Structure-function relationships in NDP-sugar active SDR enzymes: Fingerprints for functional annotation and enzyme engineering

Matthieu Da Costa, Ophelia Gevaert, Stevie Van Overtveldt, Joanna Lange, Henk-Jan Joosten, Tom Desmet, Koen Beerens

PII:	S0734-9750(21)00011-2
DOI:	https://doi.org/10.1016/j.biotechadv.2021.107705
Reference:	JBA 107705
To appear in:	*Biotechnology Advances*
Received date:	8 July 2020
Revised date:	18 December 2020
Accepted date:	27 January 2021

Please cite this article as: M. Da Costa, O. Gevaert, S. Van Overtveldt, et al., Structure-function relationships in NDP-sugar active SDR enzymes: Fingerprints for functional annotation and enzyme engineering, *Biotechnology Advances* (2019), https://doi.org/10.1016/j.biotechadv.2021.107705

Structure-function relationships in NDP-sugar active SDR enzymes: Fingerprints for functional annotation and enzyme engineering.

Matthieu Da Costa[a], Ophelia Gevaert[a], Stevie Van Overtveldt[a], Joanna Lange[b], Henk-Jan Joosten[b], Tom Desmet[a]*, Koen Beerens[a]*

a Centre for Synthetic Biology – Unit for Biocatalysis and Enzyme Engineering, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Gent, Belgium.

b Bio-Prodict BV, Nieuwe Marktstraat 54E, 6511 AA, Nijmegen, The Netherlands

tom.desmet@ugent.be (TD); koen.beerens@ugent.be (KB)

Abstract

Short-chain Dehydrogenase/Reductase enzymes that are active on nucleotide sugars (abbreviated as NS-SDR) are of paramount importance in the biosynthesis of rare sugars and glycosides. Some family members have already been extensively characterized due to their direct implication in metabolic disorders or in the biosynthesis of virulence factors. In this review, we combine the knowledge gathered from studies that typically focused only on one NS-SDR activity with an in-depth analysis and overview of all of the different NS-SDR families (169,076 enzyme sequences). Through this structure-based multiple sequence alignment of NS-SDRs retrieved from public databases, we could identify clear patterns in conservation and correlation of crucial residues. Supported by this analysis, we suggest updating and extending the UDP-galactose 4-epimerase "hexagonal box model" to an "heptagonal box model" for all NS-SDR enzymes. This specificity model consists of seven conserved regions surrounding the NDP-sugar substrate that serve as fingerprint for each specificity. The specificity fingerprints highlighted in this review will be beneficial for functional annotation of the large group of NS-SDR enzymes and form a guide for future enzyme engineering efforts focused on the biosynthesis of rare and specialty carbohydrates.

Keywords: NDP-sugars, Specificity fingerprints, Epimerase, Dehydratase, Reductase, Decarboxylase, Structure-function relationship, Short-chain Dehydrogenase/Reductase, SDR superfamily.

Contents

Table of Contents

1. Introduction

Carbohydrates are the most abundant biomolecules in nature and found in every domain of life (Varki, et al., 2015). No less than 344 distinct appended carbohydrates have been found in glycosylated bacterial secondary metabolites (such as antibiotics, antivirals, anticancer molecules, etc.) and more than 100 sugar moieties have been encountered in bacterial polysaccharides (Elshahawi et al., 2015; Thibodeaux et al., 2007). Glycan profiles have the ability to greatly influence the physicochemical and biological properties of protein therapeutics and both natural and man-made organic compounds (e.g. solubility, stability, binding to receptor, activity and/or pharmacokinetics) (Thorson et al., 2005). Leveraging this extraordinary pool of bacterial carbohydrates could undoubtedly lead to a unique opportunity for glycodiversification intentions (Elshahawi et al., 2015). Nonetheless, these carbohydrates are not abundant in nature, which renders their large-scale production challenging (Beerens et al., 2012).
The modifications of common sugars to their rare counterparts often happen at the level of the nucleotide sugar, mostly represented by nucleotide-diphosphate (NDP) sugars, before being used alone or built into glycosides or polysaccharides (Thibodeaux et al., 2008, 2007). A close examination of the biochemical reactions catalyzed by NDP-sugar active enzymes suggests that the majority of these enzymes belong to one of the largest known protein families: the Short-chain Dehydrogenase/Reductase (SDR) superfamily (Thibodeaux et al., 2007). The subgroup of SDRs active on NDP-sugars (NS-SDRs) has been found to execute a very diverse range of chemical catalysis, including epimerization (EC 5), dehydration and decarboxylation (both EC 4) and oxidoreduction (EC 1) (Fig. 1.A) (Borg et al., 2021). This functional diversity is noteworthy knowing that they share the same protein fold and that the core catalytic mechanism is driven by two highly conserved residues: a tyrosine serving as Brønsted acid/base for reduction/oxidation and four amino acids distanced to that, a conserved lysine essential for binding the nicotinamide ribose and lowering the tyrosine's hydroxyl pKa (Fig. 1.A&B). This catalytic dyad Yx_3K is basically conserved among all SDRs family members, although some variations have been reported (e.g. Mx_3K in Pglf (Riegert et al., 2017). It has been demonstrated that a conserved serine or threonine plays an essential role in the mechanism by facilitating the proton abstraction and its transfer back to the oxygen since the tyrosine is often positioned too far away (4.9 Å) to directly abstract the proton (Liu et al., 1997), hence these authors refer to a catalytic triad ($[ST]x_nYx_3K$). Not less important, the [ST] also stabilizes and polarizes the carbonyl substrate group generated after oxidation.

Figure 1. Functional diversity in NDP-sugar active SDR enzymes. A. The epimerases, dehydratases and decarboxylases introduce the 4-keto-functionality in their first step, which is catalysed by the conserved Yx3K diad and NAD cofactor (black box). The dehydratase reaction product already contains the 4-keto functionality that is necessary for the reductase reactions (green box). **B.** CDP-paratose 2-epimerase (CPa2E) is an exception as it starts with oxidation at C2 (dark blue box). The functional groups on C2 (-OH or -NHAc) and C3 (-OH) are not shown in the first steps (black and grey) for simplicity reasons. R = -CH2OH for UDP-Glc, UDP-Gal, UDP-GlcNAc, GDP-Man and R = -CO2- for UDP-GlcA and UDP-GalA.

Engineering of these SDR members towards alternative substrates and/or reactions would be highly valuable to create new pathways for the production of rare sugars and glycosides. However, our knowledge about their structure-function relationships is currently too limited to enable efficient rational design. Another impediment is the constant increase of uncharacterized enzymes in public databases that makes their functional assignments challenging (Jacobson et al., 2014). Most prediction tools for genome functional annotations rely on sequence and/or structural homology-based algorithms. For well-studied enzyme families, in which annotation has been experimentally confirmed by biochemical and/or mutagenesis studies, a homology-based approach is often accurate and leads to correct annotation (Schnoes et al., 2009). However, in other cases, this

method might be hampered by the lack of information within an enzyme superfamily leading to annotation errors (Loewenstein et al., 2009; Schnoes et al., 2009). Recently, the UDP-glucuronic acid 5-epimerase specificity description present in public sequence database was debunked and the enzymes were shown to be actually UDP-glucuronic acid 4-epimerase (Gevaert et al., 2020).
One solution to this problem is to make use of sequence fingerprints that are specific for an enzyme subfamily and associate them with functionally related proteins (Dudek et al., 2017). Recently, Gräff *et co.* established an SDR database that allowed the classification of enzyme sequences based on superfamily and cofactor-specific motifs (Gräff et al., 2019). This approach allowed them to refine the cofactor-binding motif, previously referred to as Glycine motif ($Gx_{2-3}Gx_{1-2}G$) into [LVI][VI]TG[AG]x_2G[IL]G or L[VI]TG[GA]xGx[IVL]G specific to classical or extended SDRs, respectively, and identified a novel signature motif (i.e. NNAG or HxAA, respectively).
To the best of our knowledge, detailed analysis of the specificity fingerprints of the group of NS-SDR enzymes has never been reported. It would be highly relevant and could define distinct characteristics of the different reactions for functional assignments of new sequences added to the subfamily or in light of engineering for rare carbohydrate synthesis and could perhaps even reveal novel activities and/or specificities. In the sequence and literature review hereunder, we intend to provide in-depth information regarding the sequence-function relationships among the NS-SDR enzymes by taking advantage of an SDR superfamily wide structure alignment and known experimental studies.

2. Methodology

In order to review published enzyme sequences/structures and facilitate the comparison between structurally equivalent positions, the sequence numbering of NS-SDR enzymes had to be standardized. To do so, we used 3DM (https://3dm.bio-prodict.com), which is a protein (super)family analysis platform (Kuipers et al., 2010). At the base of each 3DM system lies a highly accurate structure-based multiple sequence alignment (MSA) that leverages known three-dimensional crystalized structures to align sequentially distinct regions of a protein family. Detailed explanation on how data are collected, and how the database is generated can be found in the review by Kuipers et al. (2010), Steffen-Munsberg et al. (2015) and Van den Bergh et al., (2017). In a nutshell, all structures of a superfamily, sharing the same fold, are superposed and an alignment of the common structurally equivalent positions, called core positions, is determined. This structural alignment serves as a starting point to generate the structure-based MSA. A 3D numbering scheme is then applied to all residues that belong to the structurally conserved core to allow for easy data transfer between core positions of different proteins in the alignment. Once generated, the 3DM database provides a myriad of protein-related information (e.g. amino acid conservation, hydrophobicity, ligand contacts, mutations extracted from literature and patents, correlated mutation data extracted from the alignments, etc.), which has proven beneficial for enzyme engineering and discovery projects (Junker et al., 2018; Kourist et al., 2010; Lanfranchi et al., 2017; Steffen-Munsberg et al., 2015).

With this in mind, a database comprising all currently known SDR enzymes was constructed and included remotely related proteins. Because this selection of SDR enzymes comprised crystal structures that are very diverse (3165 structures in the PDB), only a relatively small structural core corresponding to the Rossmann-fold could be generated. Therefore, we refined the alignment by selecting only crystal structures of enzymes active on NDP-sugars (173 structures in the PDB), giving rise to 28 subfamilies (Table 1). In light of our long-standing interest in carbohydrate epimerases (Beerens et al., 2017; Gevaert et al., 2019; Van Overtveldt et al., 2015) and more specifically in epimerases active on the second asymmetric carbon (Rapp et al., 2020; Van Overtveldt et al., 2020), which still hold unraveled potential, we targeted the CDP-paratose 2-epimerase from *Salmonella* *typhi*, *st*TyvE (PDB: 1ORR) as template for the new alignment, leading to a core composed of every position of the *st*TyvE protein sequence (Fig. 2). Consequently, from now on, the sequence positions mentioned in this review will be the ones of *st*TyvE, and if not, it will be specified in between brackets.

Figure 2. Protein topology of CDP-paratose 2-epimerase (PDB: 1ORR) used as template for 3DM alignment. Topology edited from PDBSum: http://www.ebi.ac.uk/pdbsum/.

We manually clustered the enzymes in subsets, which can be considered to be 'mini' 3DM systems within the database, based on each NS-SDR subfamily provided by the general alignment. All alignment statistics are regenerated and can be separately analyzed. Subset data can also be compared against other subsets or against the full set of sequences, thus giving information regarding subset specific conservation and correlated positions. We therefore created several subsets by sorting the 28 subfamilies based on acknowledged specificities highlighted in literature (Table 1). In epimerases, we sorted the four different stereocenter specificities (i.e. C2-, C6-, C4- and C3,5-epimerization). Despite catalyzing nearly identical reactions, NDP-sugar 4,6-dehydratases comprise different nucleotide specificities and sometimes exhibit a side 5-epimerase activity. This led us to splitting this subgroup in four distinct subsets and an additional one for the 5,6-dehydratase activity. Reductases were split in three subsets based on the product formed, whereas decarboxylases were differentiated by their ability to synthesize either only UDP-xylose or both UDP-apiose and UDP-xylose.

In order to have a more accurate view on which position could be important for influencing specificity and/or reactivity, we also created a subset for each reactivity (I-IV) on top of the above-mentioned specificities (A-O). The alignment is publicly available (https://3dm.bio-prodict.nl).

Table 1. Subsets created in this study.

	Subfamily (PDB)	Proteins Sequences	Crystal structures
I. NDP-sugar epimerases			
A- UDP-galactose 4-epimerase – GalE – EC 5.1.3.2	4ZRN	16918	65
B- UDP-glucuronic acid 4-epimerase – UGAE – EC 5.1.3.6	6KVC	10566	3
C- GDP-mannose 3,5-epimerase – GME (GM35E) – EC 5.1.3.18	2C5A	1959	4
D- CDP-paratose 2-epimerase – CPa2E – EC 5.1.3.10	1ORR	2391	1
E- ADP-L-glycero-D-manno-heptose-6-epimerase – AGMH6E – EC 5.1.3.20	1EQ2	5750	4
	3SXP	1823	1
II. NDP-sugar 4,6-dehydratases			
F- CDP-glucose 4,6-dehydratase – CGD – EC 4.2.1.45	1RKX	5367	2
G- dTDP-glucose 4,6-dehydratase – dTGD – EC 4.2.1.46	2HUN	1810	2
	1OC2	4001	4

	6BI4	3017	1
	1KEW	16294	4
H- GDP-mannose 4,6-dehydratase – GMD – EC 4.2.1.47	1N7H	7283	9
	2Z1M	2336	3
	1DB3	7433	2
I- UDP-GlcNAc 4,6-dehydratase (5-inverting) – UGNacD – EC 4.2.1.115	3PVZ	1888	1
	4G5H	11179	11
	4TQG	838	1
	2GN4	9157	5
J- UDP-GlcNac 5,6-dehydratase – UGNac56D – EC 4.2.1.x	3VPS	3157	1
III. NDP-keto-sugar reductases			
K- dTDP-**D**-rhamnose synthase (4-reduct.) – RMD – EC 1.1.1.133	1VL0	9446	10
	1N2S	7511	4
	3SC6	2117	1
L- GDP-**L**-fucose synthase (3,5-epim. 4-reduct.) – GMER – EC 1.1.1.271	4E5Y	1232	7
	1E6U	11906	8
M- GDP-4-keto-6-deoxy-**D**-mannose reductase – EC 1.1.1.281	2PK3	4493	2
IV. NDP-sugar acid decarboxylase			
N- UDP-glucuronate decarboxylase – UGADC (UXS) – EC 4.1.1.35	2B69	15388	4
	2BLL	2803	11
O- UDP-apiose/xylose synthase – UGADC (UAXS) – EC 4.1.1	6H0N	561	2

3. Distinctive characteristics of NDP-sugar active SDR enzymes

3.1. Conserved structural signatures in NDP-sugar active SDR enzymes.

All SDR enzymes share a conserved fold domain despite their low sequence identity (often around 15-30%) (Kallberg et al., 2002). The Rossmann-fold is a widely distributed tertiary structure that is structurally characterized by its two sets of β/α motifs with a crossover between the third and fourth strand shaping a cavity for cofactor binding (Kavanagh et al., 2008) (Fig. 2). Sequence-wise, a Rossmann-fold can be detected by the abovementioned conserved glycine-rich motif signature, which was exploited to further classify SDR enzymes into classical SDRs and extended SDRs (Kavanagh et al., 2008). Although the Rossman-fold is known to accommodate dinucleotide cofactors such as FAD and NAD(P), only the latter is observed in SDRs (Medvedevid et al., 2019).

Nonetheless, the Rossmann-fold is not the sole structural signature identified in the general three-dimensional structure of NS-SDR. A distinct structurally conserved α -helix emerges after aligning the representative structures of both NS-SDRs and other SDR enzymes (Fig. 3.A). Notwithstanding the fact that NS-SDR and other SDR enzymes share this helix downstream of the Rossmann β8-strand (*st*TyvE), both helices show a different orientation (Yellow and Red helices in Fig. 3.A). This difference of direction could serve as anchoring region for the NDP-sugar binding, as the α8-helix of the human UDP-glucose 4-epimerase (*h*GalE, PDB: 1EK6) is located next to the UDP-glucose (Fig. 3.B). Whereas the residue R236 (R239 in *h*GalE) located in the loop facing the α-helix (Fig.3.C) has been shown to be important for the binding of the NDP moiety (Sun et al., 2020), no residue from the aforementioned helix has yet been proven experimentally to be important for NDP binding. Nonetheless, molecular dynamics studies performed in GalE have revealed the movement of the loop downstream to this helix when scrutinizing the atomistic motions in both the *apo* and *holo* states, suggesting an important role of this region for NDP-sugar binding (Friedman et al., 2012).

Figure 3. A structurally conserved α-helix in NDP-sugar active SDR-enzymes. A. Structural overlay of ten SDR representatives. In red, the α8-helix of NS-SDR enzymes and in yellow, the α8-helix of non NDP-sugar active. NS-SDR enzymes: PDB: 1N7H (GDP-mannose 4,6 dehydratase), 1N2S (dTDP-glucose 4,6 dehydratase), 1GY8 (UDP-galactose 4-epimerase), 1VL0C (dTDP-4 rhamnose reductase), 2B69 (UDP-glucuronic acid decarboxylase. Non-NDP-sugar active: 1X1T (3-hydroxybutyrate dehydrogenase), 2PH3 (3-oxoacyl reductase), 2RH8 (anthocyanidin reductase), 2R6J (eugenol synthase), 3A28 (2,3-butanediol dehydrogenase). **B.** UDP-Glucose in close proximity to the α8-helix (red) in the human UDP-glucose 4-epimerase (PDB:1EK6). **C.** Close-up view of *h*GalE binding pocket.

3.2. Activity and specificity determinants of NDP-sugar enzymes share common positions in 3D structure: the heptagonal box.

3.2.1. Association of conserved regions and correlated mutations

Throughout their enzyme sequences, the four NDP-sugar reactivities (epimerase, dehydratase, reductase and decarboxylase) exhibit 8 regions that show a higher degree of conservation compared to others (Fig. 4. A). Fingerprints of (extended) SDRs such as the glycine-rich motif for cofactor binding (alignment position 4-15), the catalytic residues (Y164 and K168) and also the HxAA motif (alignment position 73-81) are clearly noticeable and are part of a conserved region.

Considering the common fold and basic mechanism of the four reactivities, it has been postulated that these enzymes have evolved from one another, with UDP-galactose 4-epimerase (GalE) as ancestor (Martinez Cuesta et al., 2014; Martínez Cuesta et al., 2015). In large protein superfamily alignments, co-evolving residues or correlated mutations are almost always functionally related (Franceus et al., 2017; Kuipers et al., 2009). This means that a mutation leading to a new function is often accompanied by a mutation at another position in order to structurally compensate for it. Broadly speaking, this natural occurrence is also true at the level of proteins in metabolic pathways or between species and is thought to be the driving force towards biological diversity (Ehrlich and Raven, 1964). In NS-SDR, these amino acids are located in the conserved regions indicating their potential role in specificity (Fig. 4. B).

Figure 4. Amino acid conservation and correlation in NDP-sugar active enzyme subfamily. A. Conservation. Conservation percentage for each alignment core position is depicted with a color gradient. Blue= 0 %, Red=100%. Positions involved in the hexagonal box model in *e*GalE are also represented. The number in asterisk represents the conserved Arg231 (*e*GalE) part of the "grey wall" of the heptagonal-box model. **B.** Correlated position network. Nodes represent the alignment positions with the node sizes indicating the number of edges. Edge colours indicate the strength of the pair-wise correlation from yellow to red.

3.2.2. From a hexagonal to a heptagonal box specificity model

Interestingly, 6 out of 8 of these conserved regions are located closely around the NDP-sugar substrate (Fig. 4. A). Surprisingly, it is reminiscent of a substrate specificity model in UDP-hexose 4-epimerase. Indeed, some of these residues are part of a model known as "the hexagonal box", which explains GalE's promiscuity towards different sugars. The hexagonal box model suggests that the substrate sugar is surrounded by six walls of amino acids in the active site (Ishiyama et al., 2004). As illustrated in Fig. 5.A depicting the hexagonal box of *Escherichia coli* GalE (*e*GalE), the yellow wall contains the conserved proton-shuttle S124 (*e*GalE), whereas the catalytic Y149 (*e*GalE) makes the cyan wall directly next to the yellow wall. K84 (*e*GalE) as part of the red wall flanks the C2-C3 site of the sugar guarding the diphosphate of the NAD$^+$ moiety, while N199 (*e*GalE) in the purple wall points towards the β-phosphorus and C1 of UDP-Glc (Beerens et al., 2015). Interestingly, the substrate spectrum of *e*GalE was altered only by mutating the so-called 'gatekeeper' Y299 composing the green wall into the smaller cysteine, found at this position in the human protein, introducing activity on UDP-GlcNAc (Beerens et al., 2015) or, vice versa, removing UDP-GlcNAc activity by introducing a large residue (Beerens et al., 2013). It is worth noting that the hexagonal box is not the sole specificity model in UDP-hexose 4-epimerases since the "297-308 belt" model (Bhatt et al., 2011) and the more recent 'two-pockets' model (Nam et al., 2019) also exist, albeit having in common most of the hexagonal box model. Indeed, these "297-308 belt", C5- and C2-pockets correspond to the (extended) walls of the hexagonal box model, more specifically to the Green, Green+Yellow and Purple walls, respectively. The hexagonal box model has recently also been used to evaluate the possible molecular evolution and functional divergence of UDP-hexose 4-epimerases, as reviewed by Fushinobu (2021).

Similarly, with the evolutionary analysis of NS-SDRs performed by Cuesta *et al.* (2014) in mind, which highlighted GalE as the common ancestor, this triggered us to evaluate whether to hexagonal box model had 'survived evolution' and could perhaps be applied or extended to all NS-SDRs. Our literature review of different NS-SDR quickly confirmed that the important residues for each activity were indeed found in the structurally corresponding regions (see next sections), namely the walls of the hexagonal box model. The hexagonal box model by Ishiyama *et al.* (2004) could thus be applied to other NS-SDRs. In order to further substantiate this hypothesis and to define the hexagonal box motifs for each specificity, we reviewed sequences from 3DM as described above. This allowed us to not only confirm that a similar hexagonal box model applies to all NS-SDRs but also to extend it. The extensions include going from walls consisting of a single residue to walls of multiple residues as well as the addition of a seventh wall (Fig. 5.B). However, a close-up view of the catalytic sites plainly exposed that the conserved regions in the sequence give rather rise to a heptagonal-shaped box with stretches of residues in close proximity to the NDP-sugar substrate (Fig.5.B and Fig. 4.A). In order to aid readers to easily locate the mentioned residues in a structure and to facilitate the comparison between the different specificities/reactivities, we will use the concept of "Heptagonal box" and keep the colors used by Ishiyama et al. (2004) to label each wall of the box. The new seventh wall containing the conserved Arg residue involved in NDP-sugar binding is assigned as the "grey" wall.

Figure 5. Specificity models. **A.** Hexagonal box. Residues shaping the hexagonal box in the Y299C *e*GalE mutant (PDB: 1LRK). **B.** Heptagonal box in our study. Each wall contains a range of conserved residues.

In the following sections, we will further discuss the conserved motifs for each specificity by linking them to literature, while also keeping an eye open for potentially new sequential and/or structural differences between specificities.

4. Functional fingerprints

4.1. NDP-sugar active epimerases

The inversion of an hydroxyl group at a specific carbon atom to generate its epimer is, in nature, catalysed by carbohydrate epimerases (CEPs) (Van Overtveldt et al., 2015). CEPs exhibit diverse substrate specificities and interconvert the orientation of a hydroxyl group at different stereocenters of their substrate using one of the five known epimerization mechanisms. Due to these varieties in CEPs, they have been classified in 14 families based on both their structural and mechanistic conservation (Van Overtveldt et al., 2015). Carbohydrate epimerase family 1 (CEP1) represents the largest epimerase collection with more than 10 sugar specificities, all being NDP-sugars and covering 5 stereocenter specificities.

In order to perform the epimerization, CEP1 family members use a transient keto intermediate mechanism (Van Overtveldt et al., 2015). The tyrosine base first abstracts a proton from the hydroxyl group at the stereocenter, which is to be epimerized. The nicotinamide ring from NAD+ is positioned in a productive distance to the stereocenter (3.0-3.7 Å) (Allard et al., 2001; Nam et al., 2019; Thoden et al., 1996) and serves as a hydride acceptor generating a keto group at the epimerization site (Allard et al., 2001; Van Overtveldt et al., 2015). Thereby, a non-chiral, more or less trigonal-planar carbon is created. Then, the keto intermediate rotates in the active site and the hydride from the NADH is retransferred to the carbon, but on the site opposite of the sugar plane. The proton from the initial deprotonation step now on the tyrosine acid is re-donated to the oxygen resulting in a hydroxyl group and regenerating the catalytic base (Van Overtveldt et al., 2015). Moreover, it is recognized that the NDP moiety induces a conformational change in GalE together with the rotation of the sugar which triggers the reduction (Allard et al., 2001). Consequently, in the case of C2- and C6-epimerases, the NDP-sugar will be oriented differently in order to allow the redox reaction and undergo the correct rotation. Although it is still not known why, the choice of NDP moiety is conserved for each stereospecificity, raising questions about different mechanisms. However, it has recently been shown that GalE or TyvE from *Thermus thermophilus (tt*GalE*)* and *Thermodesulfatator atlanticus* (*ta*TyvE) respectively, could still perform the epimerization reaction on different NDP-moieties, albeit with different degree of affinity (Rapp et al., 2020; Van Overtveldt et al., 2020). Regarding C3,5-epimerisation, the C4 is oxidized but other extra catalytic residues are needed such as C145 and K217 in *Arabidopsis thaliana* GM35E.

A particularity of epimerases compared to the three other reactivities in the NS-SDR subfamily presented herein, is that it comprises differences not only in NDP-sugar specificity but also in stereocenter specificity. Consequently, nucleotide and asymmetric carbon specificities are intertwined in correlated mutations, hence exhibiting different networks of correlated mutations (Fig.6.A).

Figure 6. Correlated mutation in NDP-sugar active epimerases. A. Correlated mutation analysis resulted in four networks. Positions that are part of the box are highlighted with a specific wall color. In dark blue, positions outside the heptagonal box. Nodes represent the alignment positions with the node sizes indicating the number of edges. Edge colours indicate the strength of the pair-wise correlation from yellow to red. **B.** Network hypothesized to be involved in immediate reduction in *e*GalE. Sequence numbering from the GALE of *Escherichia coli*.

As detailed in Fig. 6.A, the majority of the correlated mutations in the epimerase subfamily are part of the heptagonal box surrounding the substrate emphasizing the importance of each wall for the reactivity and/or specificity. The largest correlated network contains residues from each wall of the heptagonal box and the smaller networks contain at least one residue from these walls, as it is the case for the position 243, 80 and 177 in *e*GalE. Interestingly, some residues of this network were also mentioned in a study from Tiwari *et al.* where H243, Y177 and S122 (*e*GalE) have been postulated to potentially promote immediate re-oxidation of NADH in order to prevent abortive complexes (Fig. 6.B) (Tiwari et al., 2014). Also, correlated residues that are found outside the heptagonal box are often located near these walls and thus one might wonder whether these positions could assist in the correct orientation of the walls and their key residues.

It is clear that the glycine motif (L[VI]TG[GA]xGx[IVL]G) and catalytic triad [ST]x_nYx_3K is conserved in all epimerases, corresponding well with the fact that these are extended SDR. High conservation can also be observed for Phe in the glycine motif (F11 in *st*TyvE) and the Arg in the grey wall (R236 in *st*TyvE), of which the latter has been shown to make an important interaction with the substrate's di-phosphate backbone.

(landscape table)

Table 2 Heptagonal box motifs in NDP-sugar active Epimerases. The logos were generated using Weblogo.

Furthermore, high conservation of residues and motif can also be observed in the different walls of the heptagonal box (Table 2). Each asymmetric carbon specificity has a distinct conserved motif per wall, sometimes even differing in length, as is the case for the red and purple walls. It is also worth noting that the yellow wall is one of the most conserved parts of the box per specificity, showing TNK, SAA, SAC and SAA/SSS motifs for the C2-, C6-, C3,5- and C4-epimerases, respectively. This wall includes the additional catalytic Cys present in the GM35E, whereas its additional catalytic Lys (position 206) can be found in the purple wall. The Arg residue that is pointed towards the catalytic Cys and potentially assists by changing the cysteine's pKa is also highly conserved but only in the GM35E (green wall, position 305). A peculiarity of this position is that it corresponds to the so-called gatekeeper residue of the 4-epimerases and that high conservation of a large (Tyr/Phe) and small (Thr) residue can be observed for the CPa2E and AGMH6E, respectively. Similar to GalE's gatekeeper, a small residue is necessary for the larger heptose substrate, but a large residue is possible for the smaller dideoxy substrates (i.e. paratose/tyvelose). Regarding the two C4-epimerase specificities, GalE and UGAE, the main differences are found in the orange wall at the position 193 where an Asn is replaced by a Thr in UGAE and in the shorter purple wall in UGAE. The latter specificity has been proofed recently (Gevaert et al., 2020) and hence its specificity determinants still has to be determined. However, Sun *et al*. pointed out the importance of the positively charged R192 and the adjacent D194 (*Streptomyces viridosporus* UGAE), both located in the purple wall, which might interact and stabilize the negatively charged carboxylate moiety of the substrate (Sun et al., 2020). Recently, Borg et al. (2020) characterized a novel UGAE from *Bacillus cereus* and demonstrated the importance of an hydrogen bond network composed of four residues, namely T126, S127, S128 (all yellow wall) and T178 (193 in the orange wall), in the coordination of the carboxylate substrate (Borg et al., 2020; Iacovino et al., 2020).

In the CPa2E, the yellow wall contains the catalytically important Asn and Lys (personal communication). In addition, although the only two confirmed CPa2Es have a HSSM motif in the orange wall, it is the MSCM motif that is predominant. Interestingly, a Phe residue instead of the catalytic Tyr residue is also present, but only in a minor number of CPa2E subfamily representatives and their activity has never been experimentally proven. Interestingly, the newly discovered member of the CPa2E subfamily, *ta*TyvE was shown to have a promiscuous activity towards NDP-glucose (Rapp et al., 2020). Mutational analysis will have to be performed to associate any residue of the different wall's fragments to glucose accommodation.

Secondary structural elements can also influence internal motions or induce conformational changes important for an enzyme activity (Boehr et al., 2018). Although it has never been demonstrated to explain the difference in asymmetric carbon preference, each group of enzymes exhibit variations in the structure (Fig. 7). While every template shows a similar overall fold, some differences stand out, such as the size of the terminal α-helix in GM35E that was longer than in the other subfamilies. Similarly, GM35E possesses an extra loop downstream of this helix that goes up to the NAD domain. Also, the "red" wall of the heptagonal box in GM35E exhibits extra residues. Therefore, it elongates the loop and significantly diminishes the active site cavity and thus the torsional mobility of the sugar moiety could be restricted. Both C6-epimerase subfamilies have an α-helix within the "purple" wall in contrast to the other enzymes that only have a loop. Interestingly, CPa2E also has a conserved extra β-sheet (β6 and β7) that might interact with the green wall (R303) and grey wall (V235).

Figure 7. Structural differences in epimerases. Structural differences (variable regions in GM35E and AGMH6E and conserved β-sheet in 1ORR not present in other epimerase structures) are depicted in green.

4.2. NDP-sugar active dehydratases

Deoxysugars ($C_6H_{12}O_5$), dideoxysugars ($C_6H_{12}O_4$) and their derivatives are functionally important carbohydrates that are found in a myriad of virulence factors (Islam et al., 2019). Their biosynthesis necessitates a 4,6- or 5,6-dehydration of a hexose nucleotide precursor, a step that has been extensively scrutinized in the past decades as it could pave the way for targeted drugs development (Allard et al., 2002).

Similarly, to many other NS-SDR representatives, the reaction is initiated with an oxidation at the 4-hydroxyl group of the sugar and subsequent hydride transfer to the NAD(P)$^+$ cofactor. This step is followed by the elimination of water from C5 and C6 generating a 4-keto-5,6-glycosene intermediate. Eventually, the removed proton at C5 will be transferred back to its original position (retaining activity) or at the opposite location (inverting activity) with hydride transfer from NAD(P)H to the C6 forming the NDP-4-keto 6-deoxyhexose (Ishiyama et al., 2006; Li et al., 2015; Webb, 2004). Enzymes known to employ a retaining activity with the above catalytic mechanism are GDP-mannose 4,6-dehydratase, dTDP-glucose 4,6-dehydratase, CDP-glucose 4,6-dehydratases and UDP-GlcNAc 4,6-dehydratase. However, for the latter, the majority of enzymes from this subfamily are known to possess an inverting activity. We talk about 5,6-dehydratase activity when the proton and hydride are transferred back to the C4 generating an NDP-6-deoxy-5,6-glycosene, which were discovered much more recently (Wyszynski et al., 2012).

The dehydratase's heptagonal box also contains distinct motifs for each nucleotide specificity (Table 3). The "purple" wall has a much shorter loop compared to NDP-sugar epimerases with only the positions 206 and 207 (e.g. 174 and 175 in UGNacD from *Helicobacter pylori*) alignable with the epimerase loop. However, within the dehydratases, this loop is structurally similar, but with different motifs from each specificity. Despite variability in the motifs, the last position is mainly a short aliphatic residue (Val or Leu). The same characteristics are found in the purple wall's last position in GalE, GM35E and AGMH6E. The three latter enzymes have in common with dehydratase, the positioning of the 4'-hydroxyl of the sugar at a productive distance to the catalytic Tyr. Given that the only enzyme known to potentially have its 2-hydroxyl oxidized is CPa2E, bearing a conserved bulky aromatic Trp at the position 207. One might wonder the importance of this position for the productive positioning of the substrate.

Equivalently to epimerases, the dehydratase's yellow walls are also conserved per specificity, which is expected given the presence of the catalytic residues.

It is also worth noting that the catalytic cysteine in GDP-mannose 3,5-epimerase is located at the same position in the structure (126 in *st*TyvE) as the catalytic Glu and Lys, downstream the Ser/Thr from the catalytic triad, in the yellow wall of the heptagonal box. The yellow wall in epimerases could serve as benchmark to predict enzyme function since each specificity had a distinct motif. Indeed, on top of the [ST] signature, the two following residues dictates the epimerase specificity. In dehydratases, however, one needs to consider this wall but additionally look beyond it. Indeed, the orange wall shows highly conserved and distinct motifs per specificity and might thus be an additional motif to consider for specificity predictions.

Like in the epimerases, also in the dehydratase there are other representative residues than the established catalytic Tyr in the catalytic dyad (Yx_3K). Indeed, there is rather high conservation of methionine (Mx_3K), however, this Mx_3K is mostly present in the UDP-GlcNAc 4,6-dehydratase, with the methionine clearly showing up in the fingerprints (Table 3). An alternative residue that is found at this position is a leucine, albeit in a much lower amount. Recently, the mechanism behind the long-discussed Mx_3K dyad (Creuzenet et al., 2002; Miyafusa et al., 2013b, 2013a) present in the NDP-GlcNAc 4,6-dehydratase was confirmed (Riegert et al., 2017). Structural and biochemical investigation of the *Campylobacter jejuni* UDP-GlcNAc 4,6-dehydratase (PglF) revealed a new catalytic mechanism for this SDR superfamily member that excludes the need of the otherwise conserved/catalytic Tyr. In PglF, the combination of T395 and D396 removes the proton from the C-4 hydroxyl group (C4-oxidation) and the neighboring K397 is involved dehydration and reprotonation (Riegert et al., 2017). Once more, the three involved residues (TDK) are located in the yellow wall and clearly pop out in the fingerprints, showing very high conservation for the TDK (Table 3).

In the group clustered with the NDP-hexosamine 5,6-dehydratase (TunA , PDB code 3VPS), the catalytically important Cys-Glu residues of TunA (Wyszynski et al., 2012) seem to be absent, however, a closer look shows it is present but only in a smaller quantity (Table 3). It thus probably got outweighed by the overrepresented 4,6-dehydratase that clustered together with the 5,6-dehydratase. Based on the residues present (Ser-Glu and Asp-Glu), these could be GDP-mannose 4,6-dehydratase (GMD) and/or dTDP-glucose 4,6-dehydratase (dTGD), respectively. This might hint towards a closer evolutionary connection of the NDP-hexosamine 5,6-dehydratase with GMD and dTGD than with the 4,6-dehydratase active on NDP-hexosamine (and CDP-Glc) but remains to be scrutinized before making definitive statements.

(landscape table)
Table 3. Heptagonal box motifs in NDP-sugar active Dehydratases. The logos were generated using Weblogo.

4.3. NDP- keto sugar active reductases

In the palette of enzymatic reactions leading to the formation of deoxysugars, reduction is a crucial step allowing the generation of hydroxyl groups from an NDP-4-keto-sugar intermediate (Martinez et al., 2012). Therefore, considerable attention has been devoted to NDP-sugar reductases linked with the biosynthesis of pathogenicity factors. The most extensively studied pathway are the ones producing the key intermediates dTDP-L-rhamnose, GDP-D-rhamnose and GDP-L-fucose, dTDP-D-fucose. These deoxysugars are obtained after a stereoselective reduction of an NDP-4-keto-6-deoxy sugar precursor. In order to switch from D- to L-conformation, a 3,5-epimerization is required and is often catalyzed by a bifunctional 3,5-epimerase/reductase enzyme. The major difference between a strict NDP-sugar 3,5-epimerase and a bifunctional 3,5-epimerase/reductase is that the latter requires a keto-functionality at C4 on a deoxy sugar (product of the abovementioned dehydratases), whereas the former acts on a fully hydroxylated substrate and performs the initial oxidation at C4 by itself. However, the catalytic mechanism for epimerization is shared between both enzymes. Three subfamilies are part of NDP-keto sugar active reductases, one that forms GDP-D-rhamnose by ketoreduction, one that has bifunctional 3,5-epimerase/reductase activity leading to GDP-L-fucose and dTDP-L-rhamnose and the last one forming dTDP-D-fucose and dTDP-L-rhamnose (after prior 3,5-epimerization by another enzyme (Dong et al., 2003)).

The three specificities in reductases display a highly conserved heptagonal-box wall motif (Table 4.A). Strikingly, the acidic/basic residues involved in 3,5-epimerization in bifunctional enzymes are located at the same positions as in the NDP-sugar 3,5-epimerase, i.e. 206 and 126, respectively. Both enzymes share a common Cys in the yellow wall (core position 126), whereas the acid located in the purple wall has been replaced by a His (Lys in GM35E). This confirms the importance of both positions and residues for the 3,5-epimerization as much as the catalytic Yx$_3$K dyad located in the cyan wall.
As with the other reactivities, the majority of conserved residues are part of the heptagonal box emphasizing once more the validity of this specificity and activity model in NS-SDRs, including the reductases. Position 206 is a His (e.g., H158 in RfbD from *Clostridium acetobutylum*) when the enzyme exhibits a bifunctional activity but a Phe in the GDP-4-keto-6-deoxy-D-mannose reductase subfamily and Asn in the dTDP-D-rhamnose synthase subfamily when only showing a reductase activity. Similarly, the position 305, which corresponds to the gatekeeper residue in GalE and the highly important Arg in GM35E, is a Lys/Arg in the bifunctional reductases. Similar as in GM35E, this highly conserved Lys most likely assists the change of the catalytic cysteine’s pKa. In the reductases without epimerase activity this position is occupied by a highly conserved Ser in the dTDP-D-rhamnose synthase (4-reduct) subfamily and a Leu in GDP-4-keto-6-deoxy-D-mannose reductase subfamily, incapable of altering the pKa of neighboring residues.
Although in the full NS-SDR enzymes, the third amino acid of the orange wall (position 193 in *st*TyvE) is mostly an Asn, it is a conserved Trp residue for the dTDP-D-rhamnose synthase (4-reduct) subfamily. This Trp was shown to interact with the equatorial O$_3$’ of the sugar substrate (W153 in *Salmonella enterica* RmlD) and abolished the enzyme activity when mutated to an Ala (Blankenfeldt et al., 2002).

(landscape table)
Table 4. Heptagonal box motifs. A. Heptagonal box motifs in NDP-keto sugar active reductases. **B.** Heptagonal box motifs in NDP-sugar active decarboxylases. The logos were generated using Weblogo.

4.4. NDP-sugar active decarboxylases

The mechanistic complexity and diversity of NS-SDR enzymes becomes even more evident when including NDP-sugar active decarboxylases. Indeed, starting from the same substrate (UDP-glucuronic acid, UDP-GlcA), no less than three products are obtained, namely UDP-xylose, UDP-4-keto-xylose and UDP-apiose (Savino et al., 2019). An important sidenote here is that a fourth product (i.e. UDP-GalA) is obtained with the abovementioned UDP-GlcA 4-epimerases. The former decarboxylase product is a sugar donor of xylose, a major constituent of plant xylan polymers, and obtained by UDP-xylose synthase (UXS) (Bar-Peled et al., 2001). In humans, UDP-xylose initiates the glycosaminoglycan synthesis of proteoglycans, critical for cell signaling (Eixelsberger et al., 2012). UDP-4-keto-xylose is produced through oxidation of UDP-glucuronic by a bifunctional UDP-glucuronic acid decarboxylase/formyltransferase (ArnA). The latter from which only the C-terminal domain belongs to the SDR family, will catalyze another reaction step in the pathway forming a 4-amino-4-deoxy-L-arabinose (L-Ara4N), essential for the resistance of the antibiotic polymyxin (Williams et al., 2005). Given the fact that their homologous domain allows a similar catalytic reaction, they have been clustered together. On the other hand, UDP-apiose, a branched-chain pentose involved in the cell wall integrity of plants, is synthesized via a UDP-apiose/UDP-xylose synthase (UAXS), which unprecedently catalyzes the decarboxylation of the UDP-glucuronic acid coupled with pyranosyl-to-furanosyl sugar ring contraction (Savino et al., 2019).

The SDR catalytic triad (S/Tx_nYx_3K) located in the yellow and cyan walls is logically conserved (Table 4.B). The amino acids known to be characteristics of decarboxylases activity (Savino et al., 2019) are also highly conserved in the entire decarboxylase subfamily such as Y105 , E141 and R341 (numbering of UAXS from *Arabidopsis thaliana*). It is worth noting that these residues are located near the red wall and in the yellow and green wall, respectively. The differences that distinguish the UDP-apiose-synthesizing enzyme UAXS from the other decarboxylases are the presence of C100 and C140 (UAXS from *Arabidopsis thaliana*) corresponding to the positions 86 and 126 of the red and yellow wall, respectively. These residues were shown to be involved in the sugar ring contraction (Savino et al., 2019).

Interestingly, a structural variable region elongating the α-helix upstream the cyan wall is only visible in decarboxylases and bears the trademark position; the R182 (UAXS from *Arabidopsis thaliana)* (Fig. 8.A). Similarly, the purple wall loop before the positions 206 and 207 is longer in UAXS and ArnA, and shorter in UXS compared to *st*TyvE (Fig. 8. B).

Figure 8. Variable structural region in UAXS. **A.** Additional loop upstream the cyan wall. The region specific for decarboxylases (variable region in our alignment) is depicted in deepteal cyan. In red, the catalytic tyrosine and lysine of UAXS (PDB: 6H0P). **B.** Purple wall in decarboxylases and C2-epimerase. The purple wall of *st*TyvE is depicted in green, of UAXS in red, UXS in orange and ArnA in blue.

Concluding remarks

Fingerprints for the majority of specificities and reactivities within the NS-SDR enzyme subfamily were exposed in our sequence review. Seven regions surrounding the substrate nucleotide-sugar in a heptagonal-shaped manner are highly structurally and sequentially conserved. These heptagonal-box motifs contain important known catalytic residues and substrate specificity determinants, but few studies have made the link between them. While the “cyan wall” of the box possessed the characteristic signature of the SDR catalytic dyad (Yx_3K), the other fragments around the substrate are not less important. The highly conserved Ser/Thr in the “yellow wall” supports the choice of talking about a catalytic triad ([S/T]x_nYx_3K) for the NS-SDR enzymes. In contrast to UDP-galactose 4-epimerase (GalE) and dTDP-D-rhamnose synthase (RMD), that perform only oxidation/reduction, some NS-SDRs excecute more complex catalysis and the extra catalytic residues needed for this were almost always located in the “yellow wall” at the position 125 and/or 126 or at the position 206 in the “purple wall”. The gatekeeper residue located in the “green wall” – known to influence the sugar specificity of GalE – is structurally at the same position as the conserved Arg involved in the GDP-mannose 3,5-epimerase (GM35E) and important for decarboxylation or the Lys for GDP-L-fucose synthase (GMER). Furthermore, it shows high conservation in most other specificities too, hinting to also have an important role there. Although little is known about the red, orange and grey walls, the conserved motifs highlight their importance in the different subfamilies and pinpoint highly conserved – and thus potentially important – residues that could now be investigated via mutagenesis. On top of the heptagonal box, conserved structural differences for the enzymes of each reactivity can also be observed. Since catalysis depends on three important factors; (i) NDP-moiety, (ii) the correct orientation of the sugar and (iii) the positioning of the NAD cofactor; a static analysis of structures might not be enough to identify the determinants behind the four reactivities. Complementary information could be obtained through more motion analysis using molecular dynamic (MD) simulations.

Nonetheless, the data collected and our insights into the catalytic mechanism of NS-SDR enzymes have allowed us to identify fingerprint motifs that bring us a step closer to the (more) complete understanding of the structure-function relationships in NS-SDR enzymes. Furthermore, by leveraging these sequence and structural information, it could serve as the foundation for future enzyme engineering projects aiming to switch specificities and/or reactivities in these SDRs. As engineering strategies, one might explore the possibilities of introducing the most conserved residues from each wall, swapping heptagonal box motifs and/or inserting conserved structural elements in those NS-SDRs. An ultimate challenge lies in the design of new-to-nature NS-SDR biocatalysts that can be applied for the enzymatic production of novel "designer" carbohydrates and/or glycosides. For instance, the multiple stereocenters present in carbohydrates generate a huge structural diversity that is often linked with divergent biological properties. The identification of specificity fingerprints in the epimerase subfamily could pave the way for future engineering effort towards stereocenter specificity switches making conceivably possible the synthesis of virtually all rare sugars and derivatives thereof.

Author Contributions

Conceptualization: M.D.C., O.G., S.V.O., J.L., H.J.J, T.D., and K.B.; Methodology: M.D.C.; Software: J.L., H.J.J.; Validation: O.G., S.V.O., J.L., H.J.J., T.D, and K.B.; Formal analysis: M.D.C.; Investigation: M.D.C.; Data curation: M.D.C.; Writing—original draft preparation: M.D.C.; Writing—Review and Editing: T.D., S.V.O., O.G., J.L., H.J.J., and K.B.; Visualization: M.D.C.; Supervision: T.D. and K.B.; Project administration: M.D.C. All authors have read and agree to the published version of the manuscript.

Acknowledgment

This research was funded by the Fund for Scientific Research Flanders (FWO-Vlaanderen) via a doctoral scholarship for M.D.C. (grant n° 1S03020N), O.G. (grant n° 1S16817N), S.V.O. (grant n° FWO.3F0.2016.0056.01) as well as the EpiSwitch and DeoxyBioCat projects (grant n° G0F3417N and G0A7520N, respectively).

References

Allard, S.T.M., Beis, K., Giraud, M.F., Hegeman, A.D., Gross, J.W., Wilmouth, R.C., Whitfield, C., Graninger, M., Messner, P., Allen, A.G., Maskell, D.J., Naismith, J.H., 2002. Toward a structural understanding of the dehydratase mechanism. Structure 10, 81–92. doi:10.1016/S0969-2126(01)00694-3

Allard, S.T.M., Giraud, M.F., Naismith, J.H., 2001. Epimerases: Structure, function and mechanism. Cell. Mol. Life Sci. 58, 1650–1665. doi:10.1007/PL00000803

Bar-Peled, M., Griffith, C.L., Doering, T.L., 2001. Functional cloning and characterization of a UDP-glucuronic acid decarboxylase: The pathogenic fungus Cryptococcus neoformans elucidates UDP-xylose synthesis. Proc. Natl. Acad. Sci. U. S. A. 98, 12003–12008. doi:10.1073/pnas.211229198

Beerens, K., Desmet, T., Soetaert, W., 2012. Enzymes for the biocatalytic production of rare sugars. J. Ind. Microbiol. Biotechnol. 39, 823–834. doi:10.1007/s10295-012-1089-x

Beerens, K., Soetaert, W., Desmet, T., 2015. UDP-hexose 4-epimerases: a view on structure, mechanism and substrate specificity. Carbohydr. Res. 414, 8–14. doi:10.1016/j.carres.2015.06.006

Beerens, K., Soetaert, W., Desmet, T., 2013. Characterization and mutational analysis of the UDP-Glc (NAc) 4-epimerase from Marinithermus hydrothermalis. Appl. Microbiol. Biotechnol. 97, 7733–7740. doi:10.1007/s00253-012-4635-6

Beerens, K., Van Overtveldt, S., Desmet, T., 2017. The 'Epimerring' highlights the potential of carbohydrate epimerases for rare sugar production. Biocatal. Biotransformation 0, TBD. doi:10.1080/10242422.2017.1306738

Blankenfeldt, W., Kerr, I.D., Giraud, M.F., McMiken, H.J., Leonard, G., Whitfield, C., Messner, P., Graninger, M., Naismith, J.H., 2002. Variation on a theme of SDR: dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) shows a new Mg2+-dependent dimerization mode.

Structure 10, 773–786. doi:10.1016/S0969-2126(02)00770-0

Boehr, D.D., D'Amico, R.N., O'Rourke, K.F., 2018. Engineered control of enzyme structural dynamics and function. Protein Sci. 27, 825–838. doi:10.1002/pro.3379

Borg, A.J.E., Beerens, K., Pfeiffer, M., Desmet, T., Nidetzky, B., 2021. Stereo-electronic control of reaction selectivity in short-chain dehydrogenases: Decarboxylation, epimerization, and dehydration. Curr. Opin. Chem. Biol. doi:10.1016/j.cbpa.2020.09.010

Borg, A.J.E., Dennig, A., Weber, H., Nidetzky, B., 2020. Mechanistic characterization of UDP-glucuronic acid 4-epimerase. FEBS J. doi:10.1111/febs.15478

Creuzenet, C., Urbanic, R. V., Lam, J.S., 2002. Structure-Function Studies of Two Novel UDP-GlcNAc C6 Dehydratases/C4 Reductases: Variation from the SYK dogma. J. Biol. Chem. 26769-26778. doi:10.1074/jbc.M202882200

Dong, C., Beis, K., Giraud, M.F., Blankenfeldt, W., Allard, S., Major, L.L., Kerr, I.D., Whitfield, C., Naismith, J.H., 2003. A structural perspective on the enzymes that convert dTDP-D-glucose into dTDP-L-rhamnose. Biochem. Soc. Trans. 31, 532–536. doi:10.1042/BST0310532

Dudek, C.A., Dannheim, H., Schomburg, D., 2017. BrEPS 2.0: Optimization of sequence pattern prediction for enzyme annotation. PLoS One 12, 1–12. doi:10.1371/journal.pone.0182216

Ehrlich, P.R., Raven, P.H., 1964. Butterflies and Plants: A Study in Coevolution. Evolution (N. Y). 18, 586. doi:10.2307/2406212

Eixelsberger, T., Sykora, S., Egger, S., Brunsteiner, M., Kavanagh, K.L., Oppermann, U., Brecker, L., Nidetzky, B., 2012. Structure and mechanism of human UDP-xylose synthase: evidence for a promoting role of sugar ring distortion in a three-step catalytic conversion of UDP-glucuronic acid. J. Biol. Chem. 287, 31349–58. doi:10.1074/jbc.M112.386706

Elshahawi, S.I., Shaaban, K.A., Kharel, M.K., Thorson, J.S., 2015. A comprehensive review of glycosylated bacterial natural products. Chem. Soc. Rev. 44, 7591–7697. doi:10.1039/c4cs00426d

Franceus, J., Verhaeghe, T., Desmet, T., 2017. Correlated positions in protein evolution and engineering. J. Ind. Microbiol. Biotechnol. 44, 687–695. doi:10.1007/s10295-016-1811-1

Friedman, A.J., Durrant, J.D., Pierce, L.C.T., Mccorvie, T.J., Timson, D.J., Mccammon, J.A., 2012. The Molecular Dynamics of Trypanosoma brucei UDP-Galactose 4′-Epimerase: A Drug Target for African Sleeping Sickness. Chem. Biol. Drug Des. 80, 173–181. doi:10.1111/j.1747-0285.2012.01392.x

Fushinobu, S., 2021. Molecular evolution and functional divergence of UDP-hexose 4-epimerases. Curr. Opin. Chem. Biol. doi:10.1016/j.cbpa.2020.09.007

Gevaert, O., Van Overtveldt, S., Beerens, K., Desmet, T., 2019. Characterization of the First Bacterial and Thermostable GDP-Mannose 3,5-Epimerase. Int. J. Mol. Sci. 20. doi:10.3390/ijms20143530

Gevaert, O., Van Overtveldt, S., Da Costa, M., Beerens, K., Desmet, T., 2020. Novel insights into the existence of the putative udp-glucuronate 5-epimerase specificity. Catalysts 10, 1–11. doi:10.3390/catal10020222

Gräff, M., Buchholz, P.C.F., Stockinger, P., Bommarius, B., Bommarius, A.S., Pleiss, J., 2019. The Short-chain Dehydrogenase/Reductase Engineering Database (SDRED): A classification and analysis system for a highly diverse enzyme family. Proteins Struct. Funct. Bioinforma. 87, 443–451. doi:10.1002/prot.25666

Iacovino, L.G., Savino, S., Borg, A.J.E., Binda, C., Nidetzky, B., Mattevi, A., 2020. Crystallographic snapshots of UDP-glucuronic acid 4-epimerase ligand binding, rotation, and reduction. J. Biol. Chem. 295, 12461–12473. doi:10.1074/jbc.RA120.014692

Ishiyama, N., Creuzenet, C., Lam, J.S., Berghuis, A.M., 2004. Crystal structure of WbpP, a genuine UDP-N-acetylglucosamine 4-epimerase from Pseudomonas aeruginosa: Substrate specificity in UDP-hexose 4-epimerases. J. Biol. Chem. 279, 22635–22642. doi:10.1074/jbc.M401642200

Ishiyama, N., Creuzenet, C., Miller, W.L., Demendi, M., Anderson, E.M., Harauz, G., Lam, J.S., Berghuis, A.M., 2006. Structural studies of FlaA1 from Helicobacter pylori reveal the mechanism for inverting 4,6-dehydratase activity. J. Biol. Chem. 281, 24489–24495. doi:10.1074/jbc.M602393200

Islam, R., Brown, S., Taheri, A., Dumenyo, C.K., 2019. The gene encoding nad-dependent epimerase/dehydratase, wcag, affects cell surface properties, virulence, and extracellular enzyme production in the soft rot phytopathogen, pectobacterium carotovorum. Microorganisms 7. doi:10.3390/microorganisms7060172

Jacobson, M.P., Kalyanaraman, C., Zhao, S., Tian, B., 2014. Leveraging structure for enzyme function prediction: Methods, opportunities, and challenges. Trends Biochem. Sci. doi:10.1016/j.tibs.2014.05.006

Junker, S., Roldan, R., Joosten, H.-J., Clapés, P., Fessner, W.-D., 2018. Complete Switch of Reaction Specificity of an Aldolase by Directed Evolution In Vitro: Synthesis of Generic Aliphatic Aldol Products. Angew. Chemie Int. Ed. 57, 10153–10157. doi:10.1002/anie.201804831

Kallberg, Y., Oppermann, U., Jörnvall, H., Persson, B., 2002. Short-chain dehydrogenases/reductases (SDRs). Coenzyme-based functional assignments in completed genomes. Eur. J. Biochem. 269, 4409–4417. doi:10.1046/j.1432-1033.2002.03130.x

Kavanagh, K.L., Jörnvall, H., Persson, B., Oppermann, U., 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: Functional and structural diversity within a family of metabolic and regulatory enzymes. Cell. Mol. Life Sci. 65, 3895–3906. doi:10.1007/s00018-008-8588-y

Kourist, R., Jochens, H., Bartsch, S., Kuipers, R., Padhi, S.K., Gall, M., Böttcher, D., Joosten, H.J., Bornscheuer, U.T., 2010. The α/β-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. ChemBioChem 11, 1635–1643. doi:10.1002/cbic.201000213

Kuipers, R.K., Joosten, H.J., Van Berkel, W.J.H., Leferink, N.G.H., Rooijen, E., Ittmann, E., Van Zimmeren, F., Jochens, H., Bornscheuer, U., Vriend, G., Martins Dos Santos, V.A.P., Schaap, P.J., 2010. 3DM: Systematic analysis of heterogeneous superfamily data to discover protein functionalities. Proteins Struct. Funct. Bioinforma. 78, 2101–2113. doi:10.1002/prot.22725

Kuipers, R.K.P., Joosten, H.J., Verwiel, E., Paans, S., Akerboom, J., Van Der Oost, J., Leferink, N.G.H., Van Berkel, W.J.H., Vriend, G., Schaap, P.J., 2009. Correlated mutation analyses on super-family alignments reveal functionally important residues. Proteins Struct. Funct. Bioinforma. 76, 608–616. doi:10.1002/prot.22374

Lanfranchi, E., Pavkov-Keller, T., Koehler, E.M., Diepold, M., Steiner, K., Darnhofer, B., Hartler, J., Bergh, T. Van Den, Joosten, H.J., Gruber-Khadjawi, M., Thallinger, G.G., Birner-Gruenberger, R., Gruber, K., Winkler, M., Glieder, A., 2017. Enzyme discovery beyond homology: A unique hydroxynitrile lyase in the Bet v1 superfamily. Sci. Rep. 7, 1–15. doi:10.1038/srep46738

Li, Z., Hwang, S., Ericson, J., Bowler, K., Bar-Peled, M., 2015. Pen and pal are nucleotide-sugar dehydratases that convert UDP-GlcNAc to UDP-6-deoxy-D-GlcNAc-5,6-ene and then to UDP-4-keto-6-deoxy-L-AltNAc for CMP-pseudaminic acid synthesis in Bacillus thuringiensis. J. Biol. Chem. 290, 691–704. doi:10.1074/jbc.M114.612747

Liu, Y., Thoden, J.B., Kim, J., Berger, E., Gulick, A.M., Ruzicka, F.J., Holden, H.M., Frey, P.A., 1997. Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from Escherichia coli. Biochemistry 36, 10675–84. doi:10.1021/bi970430a

Loewenstein, Y., Raimondo, D., Redfern, O.C., Watson, J., Frishman, D., Linial, M., Orengo, C., Thornton, J., Tramontano, A., 2009. Protein function annotation by homology-based inference. Genome Biol. 10, 207. doi:10.1186/gb-2009-10-2-207

Martinez Cuesta, S., Furnham, N., Rahman, S.A., Sillitoe, I., Thornton, J.M., 2014. The evolution of enzyme function in the isomerases. Curr. Opin. Struct. Biol. doi:10.1016/j.sbi.2014.06.002

Martínez Cuesta, S., Rahman, S.A., Furnham, N., Thornton, J.M., 2015. The Classification and Evolution of Enzyme Function. Biophys. J. 109, 1082–1086. doi:10.1016/j.bpj.2015.07.020

Martinez, V., Ingwers, M., Smith, J., Glushka, J., Yang, T., Bar-Peled, M., 2012. Biosynthesis of UDP-4-keto-6-deoxyglucose and UDP-rhamnose in pathogenic fungi Magnaporthe grisea and Botryotinia fuckeliana. J. Biol. Chem. 287, 879–892. doi:10.1074/jbc.M111.287367

Medvedevid, K.E., Kinch, L.N., Schaeffer, R.D., Grishin, N. V., 2019. Functional analysis of rossmann-like domains reveals convergent evolution of topology and reaction pathways, PLoS Computational Biology. doi:10.1371/journal.pcbi.1007569

Miyafusa, T., Caaveiro, J.M.M., Tanaka, Y., Tanner, M.E., Tsumoto, K., 2013a. Crystal structure of the capsular polysaccharide synthesizing protein CapE of Staphylococcus aureus. Biosci. Rep. 33, 463−474.

Miyafusa, T., Caaveiro, J.M.M., Tanaka, Y., Tsumoto, K., 2013b. Dynamic elements govern the catalytic activity of CapE, a capsular polysaccharide-synthesizing enzyme from Staphylococcus aureus. FEBS Lett. 587, Issue, 3824–3830. doi:10.1016/j.febslet.2013.10.009

Nam, Y.W., Nishimoto, M., Arakawa, T., Kitaoka, M., Fushinobu, S., 2019. Structural basis for broad substrate specificity of UDP-glucose 4-epimerase in the human milk oligosaccharide catabolic pathway of Bifidobacterium longum. Sci. Rep. 9, 11081. doi:10.1038/s41598-019-47591-w

Rapp, C., van Overtveldt, S., Beerens, K., Weber, H., Nidetzky, B., 2020. Expanding the enzyme repertoire for sugar nucleotide epimerization: the CDP-tyvelose 2-epimerase from Thermodesulfatator atlanticus for Downloaded from. doi:10.1128/AEM.02131-20

Riegert, A.S., Thoden, J.B., Schoenhofen, I.C., Watson, D.C., Young, N.M., Tipton, P.A., Holden, H.M., 2017. Structural and Biochemical

Investigation of PglF from Campylobacter jejuni Reveals a New Mechanism for a Member of the Short Chain Dehydrogenase/Reductase Superfamily. Biochemistry 56, 6030–6040. doi:10.1021/acs.biochem.7b00910

Savino, S., Borg, A.J.E., Dennig, A., Pfeiffer, M., De Giorgi, F., Weber, H., Dubey, K.D., Rovira, C., Mattevi, A., Nidetzky, B., 2019. Deciphering the enzymatic mechanism of sugar ring contraction in UDP-apiose biosynthesis. Nat. Catal. 2, 1115–1123. doi:10.1038/s41929-019-0382-8

Schnoes, A.M., Brown, S.D., Dodevski, I., Babbitt, P.C., 2009. Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. PLoS Comput. Biol. 5, e1000605. doi:10.1371/journal.pcbi.1000605

Steffen-Munsberg, F., Vickers, C., Kohls, H., Land, H., Mallin, H., Nobili, A., Skalden, L., van den Bergh, T., Joosten, H.-J., Berglund, P., Höhne, M., Bornscheuer, U.T., 2015. Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. Biotechnol. Adv. 33, 566–604. doi:10.1016/j.biotechadv.2014.12.012

Sun, H., Ko, T.-P., Liu, Wenting, Liu, Weidong, Zheng, Y., Chen, C.-C., Guo, R.-T., 2020. Structure of an antibiotic-synthesizing UDP-glucuronate 4-epimerase MoeE5 in complex with substrate. Biochem. Biophys. Res. Commun. 521, 31–36. doi:10.1016/j.bbrc.2019.10.035

Thibodeaux, C.J., Melançon, C.E., Liu, H., Liu, H., 2008. Natural-product sugar biosynthesis and enzymatic glycodiversification. Angew. Chem. Int. Ed. Engl. 47, 9814–59. doi:10.1002/anie.200801204

Thibodeaux, C.J., Melançon, C.E., Liu, H.W., 2007. Unusual sugar biosynthesis and natural product glycodiversification. Nature 446, 1008–1016. doi:10.1038/nature05814

Thoden, J.B., Frey, P.A., Holden, H.M., 1996. Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from Escherichia coli: Implications for the catalytic mechanism. Biochemistry 35, 5137–5144. doi:10.1021/bi9601114

Thorson, J., Hosted Jr., T., Jiang, J., Biggins, J., Ahlert, J., 2005. Natures Carbohydrate Chemists The Enzymatic Glycosylation of Bioactive Bacterial Metabolites. Curr. Org. Chem. 5, 139–167. doi:10.2174/1385272013375706

Tiwari, P., Singh, N., Dixit, A., Choudhury, D., 2014. Multivariate sequence analysis reveals additional function impacting residues in the SDR superfamily. Proteins Struct. Funct. Bioinforma. 82, 2842–2856. doi:10.1002/prot.24647

van den Bergh, T., Tamo, G., Nobili, A., Tao, Y., Tan, T., Bornscheuer, U.T., Kuipers, R.K.P., Vroling, B., de Jong, R.M., Subramanian, K., Schaap, P.J., Desmet, T., Nidetzky, B., Vriend, G., Joosten, H.-J., 2017. CorNet: Assigning function to networks of co-evolving residues by automated literature mining. PLoS One 12, e0176427. doi:10.1371/journal.pone.0176427

Van Overtveldt, S., Da Costa, M., Gevaert, O., Joosten, H., Beerens, K., Desmet, T., 2020. Determinants of the Nucleotide Specificity in the Carbohydrate Epimerase Family 1. Biotechnol. J. 2000132. doi:10.1002/biot.202000132

Van Overtveldt, S., Verhaeghe, T., Joosten, H.-J., van den Bergh, T., Beerens, K., Desmet, T., 2015. A structural classification of carbohydrate epimerases: From mechanistic insights to practical applications. Biotechnol. Adv. 33, 1814–1828. doi:10.1016/j.biotechadv.2015.10.010

Varki, Ajit; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M., E., 2015. Essentials of Glycobiology, 3rd edition, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press.

Webb, N.A., 2004. Crystal structure of a tetrameric GDP-D-mannose 4,6-dehydratase from a bacterial GDP-D-rhamnose biosynthetic pathway. Protein Sci. 13, 529–539. doi:10.1110/ps.03393904

Williams, G.J., Breazeale, S.D., Raetz, C.R.H., Naismith, J.H., 2005. Structure and function of both domains of ArnA, a dual function decarboxylase and a formyltransferase, involved in 4-amino-4-deoxy-L-arabinose biosynthesis. J. Biol. Chem. 280, 23000–23008. doi:10.1074/jbc.M501534200

Wyszynski, F.J., Lee, S.S., Yabe, T., Wang, H., Gomez-Escribano, J.P., Bibb, M.J., Lee, S.J., Davies, G.J., Davis, B.G., 2012. Biosynthesis of the tunicamycin antibiotics proceeds via unique exo-glycal intermediates. Nat. Chem. 4, 539–546. doi:10.1038/nchem.1351

	Subfamily (PDB)	Proteins Sequences	Crystal structures
V. NDP-sugar epimerases			
P- UDP-galactose 4-epimerase – GalE – EC 5.1.3.2	4ZRN	16918	65
Q- UDP-glucuronic acid 4-epimerase – UGAE – EC 5.1.3.6	6KVC	10566	3
R- GDP-mannose 3,5-epimerase – GME (GM35E) – EC 5.1.3.18	2C5A	1959	4
S- CDP-paratose 2-epimerase – CPa2E – EC 5.1.3.10	1ORR	2391	1
T- ADP-L-glycero-D-manno-heptose-6-epimerase – AGMH6E – EC 5.1.3.20	1EQ2	5750	4
	3SXP	1823	1
VI. NDP-sugar 4,6-dehydratases			
U- CDP-glucose 4,6-dehydratase – CGD – EC 4.2.1.45	1RKX	5367	2
V- dTDP-glucose 4,6-dehydratase – dTGD – EC 4.2.1.46	2HUN	1810	2
	1OC2	4001	4
	6BI4	3017	1
	1KEW	16294	4
W- GDP-mannose 4,6-dehydratase – GMD – EC 4.2.1.47	1N7H	7283	9
	2Z1M	2336	3
	1DB3	7433	2
X- UDP-GlcNAc 4,6-dehydratase (5-inverting) – UGNacD – EC 4.2.1.115	3PVZ	1888	1
	4G5H	11179	11
	4TQG	838	1
	2GN4	9157	5
Y- UDP-GlcNac 5,6-dehydratase – UGNac56D – EC 4.2.1.x	3VPS	3157	1
VII. NDP-keto-sugar reductases			
Z- dTDP-D-rhamnose synthase (4-reduct.) – RMD – EC 1.1.1.133	1VL0	9446	10
	1N2S	7511	4
	3SC6	2117	1
AA- GDP-L-fucose synthase (3,5-epim. 4-reduct.) – GMER – EC 1.1.1.271	4E5Y	1232	7
	1E6U	11906	8
BB- GDP-4-keto-6-deoxy-D-mannose reductase – EC 1.1.1.281	2PK3	4493	2
VIII. NDP-sugar acid decarboxylase			
CC- UDP-glucuronate decarboxylase – UGADC (UXS) – EC 4.1.1.35	2B69	15388	4
	2BLL	2803	11
DD- UDP-apiose/xylose synthase – UGADC (UAXS) – EC 4.1.1	6H0N	561	2

Table 1. Subsets created in this study