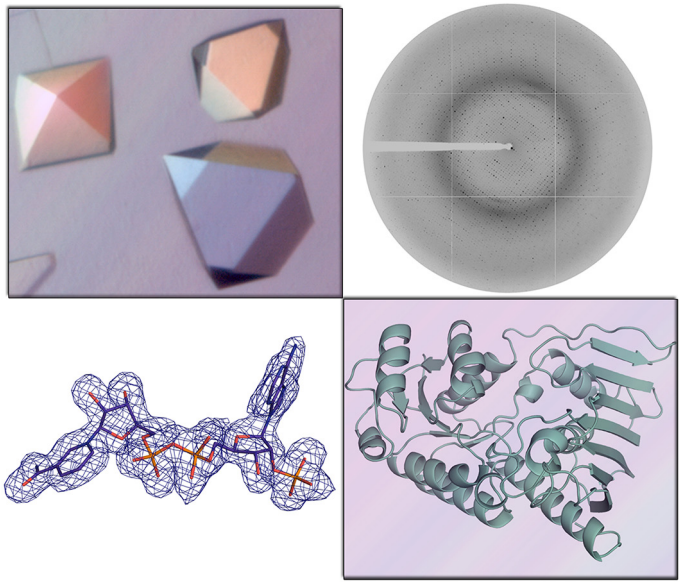
Dissertations
Department of Chemistry
University of Eastern Finland

No. 130 (2015)

Helena Taberman

Structure and function of carbohydrate-modifying enzymes



Structure and function of carbohydrate-modifying enzymes

Helena Taberman

Department of Chemistry
University of Eastern Finland
Finland

Joensuu 2015

Helena Taberman
Department of Chemistry, University of Eastern Finland
P. O. Box 111, FI-80101 Joensuu, Finland

Supervisors
Prof. Juha Rouvinen, University of Eastern Finland
Prof. Janne Jänis, University of Eastern Finland
Dr. Tarja Parkkinen, University of Eastern Finland

Referees
Docent Tommi Kajander, University of Helsinki
Docent Ossi Turunen, Aalto University

Opponent
Prof. Inari Kursula, University of Bergen

To be presented with the permission of the Faculty of Science and Forestry of the University of Eastern Finland for public criticism in Auditorium F100, Yliopistokatu 7, Joensuu, on November 27th 2015 at 12 o'clock noon.

ISSN 2242-1033
ISBN 978-952-61-1940-3 (PDF)

Grano Oy Joensuu
Joensuu 2015

ABSTRACT

Carbohydrate-modifying enzymes could be used in producing sustainable fuels and chemicals from plant biomass. The three-dimensional structure of these potential enzymes needs to be solved in order to reveal their function and to be able to develop more efficient biocatalysts and microbial routes necessary for industrial use. In this study, X-ray crystallography was used for the structural characterization of three carbohydrate-modifying enzymes: galactarolactone cycloisomerase and keto-deoxy-D-galactarate dehydratase from *Agrobacterium tumefaciens* and aldose-aldose oxidoreductase from *Caulobacter crescentus*.

Galactarolactone cycloisomerase and keto-deoxy-D-galactarate dehydratase from *Agrobacterium tumefaciens* work in the oxidative pathway of D-galacturonate, which is the main component of pectin. Galactarolactone cycloisomerase is an enzyme that catalyzes the conversion of the cyclic D-galactaro-1,4-lactone into a linear 3-deoxy-2-keto-L-*threo*-hexarate. The crystal structure of galactarolactone cycloisomerase was solved to 1.6 Å resolution. Its tertiary structure is a typical enolase family enzyme structure with two domains: an N-terminal α+β capping domain and a modified TIM-barrel domain. Furthermore, it has a long C-terminal extension packed against the N-terminal domain. The quaternary structure is octameric. The attempts to solve the complex structures of galactarolactone cycloisomerase were unsuccessful.

Keto-deoxy-D-galactarate dehydratase is a novel decarboxylating hydrolyase that follows galactarolactone cycloisomerase in the oxidative pathway of D-galacturonate. It converts the linear hexarate into α-ketoglutaric semialdehyde. The three-dimensional structures of keto-deoxy-D-galactarate dehydratase without a ligand, in complex with pyruvate, and in complex with 2-oxoadipic acid were determined to 1.7, 1.5, and 2.1 Å resolutions, respectively. The structures allowed us to propose a reaction mechanism that was based on the grounds of these unique crystal structures and mass spectrometry measurements.

Aldose-aldose oxidoreductase from *Caulobacter crescentus* is an enzyme catalyzing an NADP-dependent oxidation/reduction cycle of several aldose monosaccharides to the corresponding aldonolactones and alditols, respectively. Furthermore, it was also found to be able to oxidize 1,4-linked polysaccharides. Its complex structures with glycerol, D-xylose, D-glucose, maltotriose, and D-sorbitol were processed to 2.0, 1.8, 1.7, 1.9, and 1.8 Å resolutions, respectively. These complex structures provided new insights into the reaction mechanism of this intriguing enzyme that is able to modify a panel of biomass sugar substrates. The structural and functional features of the Gfo/Idh/MocA protein family of *Caulobacter crescentus* aldose-aldose oxidoreductase were also studied, and an extensive description is presented.

LIST OF ORIGINAL PUBLICATIONS

This dissertation is a summary of publications I-III, the submitted manuscript IV, and some unpublished data.

I Taberman, H., Andberg, M., Parkkinen, T., Richard, P., Hakulinen, N., Koivula, A., and Rouvinen, J. (2014) Purification, crystallization and preliminary X-ray diffraction analysis of a novel keto-deoxy-D-galactarate (KDG) dehydratase from *Agrobacterium tumefaciens*. *Acta Crystallogr. F* **70**, 49-52.

II Taberman, H., Andberg, M., Parkkinen, T., Jänis, J., Penttilä, M., Hakulinen, N., Koivula, A., and Rouvinen, J. (2014) Structure and function of a decarboxylating *Agrobacterium tumefaciens* keto-deoxy-D-galactarate dehydratase. *Biochemistry* **53**, 8052-8060.

III Taberman, H., Andberg, M., Koivula, A., Hakulinen, N., Penttilä, N., Rouvinen, J., and Parkkinen, T. (2015) Structure and function of *Caulobacter crescentus* aldose-aldose oxidoreductase. *Biochem. J.* doi: 10.1042/BJ20150681.

IV Taberman, H., Parkkinen, T., and Rouvinen, J. (2015) Structural and functional features of the Gfo/Idh/MocA protein family. *Submitted for publication.*

The author has carried out all the crystallization experiments, X-ray data collection and processing, and structure solving in the publications **I-III** and in the unpublished data. Also the mass spectrometry experiments in the publication **II** and the UV/vis measurements in publication **III** were performed by the author. The author did all the bioinformatics in the publication **IV**. The author has analyzed the results and written the manuscripts in collaboration with the coauthors.

CONTENTS

ABBREVIATIONS

AFR	1,5-anhydro-D-fructose reductase
α-KGSA	α-ketoglutaric semialdehyde
asu	asymmetric unit
At	*Agrobacterium tumefaciens*
A-zyme	α-N-acetylgalactosidase
BVR	biliverdin reductase
Cc AAOR	aldose-aldose oxidoreductase from *Caulobacter crescentus*
CID	collision-induced dissociation
Da	dalton, 1 Da = 1g/mol
DD	dihydrodiol dehydrogenase
ESI	electrospray ionization
ESRF	European Synchrotron Radiation Facility
G6PD	glucose-6-phosphate dehydrogenase
Gal80p	transcriptional inhibitor Gal80p
Gci	galactarolactone cycloisomerase
Gfo	glucose-fructose oxidoreductase
GLI	galactaro-δ-lactone isomerase
IDH	inositol dehydrogenase
KDG	keto-deoxy-D-galactarate
KijD10	C-3”-ketoreductase
MocA	rhizopine catabolism protein
MS/MS	tandem mass spectrometry
m/z	mass-to-charge ratio
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
PDB	Protein Data Bank
QIT	quadrupole ion trap
rmsd	root-mean-square deviation
Ss KDGA	keto-3-deoxygluconate aldolase from *Sulfobus solfataricus*
Udh	uronate dehydrogenase
UV/vis	ultraviolet/visible
WlbA	uridine-5’-phosphate-2,3-diacetamido-2,3-dideoxy-D-mannuronic acid dehydrogenase
Zm GFOR	glucose-fructose oxidoreductase from *Zymomonas mobilis*
Å	ångström, 1 Å = 1×10^{-10} m

1. INTRODUCTION

On Earth, the most abundant renewable carbon source is biomass. It is organic material that has stored energy through photosynthesis. The unexploited, but substantial sources of plant-based material are usually byproducts of forestry and agriculture, for example, cellulose, hemicellulose, lignin, and pectin. These can be used as cattle feed or burned to energy, but a considerable part is thought to be merely waste. In the search of finding new sustainable ways of producing fuels and chemicals, they could also be used as a source of raw materials. Cellulose-derived glucose has already been widely studied and is an applied monosaccharide source in industry[1-3]. To make the utilization of the plant biomass more effective, less explored carbohydrate sources, hemicellulose and pectin, should also be considered.

The natural polymers, hemicellulose and pectin, composed of carbohydrates are used as a carbon and energy source by many microorganisms. There are several known plant material decaying and modifying proteins, i.e., enzymes that exist in fungi and bacteria, which could be harnessed in bioproduction. Enzymes are already extensively used in food, pharmaceutical and cosmetics industries, and the application of biocatalysts has led to more economical and environmentally friendly processes[4-6].

In order to be able to apply new enzymes for the biotechnological production of platform chemicals, the enzyme function has to be determined. X-ray crystallography is an accurate and precise method for revealing the atomic-detail three-dimensional structures of macromolecules. Crystal structures will provide new insights into the protein-ligand interactions and the enzymatic reaction mechanisms. Structure-function studies are crucial for the comprehensive understanding of the microbial metabolism of carbohydrates, which could then be used in developing new applications for the conversion of biomass into fuel or chemicals.

In this work the structure and function of three carbohydrate-modifying enzymes were studied: galactarolactone cycloisomerase and keto-deoxy-D-galactarate dehydratase from the oxidative D-galacturonate pathway of *Agrobacterium tumefaciens* and aldose-aldose oxidoreductase from *Caulobacter crescentus*. Also a bioinformatical study of the Gfo/Idh/MocA protein family was performed. A short introduction on the three enzymes is presented in the following sections.

"*Enzymes are things invented by biologists*
that explain things which otherwise require harder thinking."

Jerome Lettvin 1920-2011

1.1. THE OXIDATIVE D-GALACTURONATE PATHWAY

Pectin is a complex heteropolysaccharide rich in the primary cell walls and middle lamellas of terrestrial plants, such as citrus and sugar beet[7, 8]. Several million tons of pectin containing residues are produced every year, which is currently only used as cattle feed thus creating a disposal problem[9]. The microbial conversion of these carbohydrate rich residues into value-added products requires the extensive identification of the existing pathways and the enzymes working on them.

The main component of pectin is D-galacturonate, a sugar acid, which composes approximately 70% of pectin[10]. Two catabolic bacterial D-galacturonate pathways have been found: the isomerase and the oxidative pathway[11-13]. The oxidative pathway has been described to function in *Agrobacterium tumefaciens* (*At*)[14] and *Pseudomonas syringae*[15].

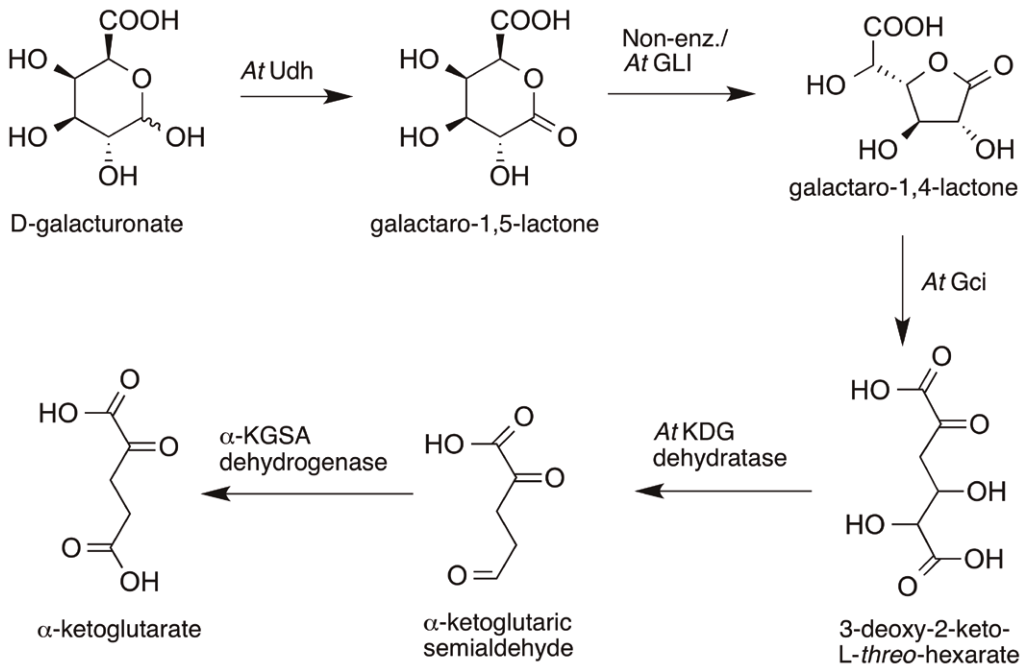


Figure 1. The oxidative D-galacturonate pathway in *Agrobacterium tumefaciens*.

In the oxidative pathway (Fig. 1), the pyranose form of D-galacturonate is first oxidized to galactaro-1,5-lactone by an NAD^+-dependent uronate dehydrogenase (Udh, EC 1.1.1.203)[16, 17]. Galactaro-1,5-lactone is then rearranged to galactaro-1,4-lactone either non-enzymatically or with the help of galactaro-δ-lactone isomerase (GLI)[18]. Next galactarolactone cycloisomerase (Gci, EC 5.5.1.-) catalyzes a novel ring opening

reaction into a linear 3-deoxy-2-keto-L-*threo*-hexarate[19], which is then dehydrated and decarboxylated to α-ketoglutaric semialdehyde by a decarboxylating hydrolyase, keto-deoxy-D-galactarate (KDG) dehydratase (EC 4.2.1.41)[20, 21]. Finally, α-ketoglutaric semialdehyde (α-KGSA) dehydrogenase (EC 1.2.1.26) oxidizes α-KGSA to α-ketoglutarate, a metabolite of the tricarboxylic acid cycle[21, 22].

1.1.1. GALACTAROLACTONE CYCLOISOMERASE

At Gci is an enzyme that catalyzes a ring opening reaction in which it converts D-galactaro-1,4-lactone into a linear 3-deoxy-2-keto-L-*threo*-hexarate. *At* Gci belongs to the divergent enolase superfamily and shows approximately 30% sequence identity to sugar acid dehydratases. According to a three-dimensional protein structure comparison within the DALI-server[23], the most similar structures are an enolase Spea_3858 from *Shewanella pealeana*[24], a putative dehydratase from *Mesorhizobium loti*[25], and a putative enolase from *Salmonella typhimurium* Lt2[26].

The enzymes that belong to the enolase superfamily have a two-domain structure with an α+β capping domain and a $(\alpha/\beta)_7\beta$ domain. The enolase superfamily can be divided into seven subgroups[27]: 1) enolase, 2) mandelate racemase, 3) muconate lactonizing enzyme, 4) 3-methylaspartate ammonia lyase, 5) D-glucarate dehydratase, 6) D-mannonate dehydratase, and 7) galactarate dehydratase subgroup. The partition of the enzymes into the subgroups is allowed by the differences in their active site structures. *At* Gci seems to belong to the mandelate racemase subgroup[19].

The enolase superfamily enzymes have similarities in their reaction mechanisms, although catalyzing different overall reactions, such as the elimination of water or ammonia and interconversion of enantiomers[28, 29]. They use a common partial reaction mechanism in which an active site base abstracts an α-proton of the carboxylic acid containing substrate. This results in an intermediate that is stabilized by the coordination to a Mg^{2+} ion. This essential Mg^{2+} ion and the catalytic residues are situated at the C-terminal side of the modified TIM-barrel fold.

1.1.2. KETO-DEOXY-D-GALACTARATE DEHYDRATASE[I, II]

In *A. tumefaciens*, the conversion of the linear C6 hexarate into a C5 compound is done by KDG dehydratase. *At* KDG dehydratase is an enzyme belonging to the Class I aldolase superfamily[30]. These enzymes share a similar $(\alpha/\beta)_8$ TIM-barrel fold and catalyze various different reactions through a Schiff-base intermediate formed between a conserved active site lysine and the α-keto acid of the substrate[31]. Members in this family catalyze, for example, condensation, dehydration, and retro-aldol condensation reactions[32, 33].

There are several crystal structures solved from the Class I aldolase family, which include: N-acetylneuraminate lyase from *Escherichia coli*[34], D-fructose-1,6-bisphosphate aldolase from *Oryctolagus cuniculus*[35], and 2-keto-3-deoxy-6-phosphogluconate aldolase from *E. coli*[36]. Nevertheless, no structures and only a few kinetic characterization studies exist of the D-4-deoxy-5-oxoglucarate dehydratase that catalyzes an identical reaction as *At* KDG dehydratase[21, 37, 38]. *At* KDG dehydratase is a novel type of decarboxylating and dehydrating enzyme and may be of great importance for future industrial processes.

1.2. ALDOSE-ALDOSE OXIDOREDUCTASE[III, IV]

Caulobacter crescentus aldose-aldose oxidoreductase (*Cc* AAOR, EC 1.1.99.-) is a newly found enzyme that was discovered through its sequence homology to the xylose dehydrogenase[39] during the search of novel D-xylose converting enzymes[40]. D-xylose is one of the most abundant monosaccharides and a major component of hemicellulose[41], but its potential in biotechnological applications has not yet been fully elucidated. In the preceding studies, *Cc* AAOR was found to be able to both oxidize and reduce a panel of aldose saccharides to corresponding aldonolactones and alditols, respectively[42]. No addition of an external cofactor or artificial electron acceptor was needed in this single ping-pong reaction cycle[43] in which the tightly bound NADP(H) cofactor is regenerated. *Cc* AAOR has the conserved Class I NAD(P) binding sequence-motif, Gly_{13}-Xaa-Gly-Xaa-Xaa-Ala_{18} [44]. A schematic presentation of the reaction cycle is presented in Figure 2.

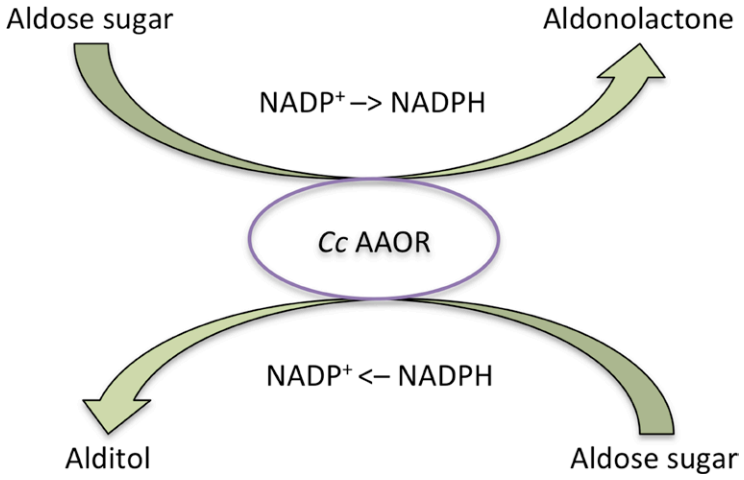


Figure 2. Schematic presentation of the *Cc* AAOR reaction cycle.

Interestingly enough, *Cc* AAOR was not only able to oxidize and reduce monosaccharides, but it was also found to be able to oxidize 1,4-linked polysaccharides[42]. Currently only a limited set of enzymes able to oxidize oligosaccharides is known[45-50]. This kind of a novel oxidoreductase having a broad substrate specificity and a dual activity could have a substantial potential in producing sugar acids and alditols from sugar hydrolysates.

Cc AAOR belongs to the diverse Gfo/Idh/MocA protein family, which has not been previously described. The abbreviations Gfo, Idh, and MocA refer to glucose-fructose oxidoreductase, inositol 2-dehydrogenase, and the rhizopine catabolism protein MocA, which could catalyze a dehydrogenase reaction involved in rhizopine catabolism. The family contains several different enzymes that catalyze reactions, such as the oxidation and reduction of carbohydrates[51-56], reduction of biliverdin[57], hydrolyzation of glycosidic bonds[58], and oxidation of *trans*-dihydrodiols[59]. Furthermore, structurally, a transcriptional inhibitor belongs to this family, although it has no enzymatic activity[60].

All the enzymes in this family catalyze NAD(P)-dependent reactions, even the transcriptional inhibitor have been shown to associate with dinucleotides[61]. Even though the sequence identity between most of the members in this family is almost absent, they all share a similar two-domain structure that has a classical N-terminal nucleotide binding Rossman fold[62, 63] and a C-terminal domain with a large β-sheet motif.

1.3. AIMS OF THE STUDY

The goal of this work was to solve the X-ray structures of different carbohydrate-modifying enzymes in order to elucidate their reaction mechanisms. The specific aims were:

1) To crystallize and determine the structures of *At* Gci, *At* KDG dehydratase, and *Cc* AAOR

2) To determine complex structures of the enzymes with their substrates or products or their analogues

3) To identify the roles of the active site amino acid residues

4) To elucidate the reaction mechanisms according to the three-dimensional structures and high-resolution mass spectrometry measurements

5) To do a thorough bioinformatical analysis of the Gfo/Idh/MocA family according to the structure and function of the proteins belonging to the family

2. MATERIALS AND METHODS

2.1. MATERIALS

The enzymes were expressed, produced, and purified at VTT Technical Research Centre of Finland. *At* Gci with a C-terminal *Strep*-tag II was expressed in *E. coli* strain BL21(DE3) transforming it with the generated plasmid, pBAT4-*At gci*. The enzyme was purified from the cell extract in a single-step using a *Strep*-Tactin Superflow column. Furthermore, *At* KDG dehydratase also with a C-terminal *Strep*-tag II and its site-directed *Strep*-tagged S211A mutant were expressed in *E. coli* and purified using essentially the same protocol as for *At* Gci. *Cc* AAOR was expressed in *Saccharomyces cerevisiae* strain CEN.PK 113-17A and purified from the cell extract in a single cation exchange chromatography step using a SP FF column.

2.2. CRYSTALLIZATION

All crystallization experiments were performed with the hanging-drop vapor-diffusion method mixing 2 µl of enzyme with an equal volume of reservoir solution and equilibrated against 0.5 ml of reservoir solution at 20°C. Initial crystallization conditions (Fig. 3A, C, and E) were found with commercial screening kits (Hampton Research) and optimized further by varying the used reagents and their concentrations, and the pH. Furthermore, changing the protein concentration was studied.

The best crystals of *At* Gci for synchrotron source measurements were grown in 0.1 M sodium acetate pH 5.0, 0.2 M calcium chloride, 4% 2-propanol after addition of a detergent C-HEGA®-10 (Detergent screen, Hampton Research) into the conditions. The crystals (Fig. 3B) grew approximately in 1 to 2 days to dimensions 0.3×0.4×0.3 mm. The crystals of *At* KDG dehydratase for X-ray diffraction measurements were obtained in conditions containing 0.1 M Bicine pH 8.5, 0.2 M sodium formate, and 15% (w/v) polyethylene glycol monomethyl ether 5000. Three-dimensional crystals in different shapes (Fig. 3D) grew in 3 to 4 days to dimensions of 0.4×0.2×0.2 mm approximately. X-ray diffraction quality crystals of *Cc* AAOR were obtained under several conditions (1.6 M ammonium/magnesium sulphate, 0.1 M MES monohydrate pH 6.5, 7-10% 1,4-dioxane/2-propanol/glycerol). Plate-like crystals (Fig. 3F) grew in 7 to 10 days to dimensions of 0.4 × 0.2 × 0.05 mm approximately.

The crystals for the complex structure determinations were prepared by soaking the crystals in a cryoprotectant solution containing the ligand in question. The crystals were soaked for approximately 0.5 minutes before flash-cooling with liquid nitrogen. In the case of *Cc* AAOR complexes with glycerol, D-xylose, D-glucose, and D-sorbitol, the ligand acted also as a cryoprotectant. For *At* Gci and *At* KDG dehydratase,

glycerol acted as a cryoprotectant and for the *Cc* AAOR in complex with maltotriose, sodium malonate was the cryoprotectant.

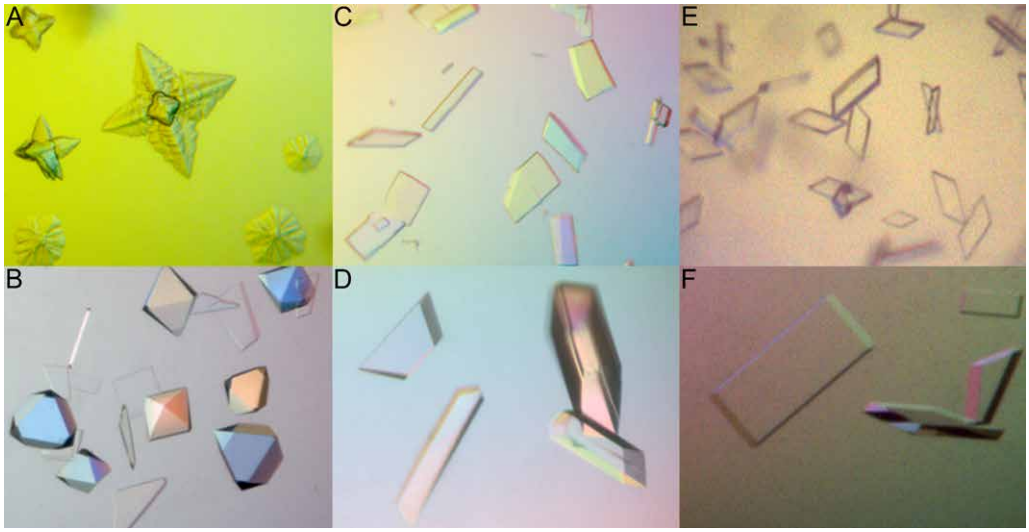


Figure 3. A) Initial *At* Gci crystals, **B)** *At* Gci crystals after optimization, **C)** Initial *At* KDG dehydratase crystals, **D)** *At* KDG dehydratase crystals after optimization, **E)** Initial *Cc* AAOR crystals, **F)** *Cc* AAOR crystals after optimization.

2.3. UV/VIS SPECTROSCOPY

UV/vis spectroscopy was used to determine the oxidization state of the NADP(H) cofactor in the *Cc* AAOR crystals. The appearance of a major absorption band at 340 nm on the UV/vis spectrum is indicative for the reduced cofactor and can be used to distinguish the oxidation/reduction state of nicotinamide cofactor[64]. The spectra were measured before the X-ray diffraction experiments, except the sorbitol complex crystal that was measured after X-ray radiation, in an offline mode at cryogenic temperature at the Cryobench laboratory ID29S at European Syncrotron Radiation Facility (ESRF, Grenoble, France)[65].

2.4. MASS SPECTROMETRY

Mass spectrometry was used to confirm the reaction end-products of the enzymatic reactions of *At* KDG dehydratase and its site-directed S211A mutant, which was prepared to study the role of the conserved Ser-211 in the reaction mechanism. The measurements were performed on a quadrupole ion trap (QIT) instrument (Esquire 3000 Plus, Bruker Daltonics, Bremen, Germany) using negative-ion electrospray

ionization (ESI)[66]. The ions for the observed enzymatic reaction end-products were further subjected to tandem mass spectrometry (MS/MS) by using collision-induced dissociation (CID).

2.5. X-RAY DATA COLLECTION AND PROCESSING

The X-ray diffraction data of the flash-cooled crystals were measured with synchrotron radiation at ESRF[67, 68] and Diamond Light Source (Didcot, United Kingdom). All data sets were processed using XDS. The space groups were determined with POINTLESS[69], and the data quality and twinning was analyzed with phenix.xtriage[70, 71]. The data sets of *At* Gci, the *At* KDG dehydratase complex structures and the *Cc* AAOR glycerol complex structure were also scaled with XSCALE[72]. The data collection statistics are summarized in Tables 1 and 2 in which the values in parentheses are for the highest resolution shell.

Table 1. Data collection and refinement statistics for *At* Gci and *At* KDG dehydratase.

Enzyme	***At* Gci**	***At* KDG dehydratase**	***At* KDG dehydratase**	***At* KDG dehydratase**
Ligand	**-**	**-**	**Pyruvate**	**2-Oxoadipic acid**
Space Group	I4	C2	C2	C2
Molecules in asu	2	4	4	4
Unit cell *a* (Å)	110.0	168.9	169.8	169.5
b (Å)	110.0	118.4	120.1	119.0
c (Å)	139.8	74.3	74.1	74.3
β (°)	90	112.3	112.0	112.2
Resolution (Å)	1.6	1.7	1.5	2.1
R_{sym} (%)	5.3 (27.1)	7.2 (57.8)	3.8 (50.9)	13.6 (71.0)
$CC_{1/2}$	99.7 (93.5)	99.6 (80.4)	99.9 (79.4)	98.8 (71.3)
Completeness (%)	96.8 (89.3)	98.1 (95.6)	97.7 (97.4)	96.0 (92.8)
No. of unique reflections	105668 (9343)	159523 (25113)	214991 (15841)	76785 (5470)
Redundancy	3.0 (2.0)	3.5 (3.4)	2.8 (2.6)	3.5 (3.5)
$I/\sigma I$	12.6 (2.9)	10.0 (2.0)	15.6 (2.3)	8.4 (2.0)
R_{work} (%)/R_{free} (%)	15.1/18.1	17.4/20.4	16.3/18.2	22.1/27.2
No. of atoms	11382	10508	10963	9603
Wilson *B*-factor	17.1	22.4	18.3	28.3
Average B-factor (Å^2)	25.3	26.9	23.7	32.5
Protein	24.1	25.8	21.9	32.5
rmsd bond lengths (Å)	0.011	0.007	0.006	0.009
rmsd bond angles (°)	1.2	1.1	1.1	1.2
Ramachadran favoured (%)	97.7	92.6	92.5	91.7
PDB code	-	4UR5	4UR7	4UR8

Table 2. Data collection and refinement statistics for *Cc* AAOR.

Ligand	**Glycerol**	**D-xylose**	**D-glucose**	**Maltotriose**	**D-sorbitol**
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Molecules in asu	6	6	6	6	6
Unit cell *a* (Å)	101.3	101.2	101.1	100.3	101.2
b (Å)	153.6	153.4	153.5	151.5	154.1
c (Å)	108.0	107.2	108.0	108.7	107.8
β (°)	109.9	109.5	109.3	108.7	110.0
Resolution (Å)	2.0	1.8	1.7	1.9	1.8
R_{svm} (%)	11.1 (50.4)	9.5 (55.8)	6.2 (59.9)	8.5 (61.9)	9.6 (67.7)
$CC_{1/2}$	99.2 (71.5)	99.5 (74.3)	99.8 (70.1)	99.7 (72.5)	99.6 (64.4)
Completeness (%)	99.5 (99.1)	96.9 (92.0)	99.0 (97.0)	98.9 (97.5)	98.6 (92.7)
No. of unique reflections	208078 (15236)	256341 (39306)	338584 (53566)	239447 (38100)	264033 (39986)
Redundancy	3.2 (3.2)	3.4 (3.3)	3.5 (3.5)	3.5 (3.5)	4.3 (3.9)
$I/\sigma I$	7.4 (2.0)	10.6 (2.1)	12.7 (2.0)	11.7 (2.1)	11.4 (1.9)
R_{work} (%)/R_{free} (%)	16.0/18.7	16.7/19.5	16.0/18.1	15.2/17.8	16.0/18.7
No. of atoms	17544	17880	18071	17857	18236
Wilson *B*-factor	22.7	21.4	21.3	23.4	22.2
Average B-factor (Å^2)	25.1	24.1	24.2	26.0	24.6
Protein	24.2	23.1	22.9	25.0	23.1
rmsd bond lengths (Å)	0.008	0.007	0.007	0.007	0.008
rmsd bond angles (°)	1.1	1.1	1.1	1.1	1.1
Ramachadran favoured (%)	97.2	97.7	97.7	97.5	97.5
PDB code	5A02	5A03	5A04	5A05	5A06

2.6. STRUCTURE DETERMINATION

Molecular replacement was used to solve the phase ambiguity of all the structures. The structures of *At* Gci and *At* KDG dehydratase without a ligand were solved using MOLREP[73] from the CCP4 software suite[74, 75] followed by a REFMAC5[76] rigid body refinement. Phaser[77] from the PHENIX software suite[78] was used for the molecular replacement of the complex structures of *At* KDG dehydratase and all the *Cc* AAOR structures. AutoBuild was used to improve the initial atomic models with iterative density modification, refinement, and model building[79].

The models for the atomic structures were completed manually with COOT[80], and the refinements were performed with phenix.refine[81]. The atomic coordinates and the structure factors of *At* KDG dehydratase and *Cc* AAOR structures have been deposited in the Protein Data Bank (PDB) in Europe (http://www.ebi.ac.uk/pdbe/)[82]. The summarized statistics for the data sets and the PDB codes are presented in Tables 1 and 2.

2.7. STRUCTURE COMPARISON AND ANALYSIS

The number of molecules in the asymmetric units was analyzed by calculating the Matthews coefficients[83, 84]. The oligomerization and the total interface areas were determined with the PISA server at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)[85] confirming our results from the gel filtration, dynamic light scattering measurements, and crystal structures. The enzymes were compared with the homologous structures in PyMOL[86], and the sequence alignments were done with MacVector[87]. The validation of the structures deposited to PBD was done with MolProbity[88] and the wwPDB Validation Server[89].

3. RESULTS AND DISCUSSION

3.1. GALACTAROLACTONE CYCLOISOMERASE

3.1.1. STRUCTURE DETERMINATION

The diffraction of the tetragonal *At* Gci crystals was measured with synchrotron radiation. The crystals belonged to space group I4, and the data was processed to 1.6 Å resolution. The structure of an enolase Spea_3858 from *S. pealeana* (PDB code: 3SJN) with a 32% sequence identity was used as a template in the molecular replacement[24]. The crystals were merohedrically twinned determined with phenix.xtriage, and therefore a twin operator -h,k,-l was included in the refinements. Matthews's coefficient V_M was determined being 2.5 $Å^3$/Da with solvent content V_S of approximately 51% corresponding to two molecules in the asymmetric unit. The sequence of *At* Gci monomer, excluding the *Strep*-tag, had 379 residues. The solved structure contained currently 700 amino acid residues and 827 water molecules. The root-mean-square deviation (rmsd) between the monomers was only 0.2 Å. The structure had gaps in the electron density maps of the flexible loops around the active site, in the loop between β1 and β2, the loop between β4 and α2, and the loop between β11 and β12. The residues in the gaps were left out of the structure.

3.1.2. OVERALL STRUCTURE

At Gci had the common enolase family two-domain structure that has an N-terminal α+β capping domain with three antiparallel β-strands (β1−3) packed against a bundle of four α-helices (α1-4) and a modified TIM-barrel that has eight parallel β-strands (β4−11) surrounded by seven α-helices (α5−α11). At the C-terminus, a long extension with two β-strands (β12−13) and an α-helix (α12) packed against the N-terminal α+β capping domain exists. The overall fold has been presented in the Figure 4A.

The oligomerization of *At* Gci was analyzed also with the interactive PISA tool suggesting that the enzyme formed a stable octamer (Fig. 4B and C). The octamer was formed by packing the α+β capping domains of two monomers together to form a dimer around a 2-fold rotation axis with an interface area of approximately 1686 Å^2, with the C-terminus of the monomer A extending all the way to the dimer-forming interface next to the C-terminus of the monomer B. This was followed by formation of an octamer around 4-fold rotation axis that formed three new monomer-monomer interfaces between the α+β capping domains of monomer A and monomer B from another dimer having an interface of ~1365 Å^2, and between the modified TIM-barrel domains of B monomers and A monomers from different dimers with interfaces ~504 Å^2.

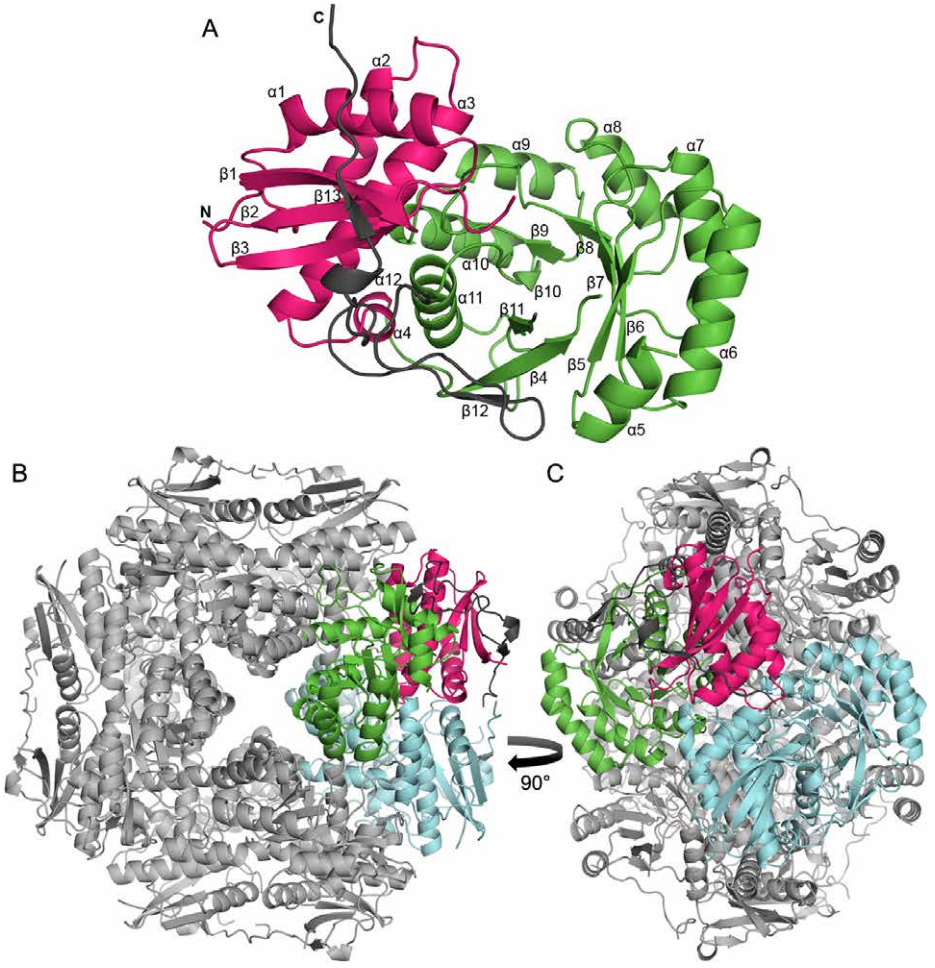


Figure 4. A) Overall structure of *At* Gci monomer having a modified TIM-barrel in *green*, α/β-capping domain in *pink*, and C-terminal extension in *gray*. **B)** The *At* Gci octamer. Monomer A colored as in Figure A, which forms the dimer with the monomer B in *blue*. **C)** The octameric structure turned 90° to the left.

The active site of *At* Gci was situated on the C-terminal side of the α/β-barrel (Fig. 5.) and was solvent exposed in every monomer at the octameric oligomerization state. It had the typical enolase family enzyme amino acid residues such as ligands for binding the Mg^{2+} cation at the ends of the strands β6 (Asp-195), β7 (Glu-221), and β8 (Glu-247). It also had the common mandelate racemase subgroup His-Asp dyad, of which, the His-297 was located at the end of the β10-strand and the Asp-270 at the end of the β9-strand. Usually in the mandelate racemase group enzymes, the histidine residue of the dyad acts as a general base and the general acid can be located at the end of β5, β6, or β8[28].

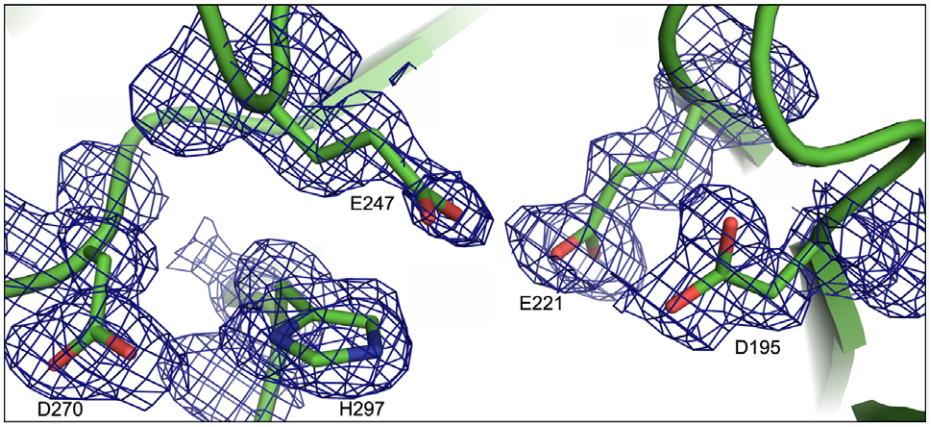


Figure 5. The active site of *At* Gci. The conserved active site residues have been shown as stick models with their $2F_o - F_c$ (in *blue*) electron density maps contoured at 1σ.

3.1.3. ATTEMPTED COMPLEX STRUCTURE DETERMINATION

The essential magnesium ion is not tightly bound into the enzyme. Therefore, as there was no magnesium in the enzymes' buffer or in the crystallization conditions, there was no density for the Mg^{2+} ion at the active site after the molecular replacement. Several magnesium-containing conditions were tested for crystallization, and adding magnesium in the original conditions and into the protein solution was also tried, but no optional crystallization conditions were found.

Complex formation was also studied in various conditions by soaking or co-crystallization with a substrate or a product and with several different analogues, but none of these conditions was successful. Without complex structures, no elucidations of the substrate binding or the reaction mechanism can be done, which is the reason the structure has not yet been published.

3.2. KETO-DEOXY-D-GALACTARATE DEHYDRATASE[I, II]

3.2.1. STRUCTURE DETERMINATION

The X-ray diffraction of the monoclinical crystals of *At* KDG dehydratase belonging to space group C2 was measured, and the three-dimensional structures of the enzyme without a ligand was solved to 1.7 Å, in complex with pyruvate to 1.5 Å, and in complex with 2-oxoadipic acid to 2.1 Å resolutions. The structure of *Bacillus clausii* dihydrodipicolinate synthase (PDB code: 3E96) that has only a 28% sequence identity to *At* KDG dehydratase was used as a search model in the molecular replacement[90]. The model was chosen according to the highest expectation value of a sequence search using BLAST[91] from the Research Collaboratory for Structural Bioinformatics PDB (www.rcsb.org)[92]. Calculated Matthews coefficient V_M 2.5 $Å^3$/Da gave us an estimated solvent content V_S of 51% corresponding to four molecules in the asymmetric unit. The sequence of *At* KDG dehydratase monomer without the *Strep*-tag consisted of 303 residues. The final solved structure of the enzyme without a ligand was composed of 1219 amino acid residues, 1216 water molecules, seven glycerol molecules, and three formate molecules. There was only an approximate 0.3 Å rmsd between the monomers.

3.2.2. OVERALL STRUCTURE

The main fold of *At* KDG dehydratase is a typical $(\alpha/\beta)_8$-barrel with eight parallel β-strands (β1−β8) surrounded by eight α-helices (α1−α8), and the active site was situated at the C-terminal side of the α/β-barrel. Furthermore, it had an N-terminal α-helix (αN) that packed against the N-terminal side of the α/β-barrel, and a C-terminal extension of three α-helices (αC1−3) that was involved in the formation of the quaternary structure. The overall fold has been presented in Figure 6A.

The oligomeric state of *At* KDG dehydratase was again confirmed with the interactive PISA server suggesting that the enzyme formed a tetramer with 222 symmetry (Fig. 6B) that was also shown in the crystal structures. The tetramer was formed packing first two monomers tightly together to form a dimer with an interface area of approximately 1580 $Å^2$ followed the by formation of a dimer of dimers with two new relatively flat monomer-monomer interfaces of ~850 Å.

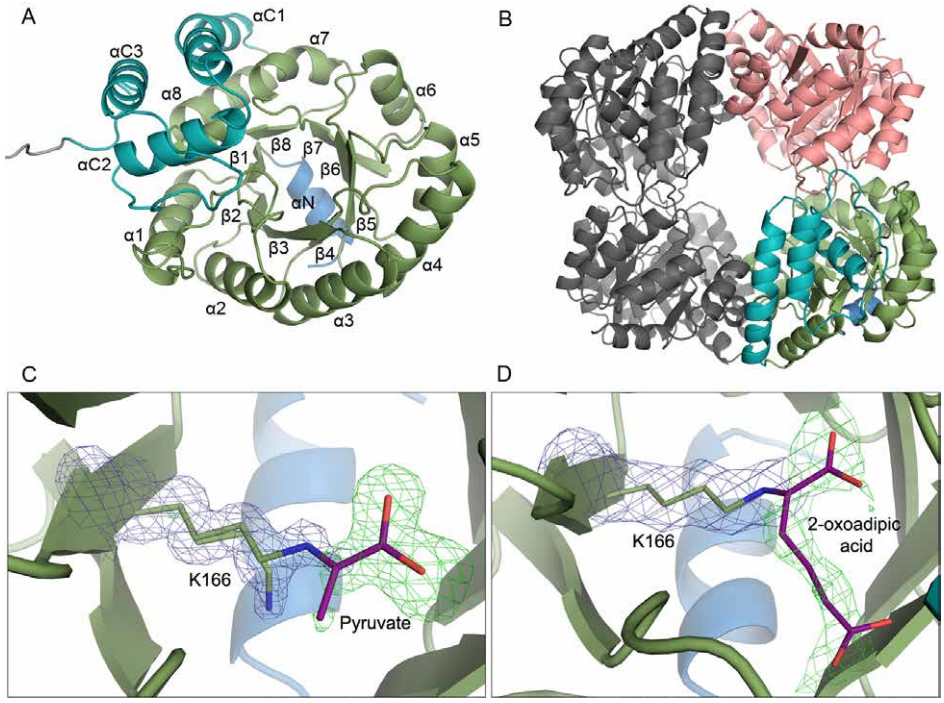


Figure 6. A) Overall structure of At KDG dehydratase monomer: α/β-barrel in *green*, N-terminal helix in *blue*, C-terminal helices in *cyan*, and Strep-tag in *gray*. B) Tetrameric structure of At KDG dehydratase. One of the monomers colored as in panel A, which forms the tight dimer with the monomer in *pink*. This dimer forms again a dimer of dimers with the dimer in *grey*. Active sites of the complex structures: C) Schiff-base formed between Lys-166 and pyruvate showing also the unbound conformation of Lys-166, and D) Schiff-base between Lys-166 and 2-oxoadipic acid. The 2Fo − Fc (in *blue*) electron density maps around the Lys-166 contoured at 1σ and the Fo − Fc omit maps (in *green*) around the ligands are contoured at 3σ.

3.2.3. LIGAND BINDING

To study the enzyme-substrate interactions, complex structures of *At* KDG dehydratase were determined from crystals soaked into a cryoprotectant containing 10 mM pyruvate or 50 mM 2-oxoadipic acid in addition. Clear electron densities for the ligands forming a Schiff-base intermediate with the conserved Lys-166 at the active site were detected after molecular replacement (Fig. 6C and D). In the structure without a ligand, the Lys-166 was able to form a hydrogen bond from its ε-amino group to the Tyr-140 (Fig. 7A).

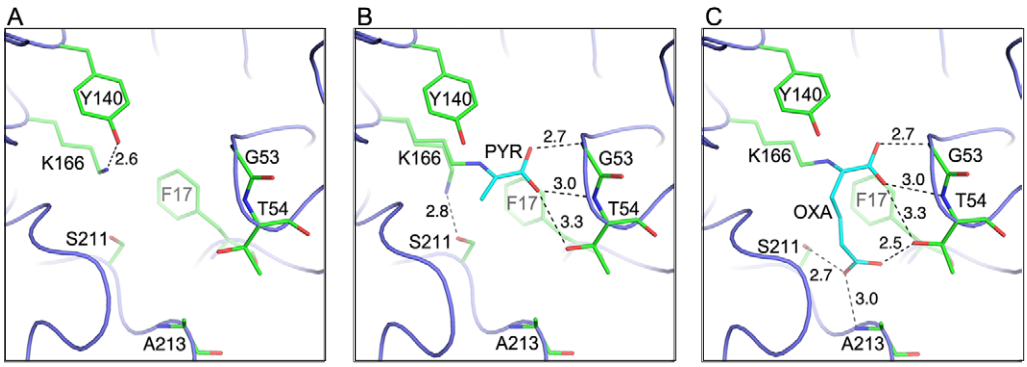


Figure 7. Comparison of the active sites of *At* KDG dehydratase structures. **A)** Without ligand, **B)** pyruvate (PYR) complex, and **C)** 2-oxoadipic acid (OXA) complex. Hydrogen bonds are shown as dashed lines with the bond lengths in Å.

In the complex structures, the carboxylate groups of the keto acid ends of the ligands formed hydrogen bonds with the main-chain nitrogens of Gly-53 and Thr-54 and with the side-chain hydroxyl group of Thr-54 (Figures 7B and C). Because the occupancy of the pyruvate was only partial at the active site, the Schiff-base forming Lys-166 was able to adapt a second conformation in which it formed a hydrogen bond to the Ser-211 side-chain. The aliphatic chain of 2-oxoadipic acid was found to exist in several conformations. In the monomer C, it was bound in a crevice formed by the loops between β7 and α7 and between β8 and α8, and in a conformation in which the aliphatic chain had the most extensive electron density. In this conformation (Fig. 7C), the second carboxylate group of the ligand made hydrogen bonds with the main-chain amino group of Ala-213 and the side-chains of Thr-54 and Ser-211.

3.2.4. REACTION END-PRODUCT ANALYSIS

In the reaction catalyzed by *At* KDG dehydratase, the KDG substrate is decarboxylated and dehydrated into an α-ketoglutaric semialdehyde. The enzymatic reactions of *At* KDG dehydratase and the S211A mutant were analyzed by negative-ion ESI-QIT mass spectrometry. Reaction mixtures of the substrate, KDG (192 Da), with both the wild-type enzyme and the site-directed mutant, were prepared and left to stand at room temperature for 30 min. The mass spectrum of the pure substrate and the reaction mixtures were measured. The spectra measured are shown in Figure 8.

In the spectrum of the pure substrate, only a single intense peak at *m/z* 191, which corresponded to its singly charged $[M-H]^-$ ion, was observed. In the spectra of the reaction mixtures, a peak at *m/z* 129 appeared as the most abundant signal corresponding to the expected singly charged ion of the reaction product, α-ketoglutaric semialdehyde (130 Da). The intensity of the product ion was not as high

for the enzymatic reaction of the S211A mutant as it was for the enzymatic reaction of the wild-type enzyme and this observation supports the loss of activity with the mutation of this conserved amino acid residue. The 30-fold lower activity of the S211A mutant was shown also with biochemical activity measurements (data not shown).

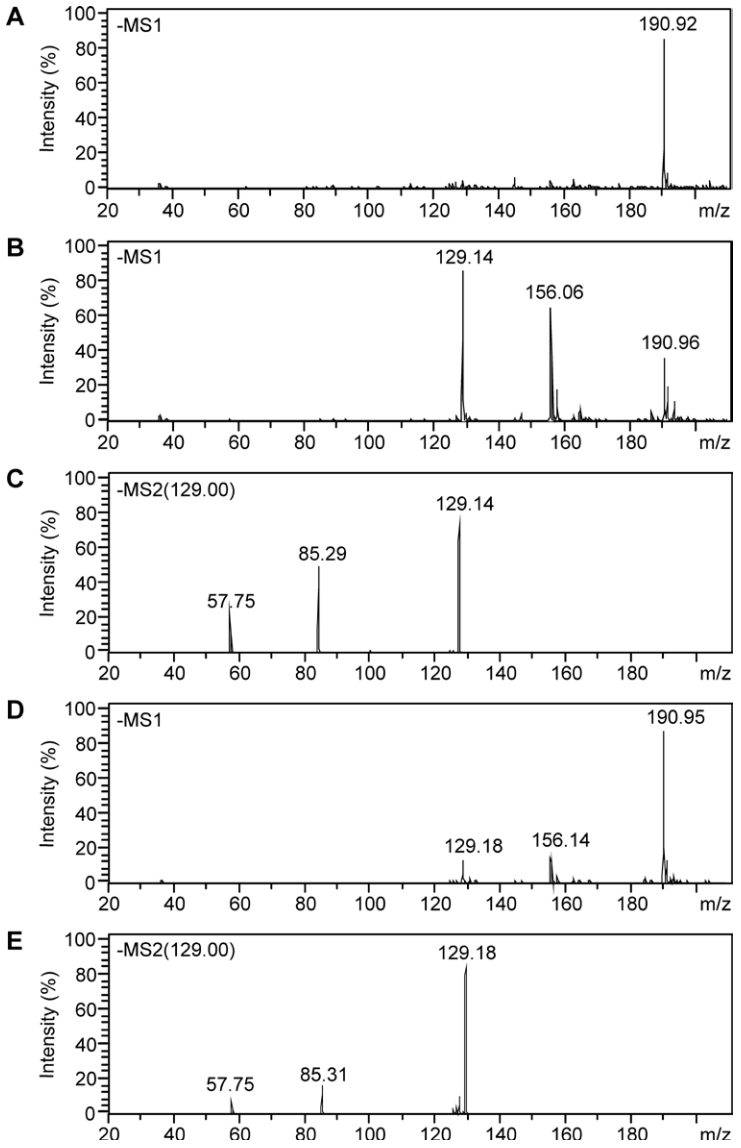


Figure 8. A) ESI-MS spectrum of KDG. **B)** ESI-MS spectrum of KDG after 30 min reaction with *At* KDG dehydratase. **C)** ESI-MS/MS spectrum of the *At* KDG dehydratase reaction product (*m*/*z* 129). **D)** ESI-MS spectrum of KDG after 30 min reaction with *At* KDG dehydratase Ser211Ala mutant. **E)** ESI-MS/MS spectrum of the *At* KDG dehydratase Ser211Ala mutant reaction product (*m*/*z* 129).

MS/MS experiments were used to verify the *m/z* 129 product ions' structure. CID activation gave fragments at *m/z* 85 that corresponded to the neutral loss of CO_2 (44 Da), and *m/z* 57 corresponded to a following neutral loss of CO (28 Da). The additional peak observed at *m/z* 156 corresponded to the deprotonated buffer salt, Tris-HCl, used as a solvent of the enzyme. These results are consistent with the enzymatic formation of α-ketoglutaric semialdehyde, and no other reaction products were observed.

3.2.5. REACTION MECHANISM

Even though the members in the Class I aldolase family catalyze various reactions, they all have common Schiff-base intermediate forming reaction mechanisms[32, 93]. *At* KDG dehydratase catalyzes the dehydration followed by decarboxylation of 3-deoxy-2-keto-L-*threo*-hexarate to α-ketoglutaric semialdehyde, which would also start with the formation of the Schiff-base, a covalent bond between the neutral side-chain of Lys-166, and the KDG substrate (Fig. 9). This happens with an imine formation (step 1 in Fig. 9) when the amine nitrogen of Lys-166 attacks the electrophilic C2 carbonyl of KDG. Isomerization to an enamine (step 2) form happens when a proton from the β-carbon of KDG is transferred by the general acid/base (B in Fig. 9) that is presumably Tyr-140 in *At* KDG dehydratase (Fig. 10A). This Lys-Tyr catalytic dyad is a characteristic feature for the enzymes in the Class I aldolase protein family[32].

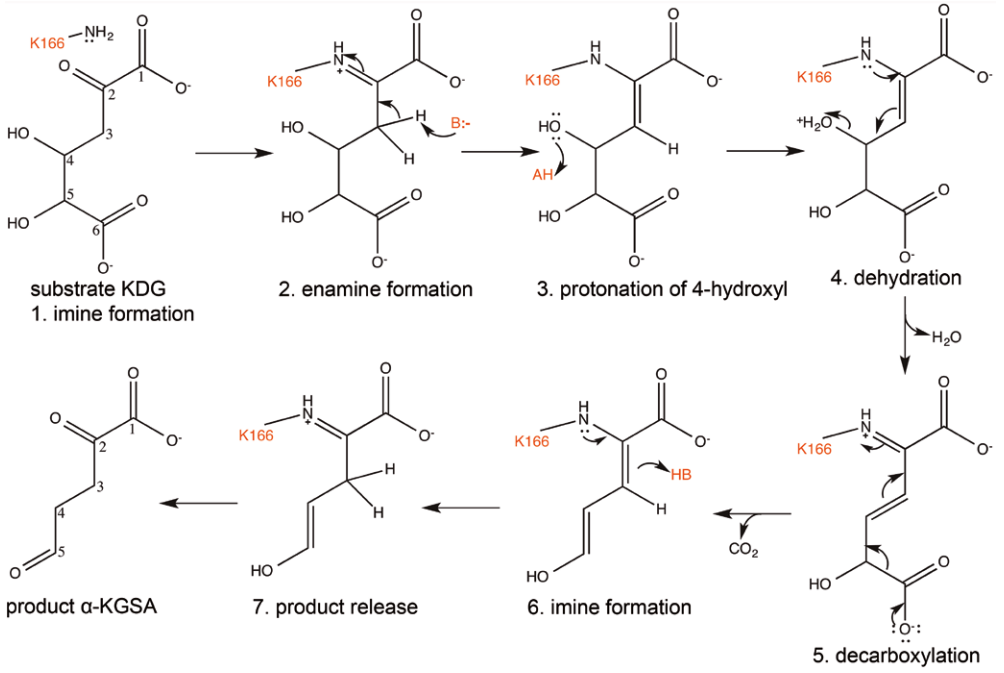


Figure 9. Proposed *At* KDG dehydratase reaction mechanism.

The formation of enamine at step 2 is a unique feature for *At* KDG dehydratase when comparing the reaction mechanism with the one of *Sulfolobus solfataricus* 2-keto-3-deoxygluconate aldolase (*Ss* KDGA). *Ss* KDGA uses the same substrate as *At* KDG dehydratase, but catalyzes the cleavage of a carbon-carbon bond by retro-aldol reaction[31]. This is due to the fact that the conformation of the bound substrate is completely different in *Ss* KDGA caused by the differences at the active site. The conformation of the substrate in *Ss* KDGA results in deprotonation from 4-hydroxyl, instead of 3-hydroxyl as in *At* KDG dehydratase, and this leads to retro-aldol reaction without enamine formation. Superimposition of the active sites of *At* KDG dehydratase complexed with 2-oxoadipic acid and *Ss* KDGA complexed with 3-deoxy-D-lyxohexonic acid (PDB code: 1W3T) is presented in Figure 10B showing the differences in the amino acid residues in the active site.

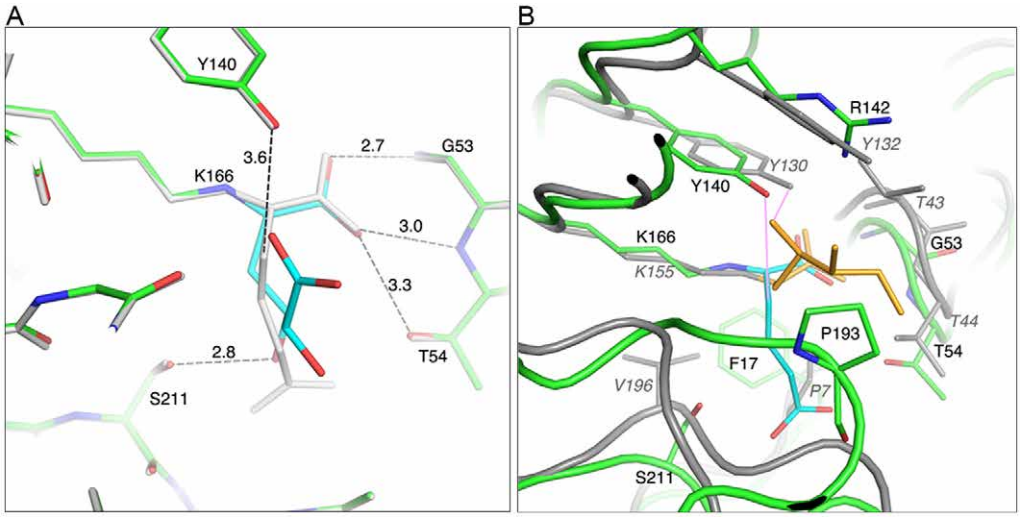


Figure 10. A) Docked enamine intermediate (in *cyan*) of the substrate in the active site of *At* KDG dehydratase (in *green*). 2-Oxoadipic acid complex structure in *gray*. The distance between the hydroxyl oxygen of Tyr-140 and β-carbon of 2-oxoadipic acid is shown with dashed line and bond length in Å. **B)** Superimposition of *At* KDG dehydratase (in *green*) complexed with 2-oxoadipic acid (in *cyan*) and *Ss* KDGA (in *gray*) complexed with 3-deoxy-D-lyxohexonic acid (in *yellow*). The purple lines show the proposed proton abstraction by a general base.

In the reaction mechanism of *At* KDG dehydratase, the enamine formation is followed by the protonation of 4-hydroxyl (step 3) that is presumably via Ser-211 (A in Fig. 9), which would cause the loss of activity in the S211A mutant. The 4-hydroxyl is then eliminated by E1-type of dehydration (step 4) and is followed by another E1-type elimination (step 5) for the decarboxylation reaction[20]. The product is then released to its imine form by breaking the bond to Lys-166 (step 6), which is finally interconverted to the enol form with common keto-enol tautomerism (step 7).

3.3. ALDOSE-ALDOSE OXIDOREDUCTASE[III]

3.3.1. STRUCTURE DETERMINATION

The *Cc* AAOR crystals belonged to the monoclinic space group $P2_1$. The measured X-ray diffraction data sets for the glycerol, D-xylose, D-glucose, maltotriose, and D-sorbitol complex structures were determined to 2.0, 1.8, 1.7, 1.9, and 1.8 Å resolutions, respectively. The structures were solved by molecular replacement using a R30K/R31K double mutated version of the precursor form of glucose-fructose oxidoreductase from *Zymomonas mobilis* (*Zm* GFOR, PDB code: 1H6A)[94] with a 49% sequence identity to *Cc* AAOR as a template. In molecular replacement, six monomers in the asymmetric unit were found, which corresponds to a calculated Matthews coefficient, V_M, of 3.6 $Å^3$/Da with a solvent content, V_S, of 65 percent. The *Cc* AAOR monomer without its signal sequence had 339 amino acid residues. The final *Cc* AAOR glycerol complex structure asymmetric unit was composed of 2010 amino acid residues, six NADPHs, 1550 water molecules, 16 glycerol molecules, and eight sulfate ions. The average rmsd was 0.2 Å when superimposing the six monomers in the asymmetric units from all the structures.

3.3.2. OVERALL STRUCTURE

Cc AAOR had the typical two-domain structure of an enzyme that belongs to the Gfo/Idh/MocA family. It had eleven α-helices and fourteen β-strands, from which the N-terminal Rossmann fold was comprised of seven parallel β-strands (β1-β7) forming a sheet surrounded by five α-helices (α1 and α2 on one side and α3-α5 on the other side). The N-terminal nucleotide-binding domain was tightly packed against the C-terminal α/β domain involved in substrate binding and dimerization. It contained an open-faced β-sheet formed of seven, mainly anti-parallel, β-strands (β8-β14) and six α-helices (α6-α11) on the one side of the β-sheet. The helices α8 and α10 that were located in the middle of long loops were especially important for packing the two domains together. The overall fold is presented in Figure 11A.

Using the PISA server, the oligomerization state of the crystal structure of *Cc* AAOR was confirmed to be a dimer with a two-fold rotation axis between the monomers. Packing the open-faced β-sheets of two monomers together, a tight dimer with a total interface area of approximately 1470 $Å^2$ was formed by creating an α/β/β/α sandwich type of a structure to the dimer interface (Figure 11B). Furthermore, *Cc* AAOR was found to be a dimer in solution defined by dynamic light scattering (data not shown).

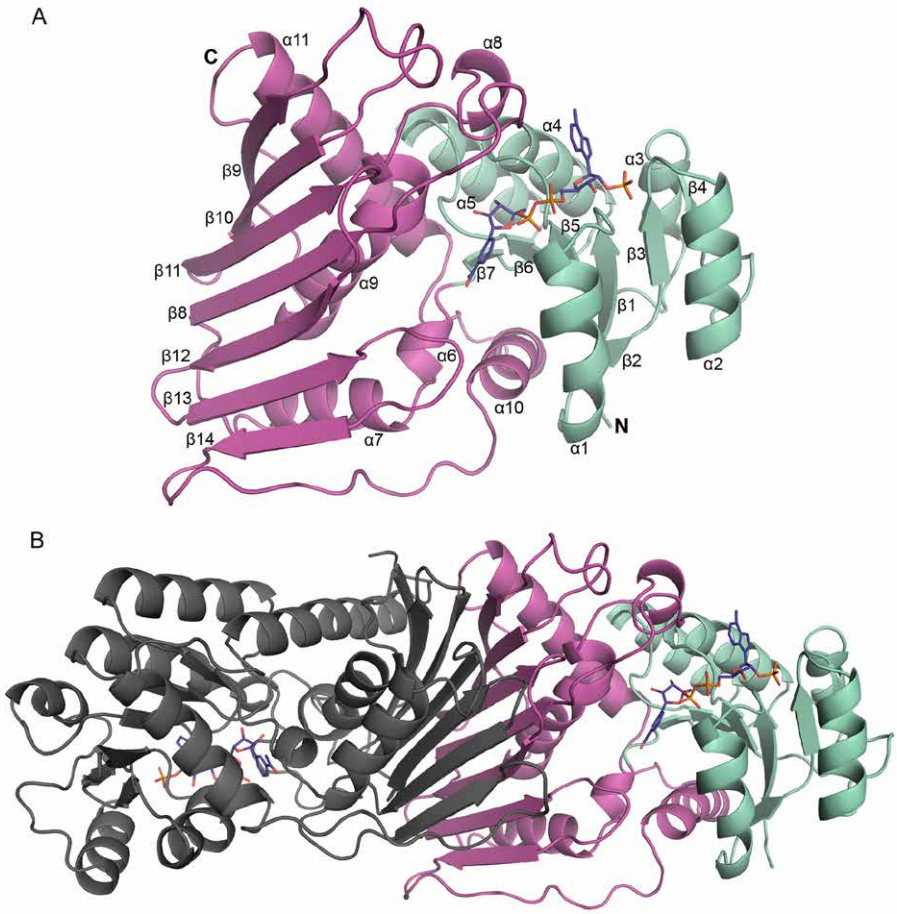


Figure 11. A) The overall structure of the *Cc* AAOR monomer. The N-terminal nucleotide-binding domain is shown in *cyan* and the C-terminal domain in *magenta*. **B)** *Cc* AAOR dimer formed by stacking the C-terminal β-sheet of monomer A (colored as in panel A) against the β-sheet of monomer B (in *dark grey*). NADP(H) cofactors are shown in *violet*.

3.3.3. NUCLEOTIDE BINDING

After molecular replacement, the electron density maps showed an evident positive difference density for the NADP(H). The cofactor was bound on the C-terminal side of the β-sheet on the Rossmann fold at the cleft between the loops from β1 and β5 in which there was a change in the direction of the β-strand order. The *syn*-conformation of the adenine moiety of NADP(H) was unusual, but it was in parallel with the bound cofactor in *Zm* GFOR[52] The cofactor formed an extensive network of interactions to the enzyme (Fig. 12) making it tightly bound even though it is relatively highly exposed to solvent.

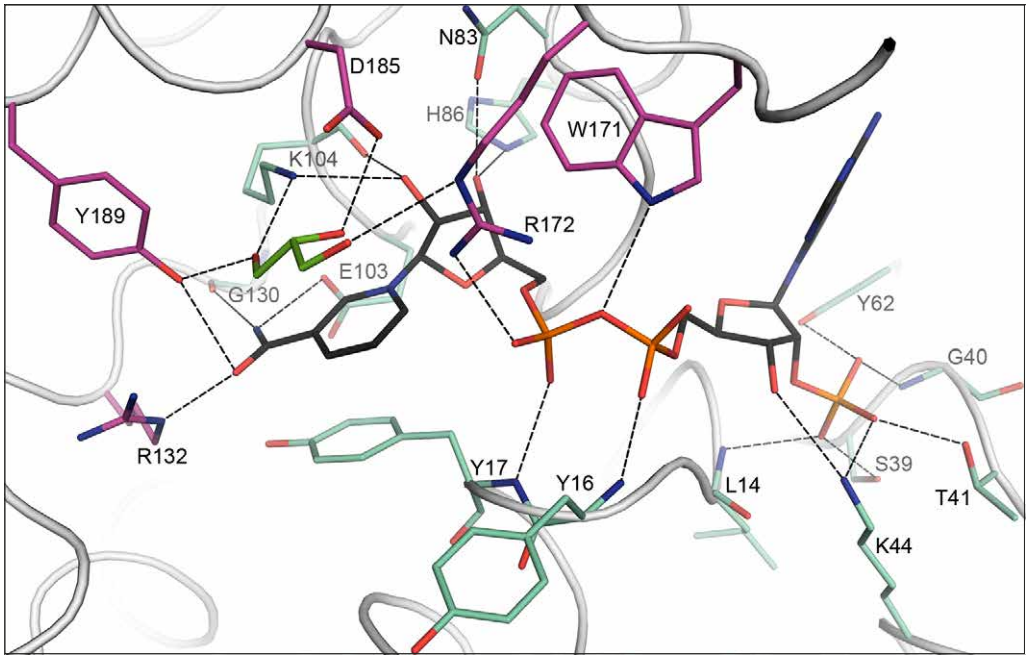


Figure 12. The active site of the *Cc* AAOR glycerol complex structure presenting the bonding of $NADP^+$ (in *black*) and glycerol (in *green*) to the active site residues (colored as in Fig. 11A).

Based on the UV/vis measurements of the crystals, the oxidized $NADP^+$ was placed in the densities at the active sites of the glycerol complex structure and the reduced NADPH in the active sites of the D-xylose, D-glucose, maltotriose, and D-sorbitol complex structures. This indicated that an electron transfer reaction had already occurred in the NADPH-containing complex structures. Because the UV/vis spectrum of the sorbitol soaked crystal was measured only after X-ray radiation, it could not be identified if the reduction of the $NADP^+$ had happened due to electron transfer reaction between the substrate and the enzyme or due to radiation damage. No conformational changes were observed at the active site between the structures with either the oxidized or reduced cofactor. The UV/vis spectra measured are shown in Figure 13.

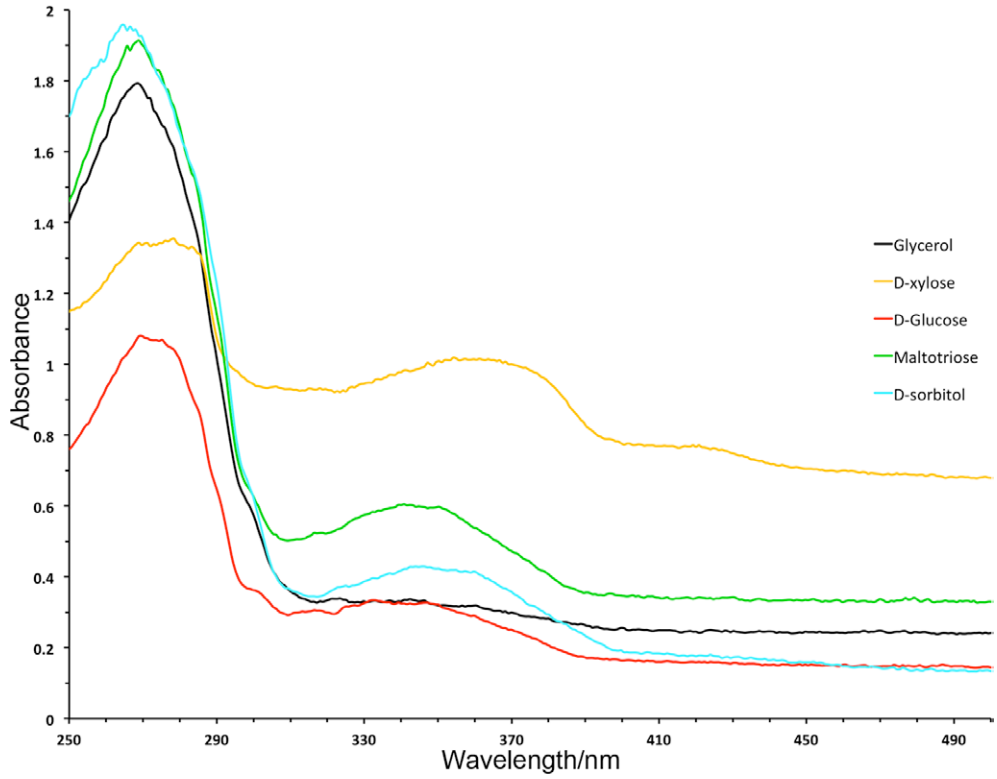


Figure 13. Measured UV/vis spectra of the oxidization state of the NADP(H) cofactor in the *Cc* AAOR crystals.

3.3.4. LIGAND BINDING

Five complex structures were solved to study the binding of different ligands in *Cc* AAOR. Clear electron densities for the two monosaccharides in pyranose forms (D-glucose and D-xylose), for the trisaccharide (maltotriose) composed of α-1,4-linked D-glucose units in pyranose form, and for the acyclic sugar alcohols (glycerol and D-sorbitol) were observed next to the nicotinamide ring of the cofactor, which is clearly a prerequisite for the substrate binding. The binding of glycerol is shown in Figure 12.

From the superimposition of the D-glucose, D-xylose, and maltotriose complex structures (Fig. 14A), we can see their similar binding where the O1 hydroxyls made hydrogen bonds with Lys-104 and Tyr-189, and the O2 and O3 hydroxyls made hydrogen bonds with Arg-172 and Asp-185. Additionally, the pyranose rings were in close contact with the phenyl ring of Tyr-267. The O6 of D-glucose was able to make a hydrogen bond with Arg-132. Furthermore, the middle glucose residue of maltotriose formed additional contacts with the enzyme: a hydrogen bond from the O6 to Asn-248, and a hydrophobic interaction by stacking clearly against the Phe-163 ring.

According to the complex structures and characterization studies[42], the configurations of the hydroxyl groups in the substrate play a big role in the substrate specificity. Substrates having a free equatorial O1 (β) hydroxyl group seemed to be the only accepted substrate for the oxidation reaction, and equatorial hydroxyl groups at the C2 and C3 were also substantially preferred. The best substrate for *Cc* AAOR was shown to be L-arabinose with the amount of β-L-arabinose decreasing faster than α-L-arabinose, which is in agreement that β-anomer would be the reactive substrate.

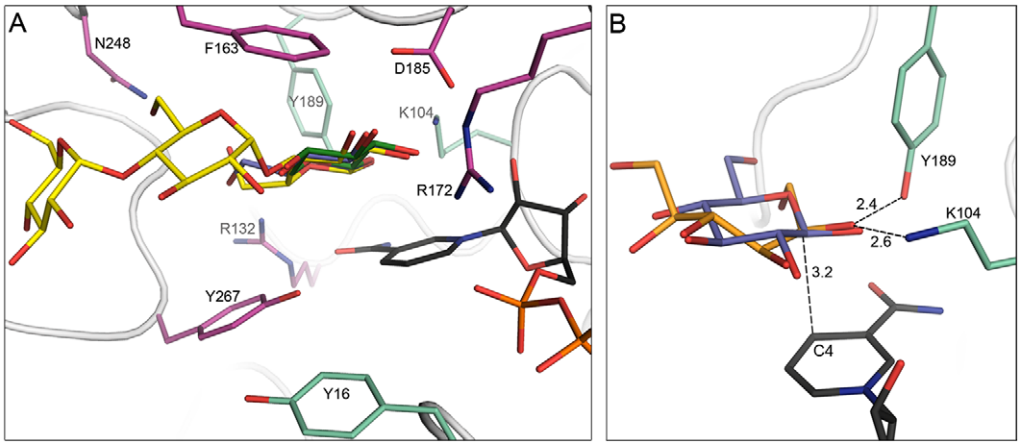


Figure 14. A) The binding of the pyranose ring substrates in the active site of *Cc* AAOR. D-xylose is shown in *green*, D-glucose in *violet*, maltotriose in *yellow*, cofactor in *black*, and active site substrate binding residues are colored as in Figure 9A. **B)** The active site of *Cc* AAOR in D-sorbitol (in *orange*) complex superimposed with D-glucose (in *violet)*. The key distances (taken from the glucose complex structure A monomer) between the ligands, the catalytic residues (in *cyan*), and the cofactor are shown (in Å).

The acyclic glycerol and D-sorbitol could bind in different conformations in the active site. For glycerol, only one predominant conformation occurred in which the three oxygen atoms of the glycerol's hydroxyl groups were located in the same positions as the O1, O2, and O3 of the pyranose ring containing substrates with the similar hydrogen bonding. On the other hand, the binding of D-sorbitol in the *Cc* AAOR active site happened in several different orientations. The most extensive electron density for D-sorbitol was in the active site of monomer A, which had the conformation presented in the Figure 14B superimposed with D-glucose. In this conformation, the C5 atom occupied approximately the same position as the C1 atom of the D-glucose ring.

In order for the hydride ion (H^-) transfer to happen between the cofactor and a substrate, the C1 carbon atom from the substrate needs to be oriented towards the reactive C4 carbon of the nicotinamide ring of the cofactor. In all the ligands, the distance between the carbon oriented towards the cofactor (C1 in the D-xylose, D-

glucose, and maltotriose complex structures, C3 in the glycerol complex structure, and C3 or C5 in the D-sorbitol complex structure) and the C4 carbon of the nicotinamide ring was 3.1-3.5 Å allowing the hydride ion to be transferred from the substrate to the $NADP^+$ thus reducing the cofactor and leading consequently to the substrate oxidation.

3.3.5. COMPARISON WITH *ZM* GFOR

Zm GFOR is a well-characterized enzyme catalyzing the NADP(H) dependent oxidation of D-glucose to glucono-1,5-lactone and reduction of D-fructose to sorbitol[95, 96]. *Zm* GFOR is the only significantly homologous sequence to *Cc* AAOR with previously published structures[52, 94, 97-99]. Despite the considerable sequence identity of 49%, *Cc* AAOR had notable differences in its three-dimensional structure and function. The average rmsd for the structures was 0.9 Å.

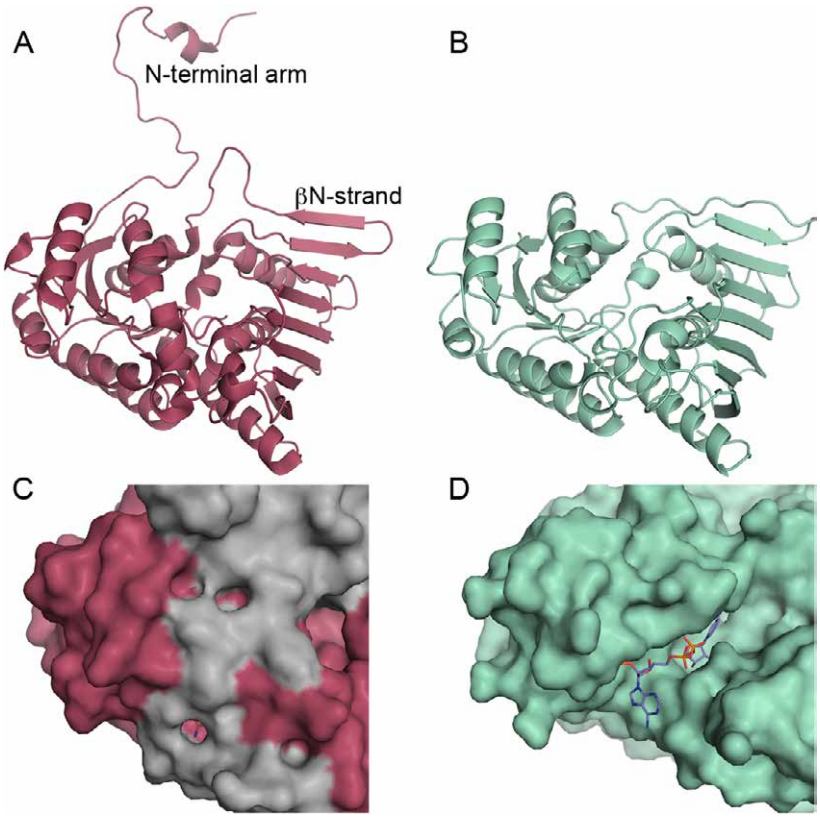


Figure 15. A) *Zm* GFOR monomer. **B)** *Cc* AAOR monomer. Surface representations of the active sites: **C)** *Zm* GFOR, where the N-terminal arm (in *grey*) of an adjacent molecule covers the active site and the cofactor of the *red* monomer, **D)** *Cc* AAOR, where we see the solvent-exposed cofactor (in *violet*).

The major difference in the structures was the 25 residues longer N-terminal of *Zm* GFOR, namely the N-terminal arm, which was the main reason for the tetrameric oligomerization state of *Zm* GFOR. This N-terminal extension wraps around an adjacent molecule covering up its active site while burying the cofactor entirely. It has been shown that the truncation of the N-terminal arm of *Zm* GFOR led its oligomerization to change to dimer and also to weaken the cofactor binding, thus leading to loss of catalytic activity[97]. Interestingly enough, even though *Cc* AAOR had the oligomeric state of a dimer and an surface exposed cofactor, there was no NADP(H) dissociation observed, and the enzyme retained its activity without addition of external cofactor. Furthermore, Phe-317 from the loop following the additional βN-strand extends towards the substrate-binding site. The structural differences have been compared in Figures 15 and 16.

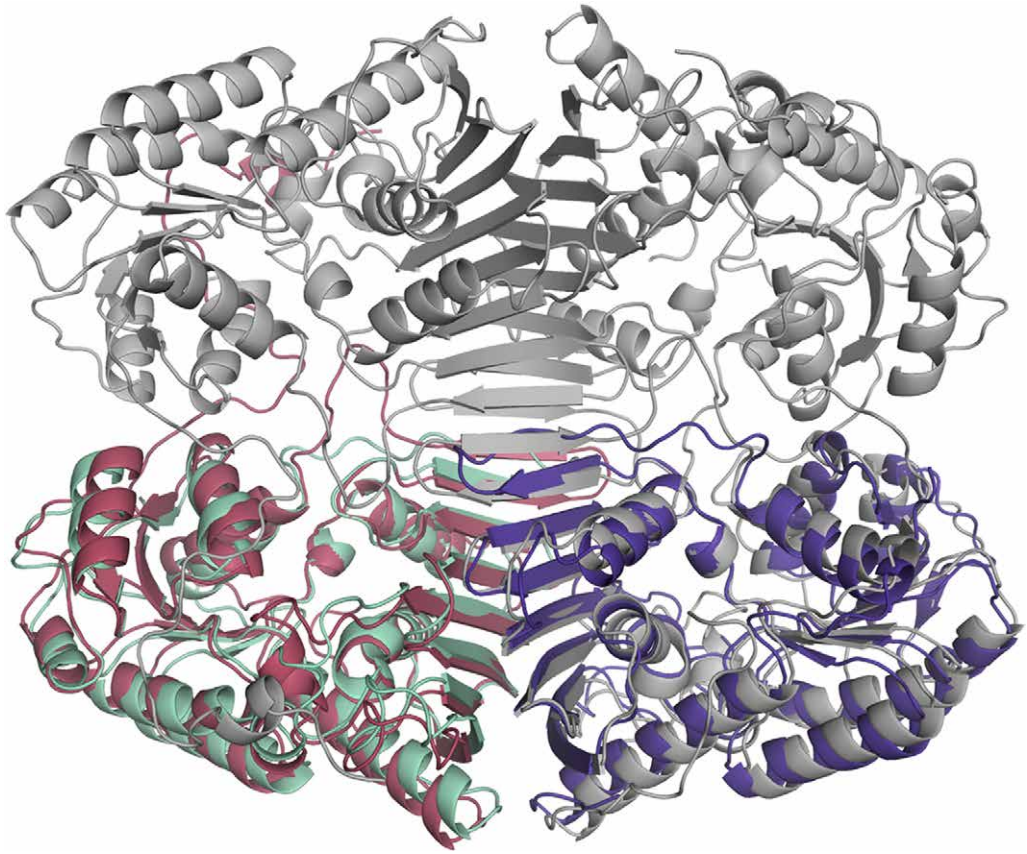


Figure 16. The superimposition of the oligomeric structures of *Cc* AAOR (dimer in *cyan* and *blue*) and *Zm* GFOR (tetramer in *red* and *grey*).

3.3.6. REACTION MECHANISM

Based on our results, we presented a general scheme for the oxidation and reduction reactions of D-glucose catalyzed by *Cc* AAOR (Fig. 17).

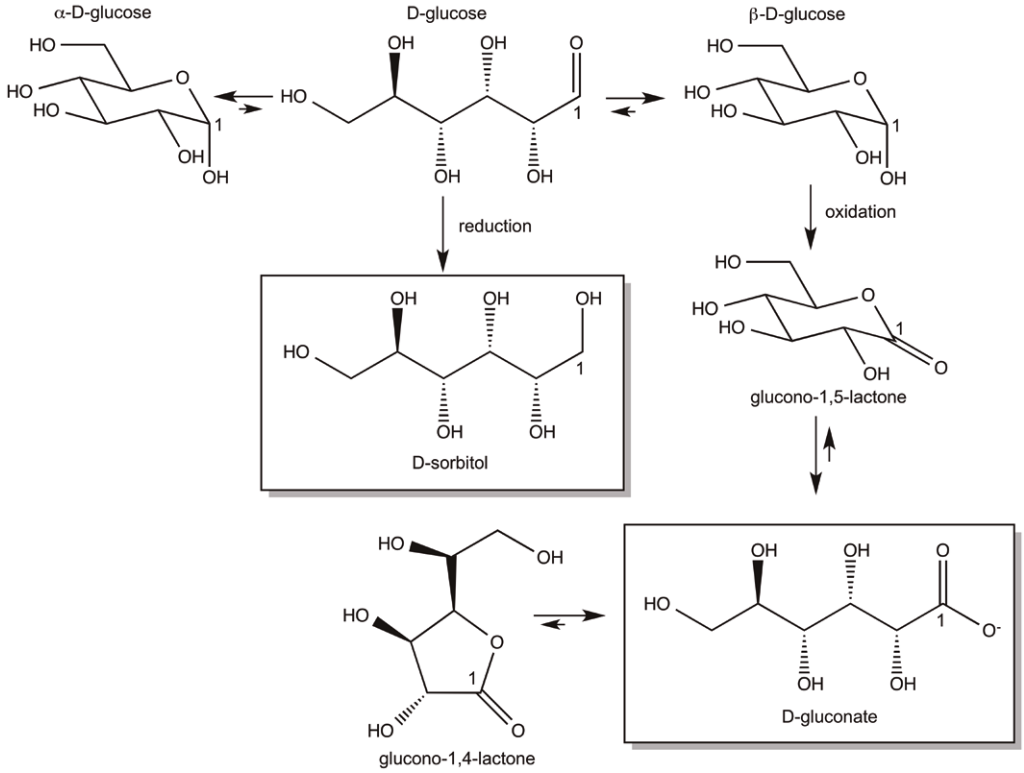


Figure 17. The overall reaction scheme of *Cc* AAOR for D-glucose in which the β-anomer is oxidized with $NADP^+$, and the small amount of free aldehyde form existing in solution can be reduced by NADPH. Number 1 indicates the carbon C1 in the molecules.

The structures of *Cc* AAOR with relevant ligands gave us novel information about the roles of the active site amino acid residues in the enzymatic catalysis. NADP(H) played a key role in the hydride ion transfer to the C1 position of the substrate. The structures unambiguously showed very short distances from the O1 of the ligand to the phenolic oxygen Oη of Tyr-189 and to the Nζ atom of Lys-104 (2.4-2.6 Å). This means both could, in theory, accept or donate a proton that is needed in the reaction after the hydride ion transfer. The mutual distance between Oη atom of Tyr-189 and Nζ atom of Lys-104 is clearly longer, 4.2 Å, indicating that they do not form a catalytic dyad even though suggested by previous studies[94].

The roles of the catalytic amino acid residues have been widely studied, for example, by site-directed mutations, but they are not unambiguously clear[55]. Nevertheless, the crystal structures suggest that both amino acid residues, Lys-104 and Tyr-189, are essential residues for *Cc* AAOR to catalyze the reduction of a linear form of a monosaccharide substrate and the oxidation of a pyranose form of the same substrate in a reaction cycle, during which the bound cofactor is regenerated, but their mutual role in the reaction mechanism was difficult to dissect.

3.4. GFO/IDH/MOCA PROTEIN FAMILY[IV]

The proteins belonging to the Gfo/Idh/MocA family were found according to extensive literature, structure, and sequence searches within the worldwide PDB[100] using the search engines at the PDB in Europe[82] and the Research Collaboratory for Structural Bioinformatics PDB (www.rcsb.org)[92] as well as within the DALI-server[23]. The searches were performed using the sequences of *Cc* AAOR and *Zm* GFOR and the protein family name as search items. Several structures were found, but only eleven different proteins were chosen for the final comparison excluding similar enzymes from different sources and structures without a defined function (no publication available). The used crystal structures have been listed in Table 3 with their PDB codes, EC numbers, number of amino acid residues (AA), quaternary structures, utilized cofactors, and calculated rmsds. The average rmsd values have been calculated by superimposing the secondary structures to *Cc* AAOR.

In addition to *Cc* AAOR and *Zm* GFOR, the proteins in the comparison were: 1) glucose-6-phosphate dehydrogenase (G6PD) from *Leuconostoc mesenteroides*[51], 2) biliverdin reductase (BVR) from *Rattus norwegicus*[57, 101], 3) 1,5-anhydro-D-fructose reductase (AFR) from *Sinorhizobium meliloti*[102], 4) α-N-acetylgalactosidase (A-zyme) from *Elizabethkingia meningosepticum*[58], 5) transcriptional inhibitor Gal80p from *Kluyveromyces lactis*[60], 6) dihydrodiol dehydrogenase (DD) from *Macaca fascicularis*[59, 103], 7) uridine-5'-phosphate-2,3-diacetamido-2,3-dideoxy-D-mannuronic acid dehydrogenase (WlbA) from *Thermus thermophilus*[54], 8) *myo*-inositol dehydrogenase (IDH) from *Bacillus subtilis*[56, 104], and 9) C-3"-ketoreductase (KijD10) from *Actinomadura kijaniata*[55]. The sequence identities were generally low, under 20%, but the overall three-dimensional structures were strikingly similar. The topological diagram of *Cc* AAOR, with the key residues at the active site, has been presented in Figure 18, which was similar to all the other proteins.

Table 3. Crystal structures used in the comparison

Protein	PDB code	EC number	AA count	Quaternary structure	Cofactor	RMSD Å
Cc AAOR	5A02, 5A03, 5A04, 5A05, 5A06	1.1.99.-	339	dimer1	NADP	0
AFR	4KOA	1.1.1.263, 1.1.1.292	332	dimer1	NADP	1.9
A-zyme	2IXA, 2IXB	3.2.1.49	444	dimer1	NAD	2.6
BVR	1GCU, 1LC0, 1LC3	1.3.1.24	295	monomer	NAD	2.4
DD	2O48, 2O4U, 2POQ, 3OHS	1.3.1.20	334	dimer1	NADP	2.5
Gal80p	2NVW	-	479	dimer1	-	2.4
Zm GFOR	1OFG, 1EVJ, 1H6A, 1H6B, 1H6C, 1H6D, 1RYD, 1RYE	1.1.99.28	381	tetramer1	NADP	0.9
G6PD	1DPG	1.1.1.49	485	dimer2	NADP	3.1
KijD10	3RBV, 3RC1, 3RC2, 3RC7, 3RC9, 3RCB	1.1.1.-	350	tetramer1	NADP	2.4
myo-IDH	3MZ0, 3NT2, 3NT4, 3NT5, 3NTO, 3NTQ, 3NTR, 4L8V, 4L9R	1.1.1.18, 1.1.369	344	tetramer1	NAD/NADP	2.3
WlbA	3OA0, 3O9Z	1.1.1.374 (?)	318	tetramer2	NAD	2.2

All the structures had a two-domain structure containing the N-terminal nucleotide-binding domain, and the C-terminal substrate-binding α/β-domain that was frequently also involved in the oligomerization. The C-terminal domain consisted of a two-layer α/β-sandwich in which the β-sheet was predominantly antiparallel having six to nine β-strands. Both faces of the β-sheet were usually hydrophobic with one face participating in the homodimer formation as described before for *Cc* AAOR and the other face packing against the α-helices between the domains. These helices were in the N- and C-terminal regions of the α/β-domain and in the polypeptide chain after the first β-strand of the β-sheet. The monomers of the proteins compared in the study have been presented in Figure 19.

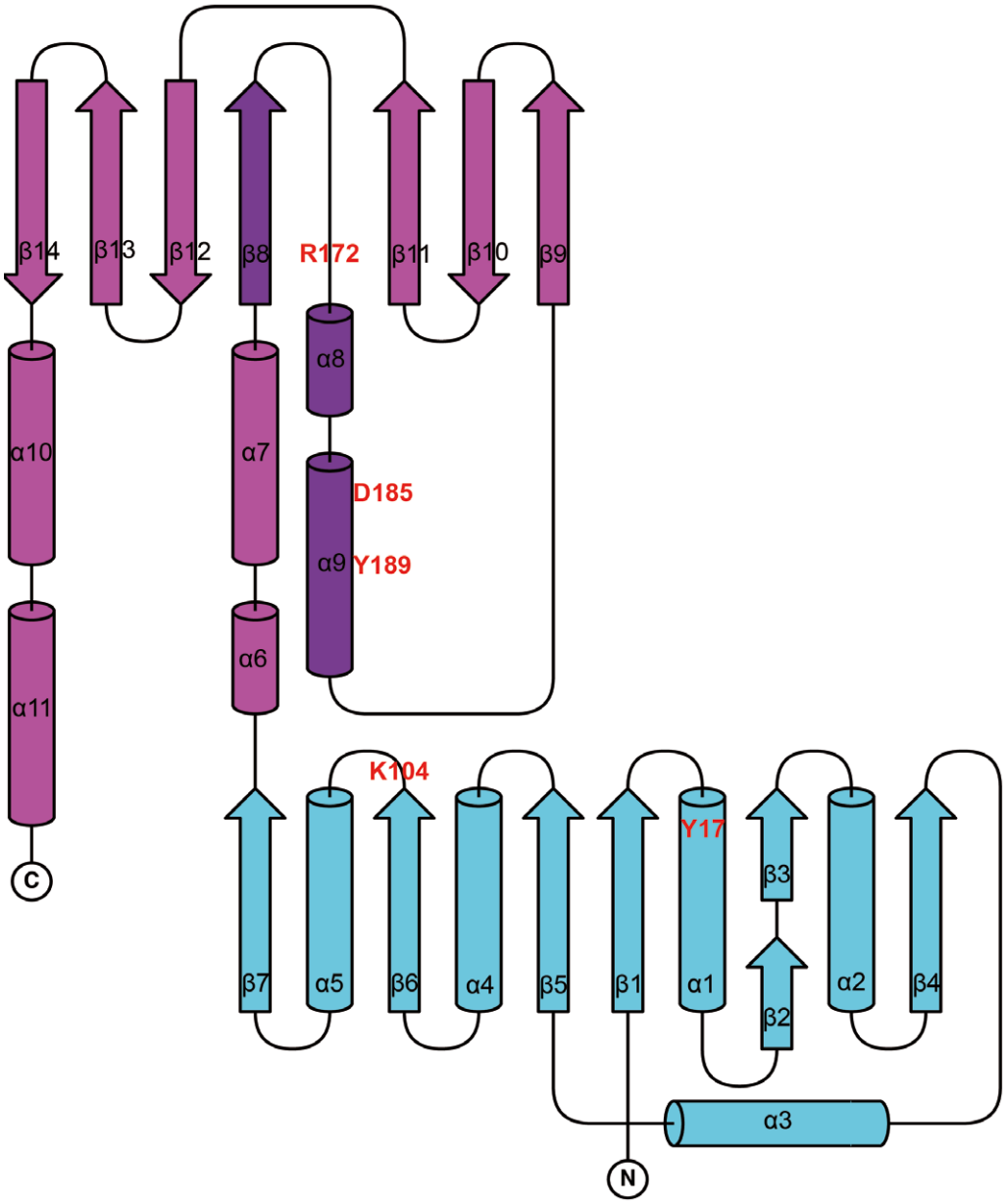


Figure 18. Topology diagram of the *Cc* AAOR monomer. The cofactor-binding domain is presented in *cyan* and the C-terminal α/β-domain in *purple*. The central βα-motif has been highlighted with darker color, and the key amino acid residues at the active site have been labeled in *red*. The figure was produced using TopDraw[105] from the CCP4 suite.

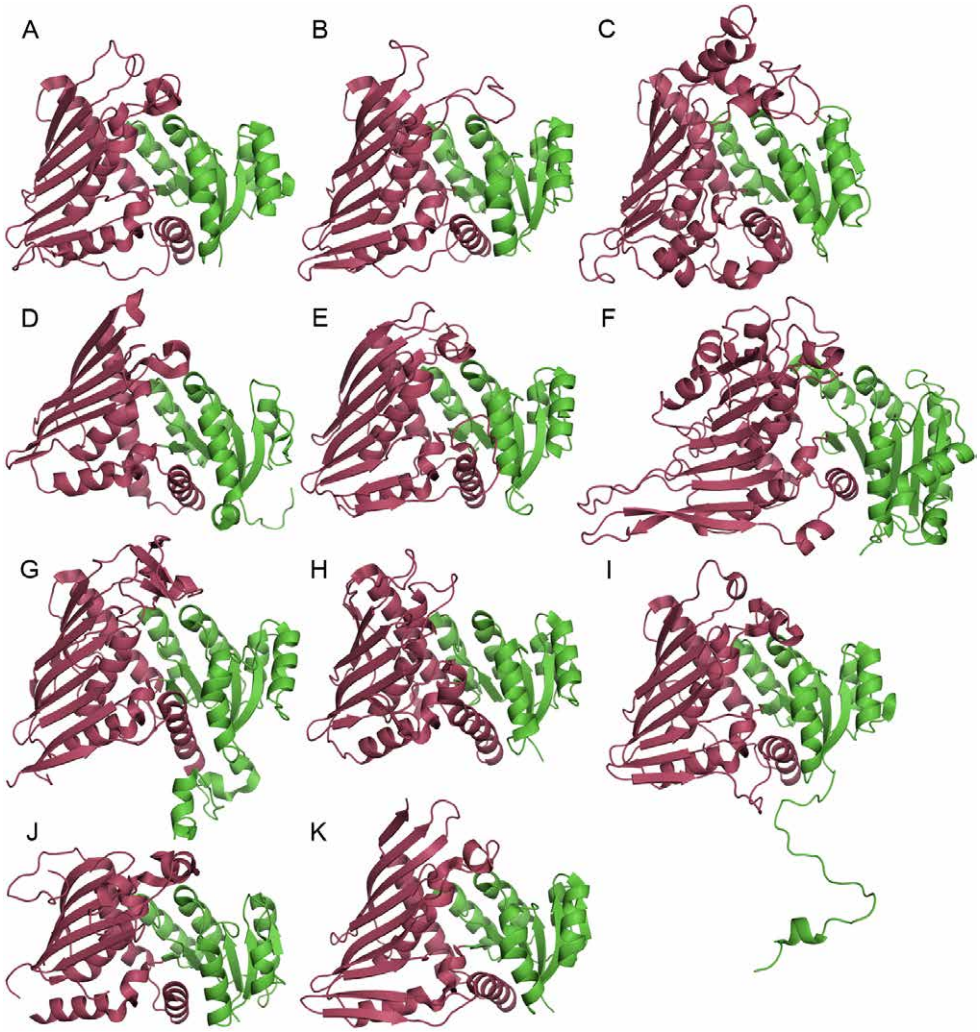


Figure 19. The monomers of the proteins used in the comparison: **A)** *Cc* AAOR, **B)** AFR, **C)** A-zyme, **D)** BVR, **E)** DD, **F)** G6DP **G)** Gal80p, **H)** IDH, **I)** *Zm* GFOR, **J)** WlbA, **K)** KijD10. The N-terminal nucleotide-binding domains are in *green* and the C-terminal domains in *red*.

The β-sheet of the α/β-domain could be divided into three sections. There were two β-hairpin-like motifs consisting of three to four adjacent antiparallel β-strands at the both ends of the β-sheet. The first strand of the β-sheet was located between them. This first strand was parallel to the first motif and antiparallel to the second motif. This first β-strand was topologically important because the catalytic residues were located in the conserved helix after this β-strand. Together these secondary structure elements formed the βα-motif, which was a characteristic feature of Gfo/Idh/MocA protein family. Figure 20 presents the simplification of the major structural elements of the Gfo/Idh/MocA protein fold.

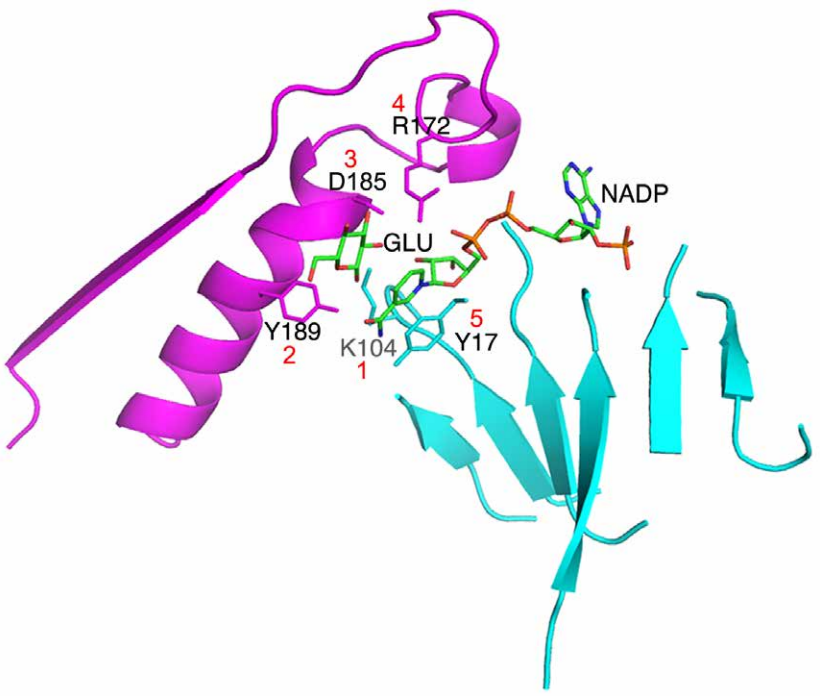


Figure 20. The simplification of the major structural elements of the Gfo/Idh/MocA protein fold (as in *Cc* AAOR). The central βα-motif is shown in *magenta* and the cofactor binding Rossman fold in *cyan*. NADP, and the glucose ligand (GLU) are shown as stick models. The side chains for the residues in the five key sites (indicated with a *red* number) are also shown as stick models.

All of the enzymes in the Gfo/Idh/MocA family catalyzed reactions using a nucleotide cofactor. In the structures where the NAD(P) was bound to the enzyme, it was binding in the expected binding-site at the C-terminal edge of the Rossmann fold, having, however, many possible conformations. To analyze the catalytic and substrate binding residues, the few available complex structures from the Gfo/Idh/MocA protein family were analyzed. Our analysis indicated positions for five residues at the active site having a key role (presented as stick models in Fig. 20 and numbered in *red*). The first two residue sites were putative catalytic residues, which could either donate or accept a proton from the substrate. The catalytic residue at the site one, which was typically a lysine, was located at the end of a N-terminal β-strand (β6 in the topology diagram of *Cc* AAOR). It made a hydrogen bond with the O2' of the ribose next to the nicotinamide ring of NAD(P) in many crystal structures. A negatively charged residue (Asp or Glu) was typically occupying the third site, and the fourth site was in general a positively charged residue (Lys or Arg). The fifth residue, located at the N-terminal domain, was packing against the reactive nicotinamide ring of the cofactor and varied considerably, having aliphatic, aromatic, or even a positively charged side chain.

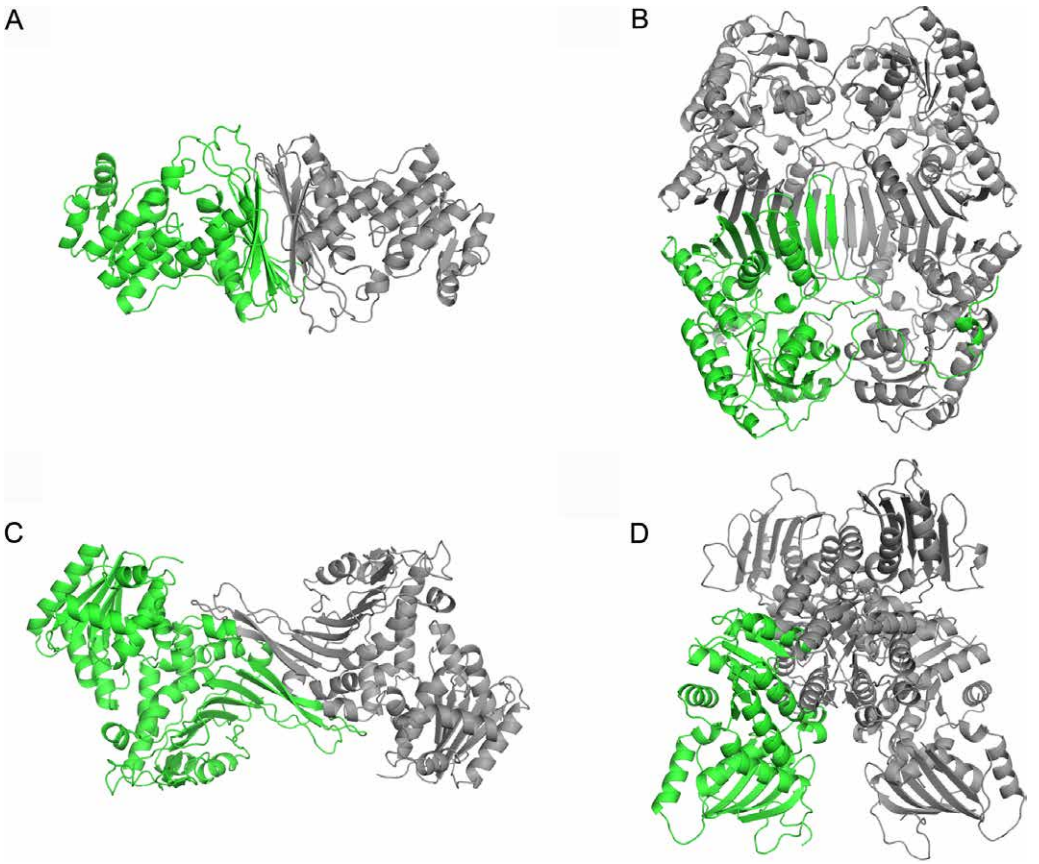


Figure 21. The quaternary structures of Gfo/Idh/MocA family proteins shown as cartoon models: **A)** AAOR dimer, **B)** GFOR tetramer, **C)** G6PD dimer, and **D)** WlbA tetramer. One monomer is presented in *green* and the other units are in *grey*.

Also the quaternary structures of the Gfo/Idh/MocA family were analyzed. BVR was the only protein found to be monomeric. The most common quaternary structure was the dimeric structure found in *Cc* AAOR (Fig. 21A), which was also found in AFR, A-zyme, DD, and Gal80p. A similar tetramer formation, as in *Zm* GFOR (Fig. 21B), was found also in KijD10 and *myo*-IDH. Also in G6PD, the β-sheet of the α/β-domain participated in the formation of the dimer, but the packing of the β-sheets against each other utilized another face of the β-sheet (Fig. 21C) because the C-terminus included a short α-helix that packed against the flat β-sheet face preventing similar packing as in *Cc* AAOR. Finally, the WlbA tetramer (Fig. 21D) represented the arrangement in which β-sheet of the α/β-domain did not have a role in the formation of the tetramer. In this WlbA structure, the monomer-monomer interfaces were predominantly located in the N-terminal nucleotide-binding domain.

4. CONCLUSIONS

The three-dimensional structures of three carbohydrate-modifying enzymes galactarolactone cycloisomeraze (Gci) and keto-deoxy-D-galactarate (KDG) dehydratase from *Agrobacterium tumefaciens*, and aldose-aldose oxidoreductase from *Caulobacter crescentus* (*Cc* AAOR) were solved by X-ray crystallography, and their function was studied by complex structure determination, UV/vis measurements, and mass spectrometry. These results provide valuable information about the protein-ligand interactions and the enzymatic mechanisms of *At* Gci, *At* KDG dehydratase, and *Cc* AAOR. These findings may be used to improve the future utilization of these enzymes in the applications for sustainable chemical production.

The novel enolase family enzyme *At* Gci was crystallized by hanging-drop vapor-diffusion method that improved the crystals for synchrotron measurements by addition of a detergent. The native structure was solved to 1.6 Å resolution, which showed us the general structure of the active site with essential amino acid residues. The attempts to solve the structure with a bound Mg^{2+} ion, required for the catalysis, and in complex with a substrate, product, or analogue of either one have so far been unsuccessful.

The structure of *At* KDG dehydratase and its complexes with pyruvate and 2-oxoadipic acid, two substrate analogues, were solved to 1.7, 1.5, and 2.1 Å resolutions, respectively. These are, to our knowledge, the first crystal structures for a decarboxylating hydrolyase and led us to propose a structure-based reaction mechanism. The solved structures, along with the kinetic characterizations on this novel type of a Class I aldolase family enzyme, add to our understanding about the oxidative D-galacturonate pathway and promote the efforts of designing mutations to improve the properties and specificity of *At* KDG dehydratase.

The structure of an oxidoreductase *Cc* AAOR, in complex with its cofactor, and several saccharides and sugar alcohols, was solved. These complex structures gave us insight in the ligand-binding interactions and demonstrated an unequalled arrangement in which two active site residues, Lys-104 and Tyr-189 could both, in principle, participate in a similar way in the catalysation. *Cc* AAOR can both oxidize and reduce a panel of saccharides to the corresponding aldonolactones and alditols, respectively, and would have enormous potential in modifying biomass sugars.

The structural and functional features of the diverse Gfo/Idh/MocA protein family were also analyzed. This family contains enzymes that catalyze various different reactions, but have a similar two-domain structure even though their sequence identity is generally very low. All the enzymes in this family have a central βα-motif carrying the catalytic residues in the conserved helix, a characteristic feature of the Gfo/Idh/MocA protein family.

ACKNOWLEDGEMENTS

This work was carried out at the Department of Chemistry at the University of Eastern Finland during the years 2012 to 2015, within the National Doctoral Programme in Informational and Structural Biology (ISB) and the Doctoral Program in Chemistry of University of Eastern Finland. The work was funded by the ISB, who is gratefully acknowledged.

I owe my deepest gratitude to my supervisors Prof. Juha Rouvinen, Prof. Janne Jänis, and Dr. Tarja Parkkinen for all their guidance, advice, and encouragement. I also wish to thank the whole organic chemistry group for being always ready to help and reflect on ideas or anything mind puzzling. Our fruitful conversations about science and beyond have made the years fly by. I would especially like to thank Ritva Romppanen for her skilled technical assistance.

The work was done in collaboration with VTT, The Technical Research Centre of Finland Ltd. Dr. Anu Koivula and Dr. Martina Andberg are warmly thanked. I would also like to thank Outi Liehunen for her invaluable technical assistance during my stay at the VTT laboratory. Aura Professional English Consulting, Ltd. is thanked for the language revision of this dissertation.

I would like to thank the whole Department of Chemistry, both the staff and students, for creating a friendly, warm, and helpful working environment. Special thanks go out to all my coffee-drinking, card-playing friends who created the inventive ambience and showed me the way to success. I would also wish to warmly thank all my friends outside the Department of Chemistry; you keep me sane in the face of all the science and give me the beat and melody for my life.

Finally, my loving thanks goes out to my family. Your love and support have carried me all the way and on to my next adventure.

October, 2015
Joensuu

Helena

"*I... a universe of atoms, an atom in the universe.*"
Richard Feynman 1918-1988

REFERENCES

1. Duff, S. J. B. and Murray, W. D. (1996) Bioconversion of forest products industry waste cellulosics to fuel ethanol: a review. *Bioresource Technol.* **55**, 1−33.

2. Jenck, J. F., Agterberg, F., and Droescher, M. J. (2004) Products and processes for sustainable chemical industry: a review of achievements and prospects. *Green Chem.* **6**, 544−556.

3. Shokri, J. and Adibkia, K. (2013) Application of cellulose and cellulose derivatives in pharmaceutical industries. In: *Cellulose – Medical, Pharmaceutical and Electronic Applications*, InTech, Rijeka, Croatia, 47−66.

4. Schmid, A., Hollmann, F., Park, J. B., and Bühler, B. (2002) The use of enzymes in the chemical industry in Europe. *Curr. Opin. Biotech.* **13**, 359−366.

5. Minussi, R. C., Pastore, G. M., and Durán, N. (2002) Potential applications of laccase in the food industry. *Trends Food Sci. Tech.* **13**, 205−216.

6. Hills, G. (2003) Industrial use of lipases to produce fatty acid esters. *Eur. J. Lipid Sci-Tech.* **105**, 601−607.

7. Jayani, R. S., Saxena, S, and Gupta, R. (2005) Microbial pectinolytic enzymes: a review. *Process Biochem.* **40**, 2931−2944.

8. Mohnen, D. (2008) Pectin structure and biosynthesis. *Curr. Opin. Plant. Biol.* **11**, 266−277.

9. Grohmann, K. and Bothast, J. (1994) Pectin-rich residues generated by processing of citrus fruits, apples, and sugar beet: enzymatic hydrolysis and biological conversion to value-added products. In: *Enzymatic conversion of biomass for fuels production*. ACS symposium series 566, Washington DC, USA, 372−390.

10. Ochoa-Villarreal, M., Aispuro-Hernández, E., Vargas-Arispuro, I., and Martínez-Téllez, M. Á. (2012) Plant cell wall polymers: function, structure and biological activity of their derivatives. In: *Polymerization*, InTech, Rijeka, Croatia, 63−86.

11 Ashwell, G, Wahba, A. J., and Hickman, J. (1958) A new pathway of uronic acid metabolism. *Biochim. Biophys. Acta* **30**, 186−187.

12. Chang, Y. F. and Feingold, D. S. (1970) D-glucaric acid and galactaric acid catabolism by *Agrobacterium tumefaciens*. *J. Bacteriol.* **102**, 85−96.

13. Richard, P. and Hilditch, S. (2009) D-Galacturonic acid catabolism in microorganisms and its biotechnological relevance. *Appl. Microbiol. Biotechnol*. **82**, 597−604.

14. Zajic, J. E. (1959) Hexuronic dehydrogenase of *Agrobacterium tumefaciens*. *J. Bacteriol*. **78**, 734−735.

15. Kilgore, W. W. and Starr, M. P. (1959) Uronate oxidation by phytopathogenic pseudomonads. *Nature* **183**, 1412−1413.

16. Boer, H., Maaheimo, H., Koivula, A., Penttilä, M., and Richard, P. (2010) Identification in *Agrobacterium tumefaciens* of the D-galacturonic acid dehydrogenase gene. *Appl. Microbiol. Biotechnol*. **86**, 901−909.

17. Parkkinen, T., Boer, H., Jänis, J., Andberg, M., Penttilä, M., Koivula, A., and Rouvinen, J. (2011) Crystal structure of uronate dehydrogenase from *Agrobacterium tumefaciens*. *J. Biol. Chem.* **286**, 27294−27300.

18. Bouvier, J. T., Groninger-Poe, F. P., Vetting, M., Almo, S. C., and Gerlt, J. A. (2014) Galactaro δ-lactone isomerase: lactone isomerization by a member of the amidohydrolase superfamily. *Biochemistry* **53**, 614−616.

19. Andberg, M., Maaheimo, H., Boer, H., Penttilä, M., Koivula, A., and Richard, P. (2012) Characterization of a novel *Agrobacterium tumefaciens* galactarolactone cycloisomerase enzyme for direct conversion of D-galactarolactone to 3-deoxy-2-keto-L-*threo*-hexarate. *J. Biol. Chem.* **287**, 17662−17671.

20. Jeffcoat, R., Hassall, H., and Dagley, S. (1969) Purification and properties of D-4-deoxy-5-oxoglucarate hydro-lyase (decarboxylating). *Biochem. J.* **115**, 977−983.

21. Aghaie, A., Lechaplais, C., Sirven, P., Tricot, S., Besnard-Gonnet, M., Muselet, D., de Berardinis, V., Kreimeyer, A., Gyapay, G., Salanoubat, M., and Perret, A. (2008) New insights into the alternative D-glucarate degradation pathway. *J. Biol. Chem.* **283**, 15638−15646.

22. Watanabe, S., Yamada, M., Ohtsu, I., and Makino, K. (2007) Alpha-ketoglutaric semialdehyde dehydrogenase isozymes involved in metabolic pathways of D-glucarate, D-galactarate, and hydroxy-Lproline: molecular and metabolic convergent evolution. *J. Biol. Chem.* **282**, 6685−6695.

23. Holm, L. and Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucl. Acids Res.* **38**, W545−549.

24. Patskovsky, Y., Kim, J., Toro, R., Bhosle, R., Hillerich, B., Seidel, R.D., Washington, E., Scott Glenn, A., Chowdhury, S., Evans, B., Hammonds, J., Zencheck, W.D., Imker, H.J., Gerlt, J.A., and Almo, S.C. (2011) Crystal structure of enolase Spea_3858 (target EFI-500646) from *Shewanella pealeana* with magnesium bound.

25. Sugadev, R., Eswaramoorthy, S. Burley, S. K., and Swaminathan, S. (2007) Crystal structure of a putative dehydratase from *Mesorhizobium loti*.

26. Patsokovsky, Y., Sauder, J. M., Diskey, M., Adams, J. M., Ozyurt, S., Wasserman, S. R., Gerlt, J., Burley, S. K., and Almo, S. C. (2006) Crystal structure of putative enolase from *Salmonella typhimurium* Lt2.

27. Gerlt, J. A., Babbitt, P. C., Jacobson, M. P., and Almo, S. C. (2012) Divergent evolution in enolase superfamily: strategies for assigning functions. *J. Biol. Chem.* **287**, 29−34.

28. Gerlt, J. A., Babbitt, P. C., and Rayment, I. (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch. Biochem. Biophys.* **433**, 59−70.

29. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R. J., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L., and Gelrt, J. A. (1996) The enolase superfaminly: a general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids. *Biochemistry* **35**, 16489−16501.

30. Gefflaut, T., Blonski, C., Perie, J., and Willson, M. (1995) Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects. *Prog. Biophys. Molec. Biol.* **63**, 301−340.

31. Theodossis, A., Walden, H., Westwick, E. J., Connaris, H., Lamble, H. J., Hough, D. W., Danson, M. J., and Taylor, G. L. (2004) The structural basis for substrate promiscuity in 2-keto-3-deoxygluconate aldolase from the Entner-Doudoroff pathway in *Sulfolobus solfataricus*. *J. Biol. Chem*. **279**, 43886−43892.

32. Choi, K. H., Lai, V., Foster, C. E., Morris, A. J., Tolan, D. R., and Allen, K. N. (2006) New superfamily members identified for Schiff-base enzymes based on verification of catalytically essential residues. *Biochemistry* **45**, 8546−8555.

33. Lawrence, M. C., Barbosa, J. A. R. G., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C., and Marcuccio, S. M. (1997) Structure and mechanism of a sub-family of enzymes related to N-acetylneuraminate lyase. *J. Mol. Biol.* **266**, 381−399.

34. Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G., and Colman, P. M. (1994) The three-dimensional sturucture of N-acetylneuraminate lyase from *Escherichia coli*. *Structure* **2**, 361−369.

35. Blom, N and Sygusch J. (1997) Product binding and role of the C-terminal region in class I D-fructose-1,6-bisphosphate aldolase. *Nature* **4**, 36−39.

36. Allard, J., Grochulski, P, and Sygusch, J. (2001) Covalent intermediate trapped in 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase structure at 1.9 Å resolution. *PNAS* **98**, 3679−3684.

37. Groninger-Poe, F. P., Bouvier, J. T., Vetting, M. W., Kalyanaraman, C., Kumar, R., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2014) Evolution of enzymatic activities in the enolase superfamily: galactarate dehydratase-III from *Agrobacterium tumefaciens* C58. *Biochemistry* **53**, 4192−4203.

38. Jeffcoat, R., Hassall, H., and Dagley, S. (1969) Purification and properties of D-4-deoxy-5-oxoglucarate hydro-lyase (decarboxylating). *Biochem. J.* **115**, 977−983.

39. Berghall, S., Hilditch, S., Penttilä, M., and Richard, P. (2007) Identification in the mould *Hypocrea jecorina* of a gene encoding an NADP(+): D-xylose dehydrogenase. *FEMS Microbiol. Lett.* **277**, 249−253.

40. Wiebe, M., Nygård, Y., Oja, M., Andberg, M., Ruohonen, L., Koivula, A., Penttilä, M., and Toivari, M. (2015) A novel aldose-aldose oxidoreductase for co-production of D-xylonate and xylitol from D-xylose with *Saccharomyces cerevisiae*. *Appl. Microbiol. Biot.* doi: 10.1007/s00253-015-6878-5.

41. Mussatto, S. I. and Teixeira, J. A. (2010) Lignocellulose as raw material in fermentation processes. In: *Current Research, technology and education topics in applied microbiology and microbial biotechnology*, Vol. 2, Formatex, Badajoz, Spain, 897–907.

42. Andberg, M., Maaheimo, H., Kumpula, E.-P., Boer, H., Toivari, M., Penttilä, M., and Koivula, A. (2015) Characterization of a unique *Caulobacter crescentus* aldose-aldose oxidoreductase having a dual activities. *Appl. Microbiol. Biot.*

43. Cleland, W. W. (1963) The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. *Biochim. Biophys. Acta* **67**,104−137.

44. Hua, Y. H., Wu, C. Y., Sargsyan, K., and Lim, C. (2014) Sequence-motif detection of NAD(P)-binding proteins: Discovery of a unique antibacterial drug target. *Sci. Rep.* **4**, 6471.

45. Lin, S.-F., Yang, T.-Y., Inukai, T., Yamasaki, M., and Tsai, Y.-C. (1991) Purification and characterization of a novel glucooligosaccharide oxidase from *Acremonium strictum* T1. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1118**, 41–47.

46. Xu, F., Golightly, E. J., Fuglsang, C. C., Schneider, P., Duke, K. R., Lam L., Christensen, S., Brown, K. M., Jørgensen, C. T., and Brown, S. H. (2001) A novel carbohydrate:acceptor oxidoreductase from *Microdochium nivale*. *Eur. J. Biochem.* **268**,1136–1142.

47. Kiryu, T., Nakano, H., Kiso, T., and Murakami, H. (2008) Purification and characterization of a carbohydrate:acceptor oxidoreductase from *Paraconiothyrium* sp. that produces lactobionic acid efficiently. *Biosci. Biotechnol. Biochem.* **72**, 833–841.

48. Lee, M.-H., Lai, W.-L., Lin, S.-F., Liu, Y., Hsu, Y.-H., and Tsai, Y.-C. (2006) Purification and characterization of a novel cellooligosaccharide oxidase from rice pathogen *Sarocladium oryzae*. *Enzyme Microb. Technol.* **39**, 85–91.

49. Heuts, D. P., Janssen, D. B., and Fraaije, M. W. (2007) Changing the substrate specificity of a chitooligosaccharide oxidase from *Fusarium graminearum* by model-inspired site-directed mutagenesis. *FEBS Lett.* **581**, 4905–4909.

50. Whittaker, J. W. (2005) The radical chemistry of galactose oxidase. *Arch. Biochem. Biophys.* **433**, 227–39.

51. Rowland, P., Basak, A. K., Gover, S., Levy, H., and Adams, M. J. (1994) The three-dimensional structure of glucose 6-phosphate dehydrogenase from *Leunostoc mesenteroides* refined at 2.0 Å resolution. *Structure* **2**, 1073−1087.

52. Kingston, R. L., Scopes, R. K., and Baker, E. N. (1996) The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure* **4**, 1413−1428.

53. Dambe, T.R., Kuehn, A.M., Brossette, T., Giffhorn, F., and Scheidig, A.J. (2006) Crystal structure of NADP(H)-dependent 1,5-anhydro-d-fructose reductase from *Sinorhizobium morelense* at 2.2 Å resolution: construction of a NADH-accepting mutant and its application in rare sugar synthesis. *Biochemistry* **45**, 10030−10042.

54. Thoden, J.B. and Holden, H.M. (2010) Structural and functional studies of WlbA: a dehydrogenase involved in the biosynthesis of 2,3-diacetamido-2,3-dideoxy-d-mannuronic acid. *Biochemistry* **49**, 7939−7948.

55. Kubiak, R. L. and Holden, H. M. (2011) Combined structural and functional investigation of a C-3’’-ketoreductase involved in the biosynthesis of dTDP-L-digitoxose. *Biochemistry* **50**, 5905−5917.

56. van Straaten, K.E., Zheng, H., Palmer, D.R., and Sanders, D.A. (2010) Structural investigation of myo-inositol dehydrogenase from *Bacillus subtilis*: implications for catalytic mechanism and inositol dehydrogenase subfamily classification. *Biochem. J.* **432**, 237−247.

57. Kikuchi, A., Park, S.Y., Miyatake, H., Sun, D., Sato, M., Yoshida, T., and Shiro, Y. (2001) Crystal structure of rat biliverdin reductase. *Nat. Struct. Biol.* **8**, 221−225.

58. Liu, Q.P., Sulzenbacher, G., Yuan, H., Bennett, E.P., Pietz, G., Saunders, K., Spence, J., Nudelman, E., Levery, S.B., White, T., Neveu, J.M., Lane, W.S., Bourne, Y., Olsson, M.L., Henrissat, B., and Clausen, H. (2007) Bacterial glycosidases for the production of universal red blood cells. *Nature Biotechnol.* **25**, 454−464.

59. Carbone, V., Endo, S., Sumii, R., Chung, R.P., Matsunaga, T., Hara, A., and El-Kabbani, O. (2008) Structures of dimeric dihydrodiol dehydrogenase apoenzyme and inhibitor complex: probing the subunit interface with site-directed mutagenesis. *Proteins* **70**, 176−187.

60. Thoden, J. B., Sellick, C. A., Reece, R. J., and Holden, H. M. (2007) Understanding a transcriptional paradigm at the molecular level. The structure of yeast Gal80p. *J. Biol. Chem.* **282**, 1537−1538.

61. Rubio-Texeira, M (2005) A comparative analysis on the GAL genetic switch between not-so-distant cousins: *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. *FEMS Yeast Res.* **5**, 1115−1128.

62. Rao, S. T. and Rossman, M. G. (1973) Comparison of super-secomdary structures in proteins. *J. Mol. Biol.* **76**, 241−256.

63. Day, R., Beck, D. A. C., Armen, R. S., and Daggett, V. (2003) A consensus view of fold space: combining SCOP, CATH, and the Dali domain dictionary. *Protein Sci.* **12**, 2150−2160.

64. Mayevsky, A. and Rogatsky, G. G. (2007) Mitochondrial function *in vivo* evaluated by NADH fluorescence: from animal models to human studies. *Am. J. Physiol. Cell Physiol.* **292**, C615−C640.

65. von Stetten, D., Giraud, T., Carpentier, P., Sever, F., Terrien, M., Dobias, F., Juers, D. H., Flot, D., Mueller-Dieckmann, C., Leonard, G. A., de Sanctis, D., and Royant, A. (2015) In crystallo optical spectroscopy (icOS) as a complementary tool on the macromolecular crystallography beamlines of the ESRF. *Acta Crystallogr. D* **71**, 15−26.

66. Pasanen, S., Jänis, J., and Vainiotalo, P. (2007) Cello-, malto- and xylo-oligosaccharide fragmentation by collision-induced dissociation using QIT and FT-ICR mass spectrometry: a systematic study. *Int. J. Mass Spectrom.* **263**, 22−29.

67. de Sanctis, D., Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., Mairs, T., McSweeney, S., Petitdemange, S., Rey-Bakaikoa, V., Surr, J., Theveneau, P., Leonard, G. A., and Mueller-Dieckmann, C. (2012) ID29: A high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering. *J. Synchrotron Rad.* **19**, 455−461.

68. Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S., and Mitchell, E. (2010) The ID23-2 structural biology microfocus beamline at the ESRF. *J. Syncrotron Radiat.* **17**, 107−118.

69. Evans, P. (2005) Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72−82.

70. Zwart, P. H., Grosse-Kunstleve, R. W., and Adams, P. D. (2005) Characterization of X-ray data sets. *CCP4 Newsletter* **42**.

71. Zwart, P. H., Grosse-Kunstleve, R. W., and Adams, P. D. (2005) Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 Newsletter* **43**.

72. Kabsch, W. (2010) Xds. *Acta Crystallogr. D* **66**, 125−132.

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

74. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760−763.

75. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235−242.

76. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, N. A., Wimm, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355−367.

77. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658−674.

78. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213−221.

79. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. *Acta Crystallogr. D* **64**, 61−69.

80. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of *coot*. *Acta Crystallogr. D* **66**, 486−501.

81. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352−367.

82. Gutmanas, A., Alhroub, Y., Battle, G. M., Berrisford, J. M., Bochet, E., Conroy, M. J., Dana, J. M., Fernandez Montecelo, M. A., van Ginkel, G., Gore, S. P., Haslam, P., Hatherley, R., Hendrickx, P. M., Hirshberg, M., Lagerstedt, I., Mir, S., Mukhopadhyay, A., Oldfield, T. J., Patwardhan, A., Rinaldi, L., Sahni, G., Sanz-Garcia, E., Sen, S., Slowley, R. A., Velankar, S., Wainwright, M. E., and Kleywegt, G. J. (2014) PDBe: Protein data bank in Europe. *Nucleic Acids Res.* **42**, D285−91.

83. Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491−497.

84. Kantardjieff, K. A. and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **12**, 1865−1871.

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

86. The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.

87. Olson, S. A. (1994) MacVector: An integrated sequence analysis program for the macintosh. *Methods Mol. Biol.* **25**, 195−201.

88. Chen, V. B., Arendall 3rd, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12−21.

89. Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nature Struct. Biol.* **10**, 980.

90. Fedorov, A. A., Fodorov, E. V., Sauder, J. M., Burley, S. K., Gerlt, J. A., and Almo, S. C. (2008) Crystal structure of dihydrodipicolinate sunthase from *Bacillus clausii*.

91. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403−410.

92. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235−242.

93. Daniels, A. D., Campeotto, I., van der Kamp, M. W., Bolt, A. H., Trinh, C. H., Phillips, S. E. V., Pearson, A. R., Nelson, A., Mulholland, A. J., and Berry, A. (2014) Reaction mechanism of N-acetylneuraminic acid lyase revealed by a combination of crystallography, QM/MM simulation, and mutagenesis. *ACS Chem. Biol.* **9**, 1025−1032.

94. Nurizzo, D., Halbig, D., Sprenger, G. A., and Baker, E. N. (2001) Crystal structures of the precursor form of glucose-fructose oxidoreductase from *Zymomonas mobilis* and its complexes with bound ligands. *Biochemistry* **40**, 13857−13867.

95. Zachariou, M. and Scopes, R. K. (1986) Glucose-fructose oxidoreductase, a new enzyme isolated from *Zymomonas mobilis* that is responsible for sorbitol production. *J. Bacteriol.* **167**, 863−869.

96. Hardman, M. J. and Scopes, R. K. (1988) The kinetics of glucose-fructose oxidoreductase from *Zymomonas mobilis*. *Eur. J. Biochem.* **173**, 203−209.

97. Lott, J. S., Halbig, D., Baker, H. M., Hardman, M. J., Sprenger, G. A., and Baker, E. N. (2000) Crystal structure of a truncated mutant of glucose-fructose oxidoreductase shows that an N-terminal arm controls tetramer formation. *J. Mol. Biol.* **304**, 575−584.

98. Kim, Y., Arora, M., Straza, M., and Joachimiak, A. (2005) Crystal structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*.

99. Kim, Y., Arora, M., Straza, M., Donnelly, M., and Joachimiak, A. (2005) Crystal structure of the shifted form of the glucose-fructose oxidoreductase from *Zymomonas mobilis*.

100. Berman, H. M., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nature Struct. Biol.* **10**, 980.

101. Whitby, F. G., Phillips, J. D., Hill, C. P., McCoubrey, W., and Maines, M. D. (2002) Crystal structure of a biliverdin IXα reductase enzyme-cofactor complex. *J. Mol. Biol.* **319**, 1199−1210.

102. Schu, M., Faust, A., Stosik, B., Kohring, G. W., Giffhorn, F., and Scheidig, A. J. (2013) The structure of substrate-free 1,5-anhydro-D-fructose reductase from *Sinorhizobium meliloti* 1021 reveals an open enzyme conformation. *Acta Crystallogr. F* **69**, 844−849.

103. Carbone, V., Sumii, R., Ishikura, S., Asada, Y., Hara, A., and El-Kabbani, O. (2008) Structure of monkey dimeric dihydrodiol dehydrogenase in complex with isoascorbic acid. *Acta Crystallogr. D* **64**, 532−542.

104. Zheng, H., Bertwistle, D., Sanders, D. A., and Palmer, D. R. (2013) Converting NAD-specific inositol dehydrogenase to an efficient NADP-selective catalyst, with a surprising twist. *Biochemistry* **52**, 5876−5883.

105. Bond, C. S. (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311−312.

102/2010 NIEMI Merja: A molecular basis for antibody specificity – crystal structures of IgE-allergen and IgG-hapten complexes

103/2010 RASILAINEN Tiina: Controlling water on polypropylene surfaces with micro- and micro/nano structures

104/2011 SAARIKOSKI Inka: Tailoring of optical transmittance, reflectance, and hydrophobicity of polymers by micro- and nanoscale structuring

105/2011 NISKANEN Mika: DFT Studies on ruthenium and rhodium chain complexes and a one-dimensional iodine bridged ruthenium complex

106/2011 RÖNKKÖ Hanna-Leena: Studies on $MgCl_2$/alcohol adducts and a self-supported Ziegler-Natta catalyst for propene polymerization

107/2011 KASANEN Jussi: Photocatalytic TiO_2-based multilayer coating on polymer substrate for use in self-cleaning applications

108/2011 KALLIO Juha: Structural studies of *Ascomycete* laccases – Insights into the reaction pathways

109/2011 KINNUNEN Niko: Methane combustion activity of Al_2O_3-supported Pd, Pt, and Pd-Pt catalysts: Experimental and theoretical studies

110/2011 TORVINEN Mika: Mass spectrometric studies of host-guest complexes of glucosylcalixarenes

111/2012 KONTKANEN Maija-Liisa: Catalyst carrier studies for 1-hexene hydroformulation: cross-linked poly(4-vinylpyridine), nano zinc oxide and one-dimensional ruthenium polymer

112/2012 KORHONEN Tuulia: The wettability properties of nano- and micromodified paint surfaces

113/2012 JOKI-KORPELA Fatima: Functional polyurethane-based films and coatings

114/2012 LAURILA Elina: Non-covalent interactions in Rh, Ru, Os, and Ag complexes

115/2012 MAKSIMAINEN Mirko: Structural studies of *Trichoderma reesei, Aspergillus oryzae* and *Bacillus circulans* sp. *alkalophilus* beta-galactosidases – Novel insights into a structure-function relationship

116/2012 PÖLLÄNEN Maija: Morphological, thermal, mechanical, and tribological studies of polyethylene composites reinforced with micro– and nanofillers

117/2013 LAINE Anniina: Elementary reactions in metallocene/methylaluminoxane catalyzed polyolefin synthesis

118/2013 TIMONEN Juri: Synthesis, characterization and anti-inflammatory effects of substituted coumarin derivatives

119/2013 TAKKUNEN Laura: Three-dimensional roughness analysis for multiscale textured surfaces: Quantitative characterization and simulation of micro- and nanoscale structures

120/2014 STENBERG Henna: Studies of self-organizing layered coatings

121/2014 KEKÄLÄINEN Timo: Characterization of petroleum and bio-oil samples by ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometry

122/2014 BAZHENOV Andrey: Towards deeper atomic-level understanding of the structure of magnesium dichloride and its performance as a support in the Ziegler-Natta catalytic system

123/2014 PIRINEN Sami: Studies on $MgCl_2$/ether supports in Ziegler–Natta catalysts for ethylene polymerization

124/2014 KORPELA Tarmo: Friction and wear of micro-structured polymer surfaces

125/2014 HUOVINEN Eero: Fabrication of hierarchically structured polymer surfaces

126/2014 EROLA Markus: Synthesis of colloidal gold and polymer particles and use of the particles in preparation of hierarchical structures with self-assembly

127/2015 KOSKINEN Laura: Structural and computational studies on the coordinative nature of halogen bonding

128/2015 TUIKKA Matti: Crystal engineering studies of barium bisphosphonates, iodine bridged ruthenium complexes, and copper chlorides

129/2015 JIANG Yu: Modification and applications of micro-structured polymer surfaces