lys-x(2)-Pro-[Gly/Ala]-x(3)-[Gly/Ala], where Cys-Asp-Lys and Pro are conserved residues. The second sequence motif consists of twelve residues with a peptide pattern [Ser/Gly/Ala/Leu/Cys]-[Leu/Ile/Met/Phe]-[Leu/Ile/Val/Met/Phe]-Thr-Asp-[Gly/Ala]-Arg-[Leu/Ile/Val/Met/Phe/Tyr]-Ser-[Gly/Ala]-[Gly/Ala/Val]-[Ser/Thr], where Thr-Asp, Arg and Ser (without brackets) are conserved residues.

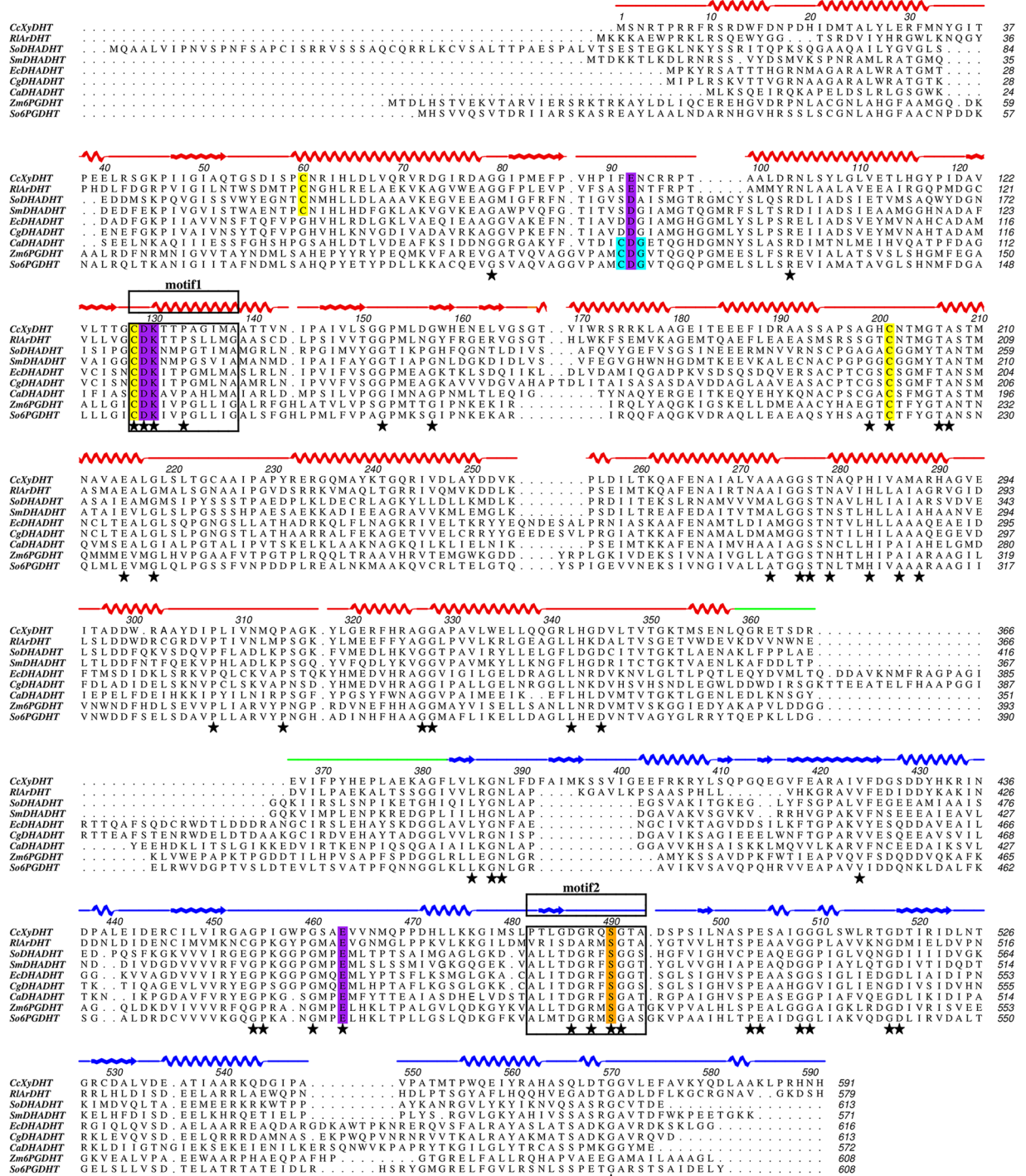
These two motifs can be found in the CcXyDHT structure (Figs 3, 4d). The first motif is Cys-Asp-Lys-Thr-Thr-Pro-Ala-Gly-Ile-Met-Ala at the N-terminal domain (residues 128–138). These eleven residues are located in the helix  $\alpha_6$  and the loop connecting the strand  $\beta_3$  and helix  $\alpha_6$ . The conserved Cys128 participates in the binding of the [2Fe-2S] cluster, Asp129 and Lys130 (in carbamylated form) in the binding of  $Mg^{2+}$ .

The second motif is Pro-Thr-Leu-Gly-Asp-Gly-Arg-Gln-Ser-Gly-Thr-Ala (residues 482–493) in CcXyDHT. These twelve residues are located at the C-terminal domain consisting of strand  $\beta_{13}$  and the loop follows. Interestingly, the first two and last two residues of this motif in CcXyDHT do not match the predicted sequence motif. The proposed conserved threonine is not conserved in CcXyDHT and is replaced by Gly485. The conserved Asp486 forms hydrogen bonds with the main chain nitrogens of Gly452 and Ala453 in the loop structure. The conserved Arg488 is located rather close (4.3 Å) to the  $Mg^{2+}$  but probably does not directly participate in the binding of magnesium; consequently, its role may be electrostatic. Ser490, which is proposed to act as a base in enzyme catalysis, as supported by experimental results from the corresponding Ser480Ala mutant of RlArDHT, is fully conserved among the IlvD/EDD protein family<sup>27</sup>.

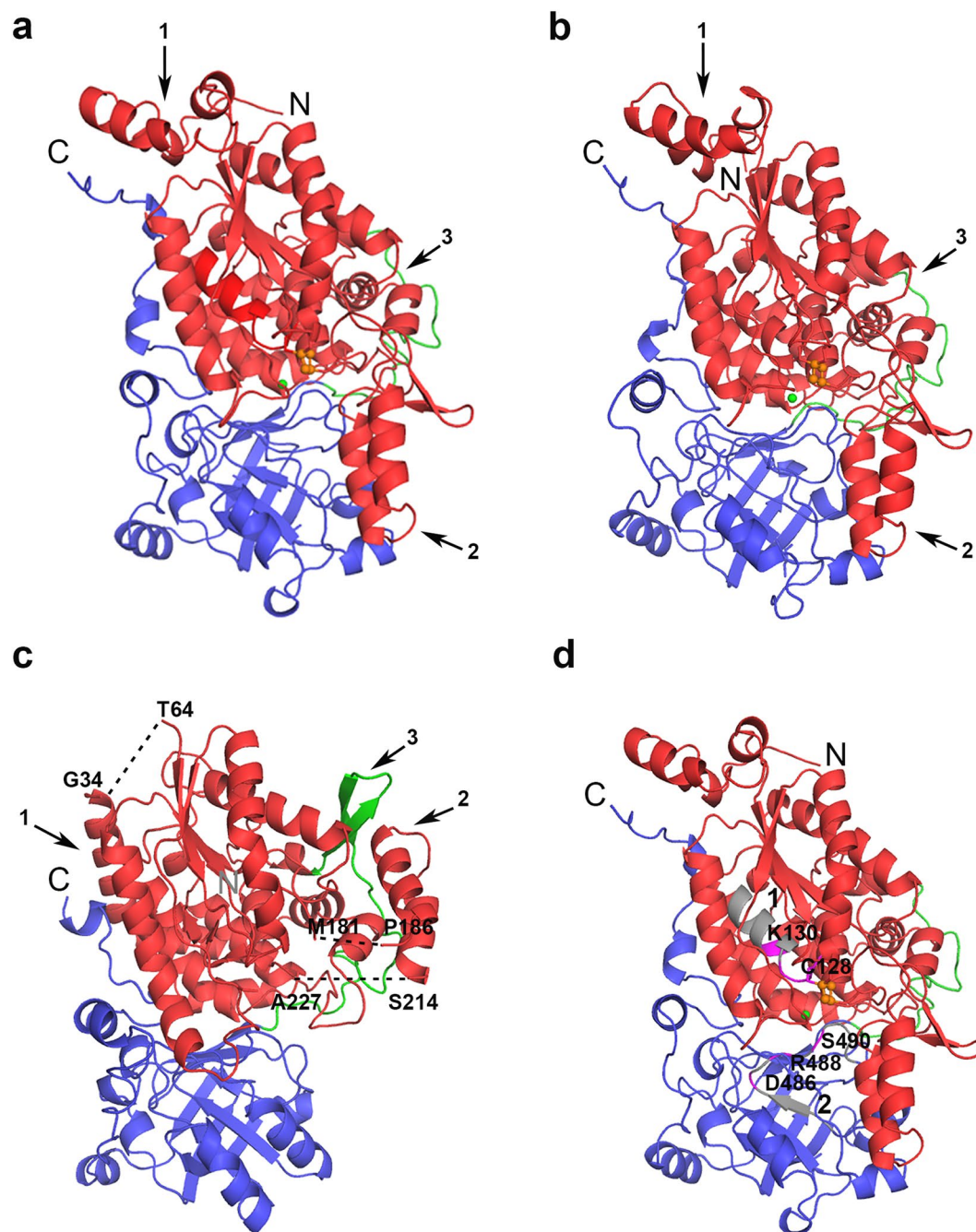
**The active site of CcXyDHT.** The subsequent analysis of the active site is based solely on the crystal structures of the holo forms of CcXyDHT and RlArDHT and not on the structure of So6PGDHT because the latter structure represents an incomplete structure in the apo form. The electron density map of the CcXyDHT crystal structure shows that the binding of the  $Mg^{2+}$  and the [2Fe-2S] cluster to the active site is similar to that observed in RlArDHT<sup>27</sup>.

The hexacoordination of the  $Mg^{2+}$  ion resembles a square bipyramid in which Glu92, Asp129, Glu463 and a water (or hydroxide ion) are approximately in the plane and another water molecule (or a hydroxide ion) and carbamylated Lys130 are in the axial positions. One water molecule in the square plane is hydrogen-bonded (2.7 Å) to Thr206. Due to the obtained resolution, we cannot fully exclude the possibility that Thr206 may also directly interact with the  $Mg^{2+}$  ion. However, the observed distance from Thr206 to the Mg is approximately 3.8 Å. Identical residues can be found in the structure of RlArDHT and all four  $Mg^{2+}$  binding residues (and Thr206) are fully conserved within the protein family (Fig. 3), suggesting that the binding of magnesium is a universal feature of the IlvD/EDD enzyme family. The crystal structure of CcXyDHT unambiguously shows that





**Figure 3.** Multiple sequence alignment of IlvD/EDD enzymes. *CcXyDHT* is a D-xylonate dehydratase from *Caulobacter crescentus* (GeneBank: ANS60449), *RIArDHT* is an L-arabinonate dehydratase from *Rhizobium leguminosarum* bv. *trifolii* (GeneBank: ANS60454), *SmDHADHT* is a dihydroxy acid dehydratase from *Streptococcus mutans* (GeneBank: KZM62800), *SoDHADHT* is a dihydroxy acid dehydratase from *Spinacia oleracea* (GeneBank: KNA20834), *EcDHADHT* is a dihydroxy acid dehydratase from *Escherichia coli* (GeneBank: CDZ22543), *CgDHADHT* is a dihydroxy acid dehydratase from *Corynebacterium glutamicum* (GeneBank: BAV23085), *CaDHADHT* is a dihydroxy acid dehydratase from *Clostridium acetobutylicum* (GeneBank: WP\_010966867), and *Zm6PGDHT* and *So6PGDHT* are 6-phosphogluconate dehydratases from *Zymomonas mobilis* (GeneBank: WP\_013934127) and *Shewanella oneidensis* (GeneBank: WP\_011072452). The N- and C-terminal domain are shown in red and blue, connecting loop in green. The [2Fe-2S] cluster coordinating residues are shown in yellow, Mg<sup>2+</sup> ion coordinating residues are in magenta and serine that play role as Lewis base is shown in orange, respectively. Cys-Asp-Gly motif is shown in cyan. Black stars represent the fully conserved residues.

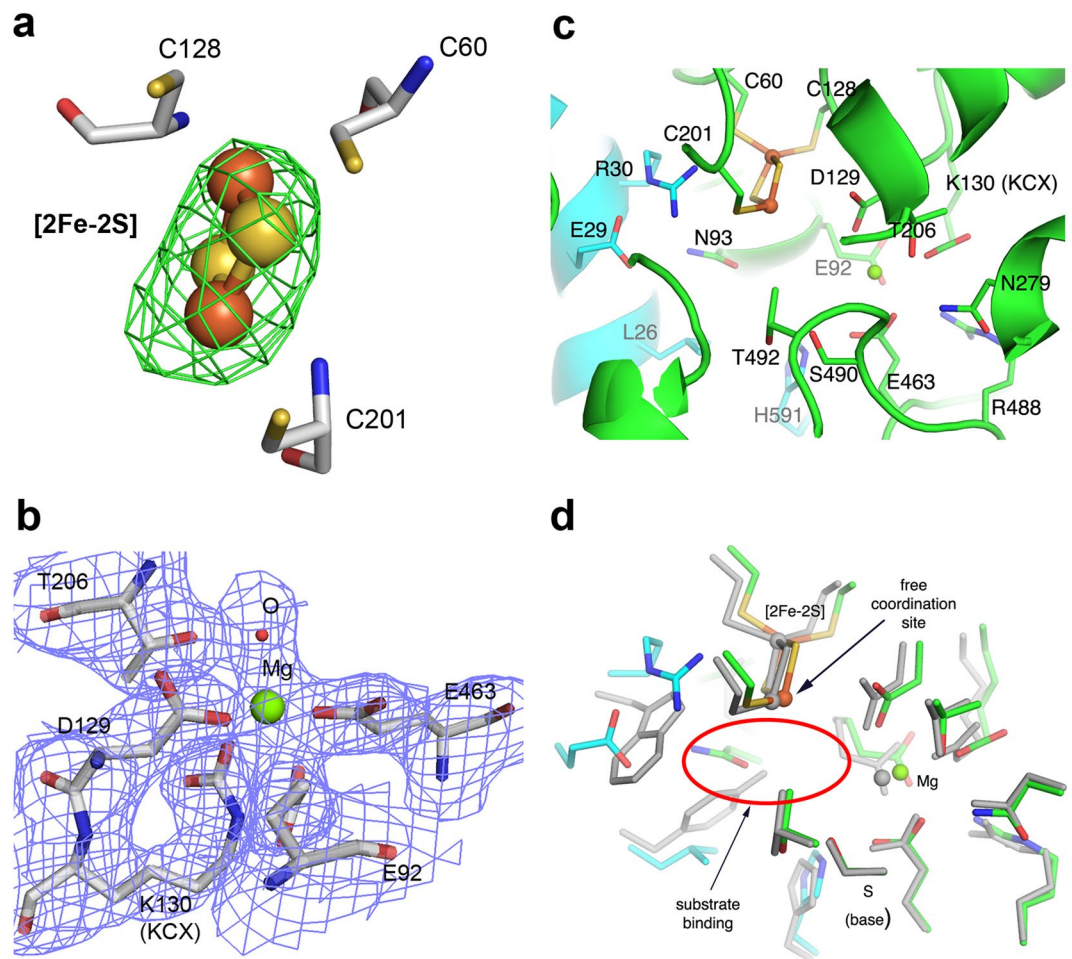


**Figure 4.** The monomeric structures of IlvD/EDD enzymes: (a) *CcXyDHT*, (b) *RlArDHT*, and (c) *So6PGDHT*. The N-terminal domain is in red, the C-terminal domain in blue, and the connecting region is shown in green. The undefined regions in the structure of *So6PGDHT* are shown as dashed lines. The arrows indicate the major differences between the structures: 1) the N-terminal region, 2) the flipped helix-loop-helix structure, and 3) the  $\beta$ -hairpin insertion of *So6PGDHT* in the connecting region. (d) The location of sequence motifs 1 and 2 in the *CcXyDHT* monomer. The conserved residues in both motifs are in purple, and other residues are in grey.

Lys130 is carbamylated (Fig. 5c,d) as is also found in the crystal structure of *RlArDHT*<sup>27</sup>. The residues Asp129 and Lys130 (or more correctly KCX130) belong to the N-terminal sequence motif. The complete structures of *CcXyDHT* and *RlArDHT* support an interpretation that carbamylation of this lysine is a common feature among IlvD/EDD enzymes.

The crystal structure of *CcXyDHT* suggests that the enzyme has a [2Fe-2S] cluster, similar to that observed in *RlArDHT*. In the cluster, the Fe1 atom is coordinated by two sulfide ions and two cysteines (Cys60 and Cys128). Cys128 belongs to the Cys-Asp-Lys motif, which is fully conserved among IlvD/EDD enzymes. Interestingly, Cys60 is conserved only among the subfamily of IlvD/EDD enzymes including pentonate dehydratases and some dihydroxy acid dehydratases<sup>19</sup>. In 6PGDHT, the putative corresponding cysteine belongs to a Cys-Asp-Gly motif





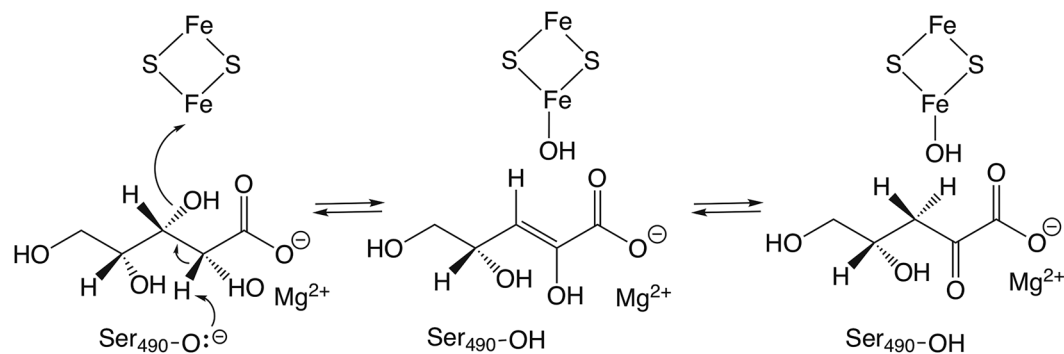
**Figure 5.** The active site of *CcXyDHT*. **(a)** The *mFo*-DFc omit maps (green) contoured at  $3\sigma$  level for the [2Fe-2S] cluster; and **(b)** The *2mFo*-DFc electron density maps (blue) contoured at  $1\sigma$  level for the  $Mg^{2+}$  ion and the binding residues; **(c)** the active site residues, which participate in the binding of the cofactors and the putative substrate binding residues are shown. The protein backbone cartoon colour is green for the first monomer and cyan for the second monomer; **(d)** The superimposition of *CcXyDHT* (in green and cyan) and *RIArDHT* (grey). The most likely binding site for the substrate is circled with a red ellipse.

(Fig. 1), which is not observed in pentonate dehydratases. This may reflect the organization of a different cluster type, for example, the binding of [2Fe-2S] versus [4Fe-4S] in the active site. In *CcXyDHT*, Fe2 atom of the cluster is coordinated by two sulfide ions and one cysteine residue (Cys200). This cysteine residue is fully conserved within the IlvD/EDD protein family.

Residues Ser490 and Thr492 from the C-terminal sequence motif participate in the formation of the substrate binding site (Fig. 5c,d). Based on the crystal structure and mutagenesis studies of *RIArDHT*, we have previously suggested that serine acts as a Lewis base in catalysis<sup>27</sup>. In *CcXyDHT*, Ser490 is approximately 6 Å away from the [2Fe-2S] cluster and 4 Å away from the  $Mg^{2+}$  ion. As in *RIArDHT*, OG1 of Thr492 forms a short (and strong) hydrogen bond (2.6 Å) with OG of Ser490 in *CcXyDHT*. In addition, a hydrogen bond (3.2 Å) exists between the main chain nitrogen of Thr492 and OG1 of Ser490. Furthermore, NE1 of Trp171 is hydrogen bonded (2.9 Å) to OG1 of Thr492. This hydrogen bond network may increase the nucleophilicity of Ser490. Thr492 in *CcXyDHT* and the corresponding Thr482 in *RIArDHT* are not conserved within the IlvD/EDD protein family (replaced by Ala/Gly) as seen in Fig. 3. However, the next amino acid after Ala/Gly in the sequence within the members of this protein family is Ser/Thr, which could play a similar role to Thr492 in the *CcXyDHT* structure. The proposed elimination reaction is shown in Fig. 6, where [2Fe-2S] cluster act as a Lewis acid and the alkoxide ion form of Ser490 is a base. In the end, the enol product is tautomerized to the final C2 oxo product.

**Molecular determinants for the substrate specificity.** To understand the substrate specificity between *CcXyDHT* and *RIArDHT*, the active sites were superimposed. The preferred substrates for *CcXyDHT* and *RIArDHT* are stereoisomers of pentonates, having different configurations at the C4 position. The  $K_m$  values for D-xylonate and L-arabinonate are in the mM range (2–10 mM) for both enzymes, but the  $k_{cat}$  values differ more significantly. The catalytic efficiency  $k_{cat}/K_m$  is 120 times higher in *CcXyDHT* for D-xylonate compared to L-arabinonate, and 10 times higher in *RIArDHT* for L-arabinonate compared to D-xylonate. In addition to





**Figure 6.** Proposed reaction mechanism for D-xylonate dehydratase.

pentonate sugar acids, various hexonate sugar acids can be dehydrated, as long as the configurations at C2 and C3 are maintained<sup>19</sup>.

The active site superimposition of *CcXyDHT* and *RIArDHT* shows that the binding of the  $Mg^{2+}$  and the [2Fe-2S] cluster is very similar, even identical in both enzymes. However, the differences in the active site cleft are significant in the N-terminal helix (residues 21–33) from the dimeric counterpart. The N-terminal helix can be therefore identified as a substrate recognition helix, where Trp31 in *RIArDHT* is replaced by Arg30 in *CcXyDHT*. This Arg residue forms a salt-bridge with Glu29 in *CcXyDHT* (the corresponding residue in *RIArDHT* is Gly30). Tyr27 in *RIArDHT* is replaced by Leu26 in *CcXyDHT*, and consequently, the conformation of Asn92 is altered. In addition, both enzymes have a C-terminal histidine residue from the dimeric counterpart (His591 in *CcXyDHT* and His579 in *RIArDHT*) in the active site (Fig. 5d). The carboxylate group of the C-terminus is salt-bridged to Arg22 and Lys422 in *RIArDHT* and to Arg174 in *CcXyDHT*. The differences between the catalytic efficiency thus probably originate from the recognition helix and C-terminus, which highlights the important role of the dimerization.

## Materials and Methods

**Protein purification and crystallization.** D-xylonate dehydratase from *Caulobacter crescentus* was heterologously expressed in *E. coli*, purified, and crystallized, and data were collected at 2.66 Å resolution using synchrotron radiation as described previously<sup>19,31</sup>. The data were processed and scaled using XDS and XSCALE<sup>32</sup>.

**Structure determination and refinement.** A crystal structure of *CcXyDHT* at 2.7 Å was determined by molecular replacement using PHASER<sup>33</sup> with the coordinates of L-arabinonate dehydratase (PDB code 5J84) as a template. The model building was performed in COOT<sup>34</sup>. The refinement of the model was performed using PHENIX refine<sup>35</sup>. The validation of the refinement was done using MolProbity<sup>36</sup>. The re-refinement of *So6PGDHT* (PDB code 2GP4) was also done using the PHENIX software<sup>35</sup>.

**UV-Vis spectroscopy.** The presence of a [Fe-S] cluster in *CcXyDHT* was analysed using UV-Vis spectroscopy. A purified protein sample was diluted with a sample storage buffer consisting of 50 mM Tris-HCl and 5 mM  $MgCl_2$  at pH 7.5. The sample was loaded into a quartz cuvette and data were collected using a UV-Vis/NIR/900 spectrophotometer (PerkinElmer, USA) at wavelengths of 260–800 nm against sample storage buffer. The spectrum is shown in Figure S1.

**Enzyme activity test.** The *CcXyDHT* activity was assayed as described previously<sup>19</sup>. The purified enzyme (6 µg in a 600 µl reaction) was incubated at 30 °C in 50 mM Tris-Cl, 10 mM  $MgCl_2$ , at pH 8.5 using 20 mM D-xylonate as substrate. At 2-min intervals, a 100-µl sample was transferred to a microcentrifuge tube and the reaction was stopped by adding 10 µl of 12% trichloroacetic acid (TCA). The product formation was measured using the thiobarbituric acid (TBA) assay<sup>37</sup>. Aliquots of 50 µl from all samples were transferred into fresh microcentrifuge tubes, after which 125 µl of 25 mM periodic acid (dissolved in 0.2 M  $H_2SO_4$ ) was added and incubated at room temperature for 20 min. To the reaction, 250 µl of 2% sodium arsenate in 0.5 M HCl was added and mixed, followed by 1 ml of 0.3% TBA. The samples were then incubated at 100 °C for 10 min. Prior to reading the absorbance at 549 nm, the samples were mixed with an equal volume of DMSO.

**Sequence alignment and structural comparisons.** The amino acid sequences of the enzymes were collected from the NCBI GeneBank<sup>38</sup>. The sequence alignment was carried out using the online multiple sequence alignment tools Clustal Omega<sup>39</sup> and MacVector<sup>40</sup>. The secondary structure analysis was carried out using the online tool STRIDE<sup>41</sup>. The alignment figure was created using ALINE<sup>42</sup>. The superimposition of the crystal structures and the figures of the three-dimensional enzyme structures were made using PyMOL<sup>43</sup>.

**Data availability.** The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 5OYN.

## References

- Lang, K., Zierow, J., Buehler, K. & Schmid, A. Metabolic engineering of *Pseudomonas* sp. strain VLB120 as platform biocatalyst for the production of isobutyric acid and other secondary metabolites. *Microb. Cell. Fact.* **13**, 2 (2014).
- Blombach, B. & Eikmanns, B. J. Current knowledge on isobutanol production with *Escherichia coli*, *Bacillus subtilis* and *Corynebacterium glutamicum*. *Bioeng. Bugs* **2**, 346–350 (2011).
- Li, S., Wen, J. & Jia, X. Engineering *Bacillus subtilis* for isobutanol production by heterologous Ehrlich pathway construction and the biosynthetic 2-ketoisovalerate precursor pathway overexpression. *Appl. Microbiol. Biotechnol.* **91**, 577–589 (2011).
- Connor, M. R. & Liao, J. C. Engineering of an *Escherichia coli* strain for the production of 3-methyl-1-butanol. *Appl. Environ. Microbiol.* **74**, 5769–5775 (2008).
- Niu, W., Molefe, M. N. & Frost, J. W. Microbial synthesis of the energetic material precursor 1,2,4-butanetriol. *J. Am. Chem. Soc.* **125**, 12998–12999 (2003).
- Guterl, J. K. *et al.* Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem* **5**, 2165–2172 (2012).
- Egan, S. E. *et al.* Molecular characterization of the Entner-Doudoroff pathway in *Escherichia coli*: sequence analysis and localization of promoters for the *edd-eda* operon. *J. Bacteriol.* **174**, 4638–4646 (1992).
- Carsten, J. M., Schmidt, A. & Sieber, V. Characterization of recombinantly expressed dihydroxy-acid dehydratase from *Sulfobus solfataricus*-A key enzyme for the conversion of carbohydrates into chemicals. *J. Biotechnol.* **211**, 31–41 (2015).
- Velasco, J. A. *et al.* Cloning of the dihydroxyacid dehydratase-encoding gene (*ILD3*) from *Saccharomyces cerevisiae*. *Gene* **137**, 179–185 (1993).
- Bromke, M. A. Amino Acid Biosynthesis Pathways in Diatoms. *Metabolites* **3**, 294–311 (2013).
- Zhang, C. *et al.* Dihydroxyacid dehydratase is important for gametophyte development and disruption causes increased susceptibility to salinity stress in *Arabidopsis*. *J. Exp. Bot.* **66**, 879–888 (2015).
- Kanamori, M. & Wixom, R. L. Studies in valine biosynthesis. V. Characteristics of the purified dihydroxyacid dehydratase from spinach leaves. *J. Biol. Chem.* **238**, 998–1005 (1963).
- Myers, J. W. Dihydroxy acid dehydrase: an enzyme involved in the biosynthesis of isoleucine and valine. *J. Biol. Chem.* **236**, 1414–1418 (1961).
- Conway, T. The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiol. Rev.* **9**, 1–27 (1992).
- Gottschalk, G. & Bender, R. D-Gluconate dehydratase from *Clostridium pasteurianum*. *Methods Enzymol.* **90**(Pt E), 283–287 (1982).
- Meloche, H. P. & Wood, W. A. The Mechanism of 6-Phosphogluconic Dehydrase. *J. Biol. Chem.* **239**, 3505–3510 (1964).
- Stephens, C. *et al.* Genetic analysis of a novel pathway for D-xylose metabolism in *Caulobacter crescentus*. *J. Bacteriol.* **189**, 2181–2185 (2007).
- Weimberg, R. Pentose oxidation by *Pseudomonas fragi*. *J. Biol. Chem.* **236**, 629–635 (1961).
- Andberg, M. *et al.* Characterization and mutagenesis of two novel iron-sulphur cluster pentonate dehydratases. *Appl. Microbiol. Biotechnol.* **100**, 7549–7563 (2016).
- Watanabe, S., Shimada, N., Tajima, K., Kodaki, T. & Makino, K. Identification and characterization of L-arabonate dehydratase, L-2-keto-3-deoxyarabonate dehydratase, and L-arabinolactonase involved in an alternative pathway of L-arabinose metabolism. Novel evolutionary insight into sugar metabolism. *J. Biol. Chem.* **281**, 33521–33536 (2006).
- Jiang, Y. *et al.* Characterization of D-xylonate dehydratase YjhG from *Escherichia coli*. *Bioengineered* **6**, 227–232 (2015).
- Kim, S. & Lee, S. B. Identification and characterization of the bacterial d-gluconate dehydratase in *Achromobacter xylosoxidans*. *Biotechnol. Bioproc E* **13**, 436–444 (2008).
- Flint, D. H., Emptage, M. H., Finnegan, M. G., Fu, W. & Johnson, M. K. The role and properties of the iron-sulfur cluster in *Escherichia coli* dihydroxy-acid dehydratase. *J. Biol. Chem.* **268**, 14732–14742 (1993).
- Flint, D. H. & Emptage, M. H. Dihydroxy acid dehydratase from spinach contains a [2Fe-2S] cluster. *J. Biol. Chem.* **263**, 3558–3564 (1988).
- Flint, D. H., Rothman, S. C., Suh, W. & Tomb, J. F. Identification and use of bacterial [2Fe-2S] dihydroxy-acid dehydratases, Patent WO2010037112 (A1) (2010).
- Rodriguez, M., Wedd, A. G. & Scopes, R. K. 6-phosphogluconate dehydratase from *Zymomonas mobilis*: an iron-sulfur-manganese enzyme. *Biochem. Mol. Biol. Int.* **38**, 783–789 (1996).
- Rahman, M. M. *et al.* The crystal structure of a bacterial L-arabinonate dehydratase contains a [2Fe-2S] cluster. *ACS Chem. Biol.* **12**, 1919–1927 (2017).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- Scopes, R. K. & Griffiths-Smith, K. Use of differential dye-ligand chromatography with affinity elution for enzyme purification: 6-phosphogluconate dehydratase from *Zymomonas mobilis*. *Anal. Biochem.* **136**, 530–534 (1984).
- Hulo, N. *et al.* The PROSITE database. *Nucleic Acids Res.* **34**, D227–30 (2006).
- Rahman, M. M., Andberg, M., Koivula, A., Rouvinen, J. & Hakulinen, N. Crystallization and X-ray diffraction analysis of an L-arabinonate dehydratase from *Rhizobium leguminosarum* bv. trifolii and a D-xylonate dehydratase from *Caulobacter crescentus*. *Acta Crystallogr. F. Struct. Biol. Commun.* **72**, 604–608 (2016).
- Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
- Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
- Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
- Buchanan, C. L., Connaris, H., Danson, M. J., Reeve, C. D. & Hough, D. W. An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. *Biochem. J.* **343**(Pt 3), 563–570 (1999).
- Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–72 (2016).
- Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
- Olson, S. A. MacVector: an integrated sequence analysis program for the Macintosh. *Methods Mol. Biol.* **25**, 195–201 (1994).
- Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500–2 (2004).
- Bond, C. S. & Schuttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 510–512 (2009).
- The PyMOL Molecular Graphics System, Version 1. 8 Schrödinger, LLC, Portland, OR (2015).

## Acknowledgements

We thank Dr. Steffi Goffart and Dr. Jaakko Pohjoismäki, Department of Biology, UEF for their assistance during the protein expression. We thank the Diamond Light Source (DLS) for providing synchrotron access. At UEF, this work received financial support from the Academy of Finland (Decisions 256937 and 292705). The work was part of the Finnish Centre of Excellence in the White Biotechnology-Green Chemistry programme (Academy of Finland Decision 118573) and the IV4SP project (Academy of Finland Decision 272598) at VTT. The work also received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under BioStruct-X (Grant Agreement N° 283570).

## Author Contributions

M.M.R. and M.A. expressed and purified the enzyme. M.M.R. and N.H. performed the crystallization, data collection, structure determination and refinement. All authors discussed the results and contributed to the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-19192-6>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018