



Published in final edited form as:

*Nat Prod Rep.* 2012 October ; 29(10): 1201–1237. doi:10.1039/c2np20039b.

## The Structural Biology of Enzymes Involved in Natural Product Glycosylation

Shanteri Singh<sup>1</sup>, George N. Phillips Jr.<sup>2</sup>, and Jon S. Thorson<sup>3</sup>

Jon S. Thorson: jsthorson@uky.edu

<sup>1</sup>Laboratory for Biosynthetic Chemistry, Pharmaceutical Sciences Division, School of Pharmacy, University of Wisconsin-Madison, 777 Highland Avenue, Madison, WI 53705

<sup>2</sup>Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706

<sup>3</sup>Center for Pharmaceutical Research and Innovation, University of Kentucky College of Pharmacy, 789 South Limestone Street, Lexington, KY 40536-0596; Tel: (859) 218-0140

### Abstract

The glycosylation of microbial natural products often dramatically influences the biological and/or pharmacological activities of the parental metabolite. Over the past decade, crystal structures of several enzymes involved in the biosynthesis and attachment of novel sugars found appended to natural products have emerged. In many cases, these studies have paved the way to a better understanding of the corresponding enzyme mechanism of action and have served as a starting point for engineering variant enzymes to facilitate production of differentially-glycosylated natural products. This review specifically summarizes the structural studies of bacterial enzymes involved in biosynthesis of novel sugar nucleotides.

### 1. Introduction

Glycosylation is a prevalent and critical reaction in cells, functioning in energy metabolism, maintenance of cell integrity, molecular recognition, pathogen virulence, molecular defense, signaling information storage and chemical defense<sup>1–4</sup>. Many bacteria use glycosylated small molecules as chemical weapons to gain a selective advantage or as signaling molecules for intra- and interspecies communication<sup>5</sup>. Scheme 1 presents just small set of representative glycosylated bacterial secondary metabolites. Given carbohydrates are capable of accessing a wide range of unique chemical space<sup>6–9</sup>, the sugars attached to these metabolites can clearly compliment and expand inherent natural product chemical diversity. Such sugar attachments also dramatically influence numerous properties of the metabolite to which they are attached including pharmacological and pharmacokinetic properties such as solubility, distribution, metabolic stability and/or tissue, cellular and/or molecular specificity<sup>2, 10, 11</sup>. Thus, the differential glycosylation of natural products has emerged as a viable strategy to produce bioactive compounds with improved activity<sup>8, 12–16</sup>.

In nature, these sugar structures are generated through enzymatic modification of the functional groups of a common sugar nucleotide precursor via complex multi-enzyme pathways<sup>13, 14</sup>. The inherent promiscuity of these enzymes, in conjunction with advances in bioengineering methodology, biochemical and structural studies, have enabled efforts to modify the glycosylation patterns of natural products through metabolic pathway

engineering<sup>17–22</sup> and enzymatic glycodiversification<sup>14, 23–29</sup>. Such studies are greatly augmented via a fundamental understanding of the intricate substrate-enzyme interactions, the substrate scope and the detailed catalytic mechanism of targeted enzyme-catalyzed transformation. The goal of this review is to summarize the structural biology studies of enzymes involved in the biosynthesis of key precursors of glycosylated microbial natural products – namely, novel sugar nucleotides. For an overview of the structural biology of the class of enzymes that catalyze the culminating glycosylation step in such pathways (glycosyltransferases, GTs), the reader is referred to recent reviews<sup>24, 25, 30–35</sup>.

## 2. Fundamentals of bacterial sugar nucleotide biosynthesis

Nucleotide 5'-diphosphosugars (NDP-sugars) represent the most common form of sugar donor employed by glycosyltransferases of microbial biosynthetic pathways<sup>36</sup>. Despite the enormous variety of glycosides found among microbial secondary metabolites, the biosynthesis of their corresponding NDP-sugar precursors are divergent and share a number of conceptual and strategic similarities. These pathways are initiated via the formation of nucleoside monophosphate (NMP) or nucleoside diphosphate (NDP) derivatives through the action of sugar-1-phosphate nucleotidyltransferases - enzymes which couple a sugar-1-phosphate (Sugar1P) with nucleoside triphosphate (NTP), via expulsion of phosphate (P<sub>i</sub>) or pyrophosphate (PP<sub>i</sub>), to provide the desired NDP/NMP-sugar (Scheme 2). The corresponding Sugar1P originates from primary metabolic intermediates such as fructose-6-phosphate and glucose-6-phosphate. Glucose-6-phosphate is a biosynthetic precursor of many bacterial thymidine diphosphate (dTDP)-, cytidine diphosphate (CDP)-, and uridine diphosphate (UDP)-sugars. Fructose-6-phosphate is converted to mannose-6-phosphate by phosphomannoisomerase (PMI) in the biosynthesis of guanosine diphosphate (GDP)-sugars and to glucosamine-6-phosphate by glucosamine-6-phosphate synthase (GlmS) in the formation of UDP-glucosamine-based analogs. In all cases, the sugar-6-phosphates are converted to the corresponding Sugar1P by distinct but related phosphohexose mutases prior to nucleotidyltransfer<sup>37</sup> (Scheme 2). While structural information for many enzymes involved in sugar phosphate biosynthesis are known<sup>38–42</sup>, they fall outside the scope of this review.

Among the NDP-sugars, dTDP-sugars are the predominant sugar donor form utilized in the biosynthesis of bacterial glycosylated natural products. Most dTDP-sugars from bacteria are 6-deoxyhexoses and many are also deoxygenated and/or functionalized at C2, C3, or C4<sup>14</sup>. dTDP-sugars derive, almost exclusively, from glucose-1-phosphate (Glc1P), which is converted to a common intermediate (dTDP-4-keto-6-deoxy-D-glucose) in the biosynthesis of many diverse sugar nucleotides by the action of two ubiquitous enzymes – Glc1P thymidylyltransferase and dTDP-D-glucose 4,6-dehydratase. While genes encoding these common enzymes are sometimes found associated with secondary metabolite-encoding gene clusters, these core enzymes also participate in primary metabolism (particularly bacterial cell wall biosynthesis). The culminating C4-keto group of the common intermediate formed by this common two enzyme conversion is key to facilitating subsequent enzyme-catalyzed functionalization at C2, C3 and C4 (Scheme 3). Typical modifications proceeding from dTDP-4-keto-6-deoxy-D-glucose in bacterial secondary metabolism (and, in some cases also primary metabolism) include: 2-, 3- or 4-deoxygenation; 2-, 3-, 4- or 5-epimerization; 3- or 4-amination; and C-, N- or O-methylation<sup>13, 14</sup>. In addition, amine oxidation, oxidation/decarboxylation, acetyltransfer, carbomoyltransfer and sulfur installation are among a range of downstream modifications known to further expand structural diversity<sup>13, 43–47</sup>.

Over the last two decades the structures of several sugar biosynthetic enzyme family members have emerged including examples for which structures of enzymes for an entire sugar nucleotide pathway have been elucidated (*e.g.*, dTDP-L-rhamnose)<sup>48–50</sup>. These

structural studies have led to new mechanistic hypotheses<sup>46, 51–53</sup>, offered a template to extend mechanistic studies via site-directed-mutagenesis<sup>48, 53–58</sup>, and served as a foundation for structure-based engineering of new catalysts<sup>35, 59–61</sup>. The structural studies highlighted within this review have been organized by reaction type with a primary focus upon the core enzymes involved in manipulating the sugar scaffold structure (specifically, sugar nucleotide formation and core sugar nucleotide modification reactions - deoxygenation, epimerization, oxidation/reduction and transamination). A section highlighting structural studies for enzymes that offer additional sugar nucleotide modification (methylation, acylation, *N*-oxidation, decarboxylation, formyl, enolpyruvyl transfer and pyranose/furanose interconversion) has also been included. Table 1 lists the structures discussed as organized in the context of this review for quick reference. For additional details regarding the chemistry and mechanism of sugar nucleotide-modifying enzymes, the reader is referred to the many excellent recent reviews<sup>13, 46–50, 55, 62–64</sup>.

### 3. Sugar nucleotide formation

#### 3.1. Nucleotidyltransferases

Glycoside biosynthesis typically begins with the conjugation of  $\alpha$ -D-glucose-1-phosphate (Glc1P) or  $\alpha$ -D-mannose-1-phosphate (Man1P) to NMP (from NTP). This reaction, which, in most cases is dependent upon divalent metal, is catalyzed by a nucleotidyltransferase (also sometimes referred to as a sugar nucleotide pyrophosphorylase, EC 2.7.7.-) and proceeds with the concomitant loss of pyrophosphate. The three-dimensional structures of many examples of this class of enzyme are known, including thymidyltransferases [dTDP-glucose pyrophosphorylases, RmlA/RfbA (EC 2.7.7.24)] from *Salmonella enterica*<sup>59</sup>, *Bacillus anthracis*, *Escherichia coli*<sup>65</sup>, *Pseudomonas aeruginosa*<sup>54</sup>; uridylyltransferases [UDP-glucose pyrophosphorylase, UGPase, (EC 2.7.7.9)] from *Escherichia coli*<sup>66, 67, 68</sup>, *Helicobacter pylori*<sup>69, 70</sup>; UDP-*N*-acetylglucosamine (UDP-GlcNAc) pyrophosphorylase GlmU (EC 2.7.7.23) from *Escherichia coli*<sup>71</sup>, *Streptococcus pneumoniae*<sup>72</sup>, *Mycobacterium tuberculosis*<sup>73, 74</sup>, *Haemophilus influenzae*<sup>75</sup>; cytidyltransferase [CDP-glucose pyrophosphorylase, CGPase, (EC 2.7.7.33)] from *Salmonella typhi*<sup>76, 77</sup>; guanylyltransferase [GDP-mannose pyrophosphorylase, GMPase, (EC 2.7.7.22)] from *Thermotoga maritima*<sup>78</sup>; adenylyltransferase [ADP-glucose pyrophosphorylase, AGPase, (EC 2.7.7.27)] from *Agrobacterium tumefaciens*<sup>79</sup>.

Despite very little sequence homology, nucleotidyltransferases share a similar domain organization and common structural features. Though they do not seem to have any obligate multimeric preference (Figure 1A & 1B), most adopt a tetrameric structure<sup>59, 66, 67, 69, 74, 78</sup> wherein multimerization is generally mediated through C-terminal domain interactions (Figure 1A & 1B). Each monomer is composed of a conserved N-terminal catalytic domain and a C-terminal auxiliary domain (colored purple in Figure 2). The N-terminal catalytic domain is composed of an  $\alpha/\beta/\alpha$  sandwich reminiscent of the dinucleotide-binding Rossmann fold (Figure 1C). It consists of a twisted mixed  $\beta$ -sheet made up of seven  $\beta$ -strands arranged in the order 3214657 (Figure 1C). The central  $\beta$ -sheet is flanked by eight  $\alpha$ -helices tightly packed against the  $\beta$ -sheet. Often there are additional secondary structure insertions, sometimes known as subdomains, between strands  $\beta$ 5 and  $\beta$ 6 and strands  $\beta$ 2 and  $\beta$ 3, depending upon the family (colored blue in Figure 2). For example, in the thymidyltransferase RmlA<sup>59</sup>, the insertion between  $\beta$ 5 and  $\beta$ 6 is a mixed two-stranded  $\beta$ -sheet flanked by  $\alpha$ -helix ( $\beta$ 5a and  $\beta$ 5b), while cytidyltransferase CGPase<sup>76</sup> has an extra mixed two stranded  $\beta$ -sheet ( $\beta$ 2a and  $\beta$ 2b) inserted between  $\beta$ 2 and  $\beta$ 3 (Figure 1F). The C-terminal auxiliary domains in this family sometimes form a left-handed  $\beta$ -helix motif (L $\beta$ H), as in the CGPase, the GMPase, in GlmU (colored purple in Figure 2D, described in detail in **section 5.2.1.**) and in AGPases<sup>72–74, 80, 81</sup>. adenylyltransferase AGPase (colored purple in Figure 2C)<sup>79, 82</sup>, and guanylyltransferase GMPase (colored purple in Figure 2E)<sup>78</sup>. The C-

terminal auxiliary domains of RmlA, UGPase and CGPase (colored purple in Figures 2A 2B & 2F) are distinct and contain two or three  $\alpha$ -helices as a main feature. In some enzymes this C-terminal domain presents a second enzymatic activity as exemplified by the bifunctional GMPase and GlnU, which exhibit phosphomannose isomerase (PMI)<sup>78</sup> and acetyltransferase activity (see **section 5.2.1.**)<sup>74</sup>, respectively. The C-terminal domain can also regulate the nucleotidyltransferase activity by mediating enzyme oligomerization<sup>59</sup> and/or by binding allosteric regulators<sup>82</sup>.

The general nucleotidyltransferase active site pocket is located in a deep cleft formed by the central  $\beta$ -sheet and flexible loops L1 and L3. Within this active site, the nucleotide base of the substrate is sandwiched between loop L1 after the strand  $\beta$ 1, which contains the canonical signature motif G11-G12-X-G14-X-R16 (numbering in *S. enterica* RmlA), and loop L3 following  $\beta$ 3. The O2, O3, and O4 hydroxyls of the sugar moiety interact with conserved residues within bacterial nucleotidyltransferases that emanate from the strand  $\beta$ 6 and flexible loops L5, L5b, L6 (Figure 1C). The phosphates of the nucleotide are hydrogen-bonded by arginine and lysine side-chains (R16, K163 and R195 in *S. enterica* RmlA), and the ribose and pyranose are also anchored by several hydrogen-bonding interactions as shown in Figures 3A and 3B. Comparative analysis of the structures of apo, NTP and NDP-sugar complexes reveals substrate and product binding are associated with significant changes in both the conformation of loop regions lining the active site and in the overall orientation of the core domain relative to the subdomain involving a  $\beta$ -sheet ( $\beta$ 5a and  $\beta$ 5b) insertion between  $\beta$ 5 and  $\beta$ 6 (colored blue in Figure 2E)<sup>78</sup>. For example, in the enzyme-product complex of GMPase, the phosphate backbone interacts with the side chains of two conserved aspartates (D109 and D260 in GMPase) via  $Mg^{2+}$  (Figure 3B)<sup>78</sup> whereas in the enzyme-NTP complex, the phosphate groups are pointed away from the sugar-binding site and interact with the main chain nitrogen atoms of the signature motif, GGXGXR. Upon Man1P binding, the subdomain containing  $\beta$ -sheet 5a and 5b (colored blue in Figure 2E) rotates by  $10^\circ$  toward Man1P, which results in the global closure of the sugar-binding site and restricts active site accessibility<sup>78</sup>.

Catalytic mechanisms for many nucleotidyltransferases have been thoroughly studied<sup>54, 59, 65, 76–78</sup>. Typically, nucleotidyltransferases utilize an ordered sequential mechanism with NTP binding first. Isotopic-labeling is consistent with a  $S_N2$ -type mechanism leading to stereochemical inversion of the NTP  $\alpha$ -phosphate upon nucleophilic attack by the Sugar1P (Scheme 4)<sup>83, 59, 65, 76, 77</sup>. This general mechanism is also supported by RmlA structural studies in complex with reactants and products which revealed the Sugar1P nucleophile and departing pyrophosphate of the NTP to bind in a manner consistent with a  $S_N2$ -type mechanism<sup>59, 76</sup>. In this complex, the essential divalent metal is believed to play a role in transition-state stabilization/orientation and activation of the pyrophosphate leaving group<sup>84, 65, 85</sup>.

## 4. Core sugar scaffold modification

### 4.1. Sugar dehydratases

Of the dehydratases (EC 4.2.1.-) involved in NDP-sugar core modification, the NDP-hexose-4,6-dehydratases are central to the biosynthesis of most deoxysugars. These enzymes catalyze a stepwise NAD(P)<sup>+</sup>-dependent C4-oxidation-C5/C6-elimination (dehydration)-C5/C6-ene reduction reaction sequence which enables the facile production of a NDP-4-keto-6-deoxysugar (most commonly of glucose origin). To date, NDP-hexose-4,6-dehydratases have been found to utilize the short-chain dehydrogenase/reductase (SDR) structural fold and include: dTDP-glucose-4,6-dehydratases (EC 4.2.1.46) RmlB from *Salmonella enterica* and *Streptococcus suis* and DesIV from *Streptomyces venezuelae*<sup>55, 86, 87</sup>; CDP-glucose-4,6-dehydratase (EC 4.2.1.45) from *Yersinia pseudotuberculosis* and *Salmonella*

*typhi*<sup>88, 89</sup>; GDP-mannose-4,6-dehydratase (EC 4.2.1.47) from *Escherichia coli*, *Pseudomonas aeruginosa* and *Aquifex aeolicus* VF5<sup>90–92</sup>; and UDP-GlcNAc-4,6 dehydratase (EC 4.2.1.115) FlaA1 from *Helicobacter pylori*<sup>93</sup>. Dehydrases which subsequently act upon NDP-4-keto-6-deoxyhexoses to facilitate pyridoxamine 5'-phosphate (PMP)-mediated dehydration/deoxygenation adjacent to the C4 carbonyl adopt an aspartate aminotransferase (AAT) fold and include: the GDP-4-keto-6-deoxy-D-mannose-3-dehydratase ColD from *Escherichia coli*<sup>94</sup>; and the CDP-6-deoxy-L-threo-D-glycero-4-hexulose 3-dehydrase E1 from *Yersinia pseudotuberculosis*<sup>53</sup>.

**4.1.1. SDR fold dehydratases**—Dehydratases of SDR family do not exhibit an obligate multimeric preference and have been found as dimeric<sup>55, 86, 87, 90, 94, 95</sup>, tetrameric<sup>88, 91</sup>, and hexameric enzymes<sup>93</sup>. The monomer-monomer association in these multimers occurs principally through hydrophobic interactions via a four-helix bundle (Figures 4A & 4B) and each monomer exhibits an  $\alpha/\beta$  structure that can be divided into two domains. Of these, the larger N-terminal NAD(P)<sup>+</sup>-binding domain consists of a seven-stranded  $\beta$  sheet in the order 3214567 flanked by  $\alpha$  helices, yielding a modified Rossmann fold (Figures 4C & 4D). The smaller C-terminal domain, responsible for binding the sugar substrate, contains four  $\beta$  strands and six  $\alpha$  helices (Figures 4C & 4D). The interdomain cavity accommodates the active-site includes a characteristic GXGXXG motif and a conserved YXXXX which, in conjunction with a conserved S/T forms a catalytic triad. Based upon a comparison of ligand bound and apo enzyme (RmlB), substrate binding does not induce substantial conformation change<sup>55</sup>.

The active site of a typical SDR dehydratase is located within the cavity formed by the junction of the N- and C-terminus within each monomer. Cofactor binding is facilitated through a range of hydrophobic and hydrogen bonding interactions (Figure 5A) to present the nicotinamide ring in the *syn* conformation, consistent with *pro-S* hydride transfer from C4 of the dTDP-D-glucose to C4' of the NAD(P)<sup>+</sup>. This cofactor-binding region contains the characteristic  $\alpha/\beta/\alpha$  motif wherein a conserved motif G8-X-G10-X-X-G13 (numbering in RmlB from *S.enterica*) contributes to interaction with nucleotide diphosphate group<sup>55</sup>. In addition, one of the hydroxyl oxygens of nicotinamide ribosyl hydroxyls often hydrogen bond to lysine (K171 in RmlB from *S.enterica*) from the conserved motif, YXXXX. A conserved S/T (T133 in RmlB from *S.enterica*) (Figure 5B) has also been implicated in transition-state stabilization as a proton shuttle or possibly via involvement in a low barrier H-bond with the substrate<sup>55</sup>. Sugar nucleotide binding is also mediated via H-bonding of phosphates of the nucleotide and arginine side-chains (Figures 5C & 5D) as well as conserved H-bonding interactions with the sugar moiety as shown in the Figures 5B 5C & 5D.

The mechanism of NDP-4-keto-6-deoxyhexose biosynthesis proceeds through three fundamental steps – C4 oxidation, C5/C6 dehydration and C5/C6 ene reduction (Scheme 5)<sup>57, 96–99</sup>. C4 oxidation involves sugar C4-OH deprotonation, mediated by a conserved tyrosine (Y167 in *S.enterica* RmlB; Y141 in FlaA1) (Figures 5B & 5D), followed by C4 hydride transfer to nicotinamide<sup>57, 96</sup>. During this step, a conserved lysine (K171 in *S.enterica* RmlB) and serine/threonine (T133 in *S.enterica* RmlB; T131 in FlaA1) (Figures 5B & 5D), stabilize the tyrosinate general base (Figures 5B & 5D). Formation of the C4 carbonyl enables a subsequent C5/C6  $\beta$ -elimination and this dehydration reaction is mediated by a conserved general base (E135 in *S.enterica* RmlB; is K133 in FlaA1) and general acid (D134 in *S.enterica* RmlB; D132 in FlaA1) pair to afford the dTDP-4-keto-glucose-5,6-ene intermediate (Scheme 5)<sup>97</sup>. The induced chair conformation reorganization, stabilized by hydrogen bonds between the C4 carbonyl oxygen and the side chains of Y167 and T133 (in *S.enterica* RmlB) (Scheme 5), enhance the electrophilicity and orientation of C6 for hydride transfer from NADH (with subsequent reprotonation at C5 from solvent).

Thus, the net reaction affords an intramolecular transfer of hydride from C4 to C6 (ultimately, with inversion of absolute stereochemistry at C6)<sup>55</sup>. While most NDP-4,6-dehydratases lead to retention of stereochemistry at C5, the UDP-GlcNAc-4,6-dehydratase FlaA1 (involved in the biosynthesis of pseudaminic acid) leads to inversion at C5 en route from UDP-GlcNAc to UDP-2-acetamido-2,6-dideoxy- $\beta$ -L-*arabino*-hex-4-ulose. (Scheme 6). The structural equivalent of RmlB E135 in FlaA1 is K133 and has been proposed to facilitate water-mediated protonation on the opposite face molecule as an alternative to Y141 in RmlB (Scheme 6)<sup>93</sup>.

**4.1.2. AAT fold dehydratases**—Dehydratases which subsequently act upon NDP-4-keto-6-deoxyhexoses to facilitate PMP-mediated dehydration or deoxygenation adjacent to the C4 carbonyl include ColD and E1 – both dimeric enzymes involved in the biosynthesis of 3,6-dideoxy sugars. While these catalysts adopt an aspartate aminotransferase (AAT) fold and, like typical aminotransferases, use PMP as a cofactor, the conserved transaminase lysine which typically forms an imine with PMP is replaced in these enzymes with a conserved histidine [H188 in ColD and H220 in E1]<sup>53, 94</sup>. Although, they have same structural fold (Figure 6A), a key feature that distinguishes E1 from ColD is the presence of a [2Fe-2S] cluster critical to the final reductive deoxygenation step, the ligands for which are cysteine and a histidine (C251-X<sub>1</sub>-C253-X<sub>7</sub>-C261-X<sub>17</sub>-H278)<sup>100</sup>. While a single mutation, H220K could convert E1 into a PLP/L-glutamate-dependent C4 aminotransferase, this E1 variant-catalyzed reaction was not catalytic due to an inability to regenerate PLP<sup>53</sup>. Further mutagenesis revealed that four active site residues of E1 (D194H, Y217H, H220K, and F345H) and two active site residues of ColD (S187N, H188K) have successfully transformed each into a PLP/L-glutamate-dependent C4 aminotransferase<sup>46, 53, 58</sup>.

Both ColD and E1 exhibit extensive dimer interface with each subunit folded into a large N-terminal cofactor binding domain and a small C-terminal domain. The N-terminal domain has a mixed  $\beta$ -sheet with seven or eight strands (strand order 17654238) surrounded by eight  $\alpha$ -helices (Figure 6A). The smaller C-terminal domain consists of an antiparallel, three-stranded  $\beta$ -sheet surrounded by helices. Similar to sugar aminotransferases (SATs, **section 4.4**), both monomers contribute residues to the active site. The PLP-binding pocket is formed primarily by one subunit and a loop (consisting of residues 240–253 in ColD and 282–293 in E1) from the second subunit (residues colored in cyan in Figures 6B & 6C). While the aromatic stacking interactions (W88 in ColD and F120 in E1) help to orient the PLP, a network of H-bonding interactions with conserved serine, histidine and glutamate/aspartate from one subunit and an asparagine (S183/S215, H188/H220, E162/D194, and N248/N288 in ColD/E1) from a second subunit further anchor the cofactor in the active site (Figures 6B & 6C). Mutagenesis studies have confirmed that the [2Fe-2S] cluster in E1 is coordinated by three cysteine residues (C251, C253 and C261)<sup>100</sup> from a flexible loop (comprising residues 253–268; boxed in Figure 6A). The lack of electron density in this loop has been attributed to the putative “open” and “closed” conformational changes of the substrate binding pocket where the loop flexibility of E1 allows facilitates the formation of a reasonably tight binary complex with the reductase E3 to support the transfer of electrons from the reduced iron–sulfur center of E3 (via the iron–sulfur cluster of E1) to reduce the E1-bound substrate<sup>46, 53, 101</sup>.

A catalytic mechanism for ColD has been proposed (Scheme 7A) in which a Schiff base, formed between the sugar C4 carbonyl and PMP, facilitates the H188 general acid/base-mediated loss of water to afford the  $\Delta^{3,4}$ -mannoseen intermediate. Intriguingly, upon hydrolysis of the cofactor in a manner reminiscent of standard transamination, the corresponding enamine product is hydrolyzed by ColD to liberate ammonia and the corresponding NDP-3,6-dideoxy-4-keto-sugar<sup>95</sup>. E1 shares many common mechanistic features with ColD (Scheme 7B) including the role of PMP and a conserved histidine

(H220) in the early stage dehydration<sup>46</sup>. However, distinct from ColD, E1 associates with a flavoprotein reductase (E3, a member of the ferredoxin-NADP<sup>+</sup> reductase family) which provides reducing equivalents from NADH in a one electron fashion to reduce the corresponding  $\Delta^{3,4}$ -glucose intermediate in E1 (mediated via the E1 [2Fe-2S])<sup>46, 102–106</sup>. Upon reduction, the hydrolysis of the Schiff base gives rise to NDP-3,6-dideoxy-4-keto-sugar product and PMP to complete the cycle<sup>46</sup> (Scheme 7B).

## 4.2. Sugar epimerases/isomerases

Sugar epimerases (EC 5.1.3.-) catalyze the stereochemical inversion of the configuration of an asymmetric carbon atom in a carbohydrate<sup>48, 62, 63</sup>. Depending on the type of chemistry, NDP-sugar epimerases can be classified into three main groups<sup>62</sup>: epimerases that epimerize at positions alpha to a carbonyl (typically the common C4 carbonyl) as exemplified by dTDP-6-deoxy-D-xylo-4-hexulose-3,5-epimerase (RmlC, EC 5.1.3.13) and GDP-4-keto-6-deoxy-mannose-epimerase/reductase (GMER, EC 1.1.1.271); those which epimerize via an oxidation/reduction at a single carbon as exemplified by CDP-tyvelose-2-epimerase (EC 5.1.3.10), UDP-galactose-4-epimerase (GALE, EC 5.1.3.2), and ADP-L-glycero-D-mannoheptose-6-epimerase (AGME, EC 5.1.3.20); and enzymes that epimerize at position adjacent to the sugar anomeric center via nucleotide elimination and re-addition as demonstrated by UDP-N-acetyl-D-glucosamine-2-epimerase (UDP-GlcNAc-2-epimerase). The structures for a range of sugar epimerases have been reported in the recent literature and cumulatively, this chemistry is accomplished by three main structural folds – a cupin fold<sup>107</sup>, an extended SDR fold and a GT-B fold (most commonly associated with sugar-nucleotide-dependent glycosyltransferases<sup>35</sup>).

**4.2.1. Cupin superfamily**—Epimerases/isomerases belonging to this structural fold invert one or two stereocenters alpha to a sugar carbonyl via simple keto-enol tautomerization<sup>48</sup>. Structures for a number of members of this family have been reported including: dTDP-4-dehydrorhamnose 3,5-epimerase, (EC 5.1.3.13) RmlC (involved in dTDP-L-rhamnose pathway of *Salmonella enterica*, *Streptococcus suis*, *Mycobacterium tuberculosis*, and *Pseudomonas aeruginosa*)<sup>62, 64, 108–112</sup>; dTDP-3-amino-4-keto-2,3,6-trideoxy-3-C-methyl-glucose-5-epimerase EvaD (involved in dTDP-epivancosamine of *Amycolatopsis orientalis*)<sup>113, 114</sup>; dTDP-6-deoxy-D-xylo-4-hexulose 3-epimerase NovW (involved in the *Streptomyces spheroides* dTDP-noviose pathway)<sup>115</sup>; and dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase FdtA (involved in the *Aneurinibacillus thermoaerophilus* dTDP-3-acetamido-3,6-dideoxy- $\alpha$ -D-galactose pathway)<sup>116</sup>. Of these FdtA, is somewhat unique as it utilizes the keto-enol tautomerization reaction to enable inversion of the initiating C4 carbonyl (in contrast to all other members which utilize a similar strategy to enable inversion adjacent to the initiating carbonyl).

‘Cupa’ in Latin means a small barrel and enzymes of this family are homodimers of ‘jelly-roll’ topology consisting of a conserved beta-barrel fold (Figure 7A). The dimer interface is created by the antiparallel interaction of the  $\beta 3$  strand of one monomer with the  $\beta 5$  strand of the other monomer (Figure 7A). Each  $\beta$  barrel comprises 13  $\beta$  strands and can be divided into three separate regions: the N-terminal, core, and C-terminal regions. The N terminus consists of an antiparallel  $\beta$ -sheet ( $\beta 1$ – $\beta 3$ ) and a two-turn  $\alpha$ -helix. The core of the monomer consists of two twisted antiparallel  $\beta$ -sheets ( $\beta 5$ – $\beta 13$ ), which form a flattened barrel. One end of the barrel is open, and the entrance is lined with polar residues. The other side is obscured by  $\beta$ -strands that fold over the entrance. A number of hydrophobic residues in this part of the polypeptide chain seal the entrance to the barrel (Figures 9A & 9B). The C-terminal region consists of two or three small helices, depending on the enzyme (Figure 7). The active site conserved residues tyrosine, lysine and a histidine-glutamate/aspartate

dyad (Y134, K74, H65-D171 in *Pseudomonas aeruginosa* RmlC) (Figure 9A) are critical to binding and catalysis.

The general mechanism of cupin-based epimerases is initiated by proton abstraction from one face of the sugar ring by the catalytic dyad (H65 and D171 in *Pseudomonas aeruginosa* RmlC) (Figures 9A & 9B) to provide an enolate intermediate which is stabilized by K74. Reprotonation on the opposite face is facilitated by tyrosine (Y134 in *Pseudomonas aeruginosa* RmlC) or solvent (Scheme 8). RmlC, EvaD and NovW catalyze 3,5-, 5- and 3-epimerase activity, respectively<sup>112, 113, 117</sup>. Based upon site directed mutagenesis of EvaD (M131F) the orientation the conserved Y133 was postulated to dictate regioselectivity<sup>113</sup>. However, the subsequent structure of NovW (which catalyzes epimerization at C3) revealed an identical orientation for this conserved tyrosine<sup>113, 117</sup>. The dual C3/C5 epimerization reaction catalyzed by RmlC proceeds in a step-wise fashion beginning with C5 inversion (Scheme 8)<sup>112</sup>.

**4.2.2. SDR enzymes**—Given the similar chemistries, it is perhaps not surprising that both dehydratases (discussed in **section 4.1.1.**) and epimerases can both adopt a SDR structural fold (Figure 8C). Epimerases within this structural family catalyze a NAD(P)<sup>+</sup>-dependent oxidation/reduction sequence that facilitates hydride removal from one face of the sugar (to afford a carbonyl) with re-introduction of the hydride from the opposing face of the sugar to ultimately afford stereochemical inversion at the target position (Scheme 9). Examples of this class include: GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase/4-reductase (GMER) also known as GDP-L-fucose synthetase from *Escherichia coli*<sup>118–120</sup>, UDP-galactose epimerase (GALE) from *Escherichia coli*<sup>56, 121–124</sup>, ADP-L-glycero-D-mannoheptose-6-epimerase (AGME) from *Escherichia coli* and *Helicobacter pylori*<sup>125–128</sup> and CDP-tyvelose-2-epimerase from *Salmonella typhi*<sup>129</sup>. As previously described for SDR folds, they do not have an obligate preference for the multimeric state (Figures 8A–C) and each monomer exhibits an  $\alpha/\beta$  structure that can be divided into two domains. The larger N-terminal domain binds the nucleotide cofactor NAD(P)<sup>+</sup> and consists of a seven-stranded  $\beta$  sheet in the order 3214567 flanked by  $\alpha$  helices, yielding a modified Rossmann fold (Figure 8C). SDR epimerases typically display a T6-G7-X-X-G10-X-X-G13 (numbering in *E. coli* GALE) cofactor binding motif and a Y149-X-X-X-K153 (numbering in *E. coli* GALE) active site motif, with the tyrosine residue of this series serving as a critical general base (Figure 9C and Scheme 11). In addition, a nearby active site serine/threonine and/or an asparagine (S124 in *E. coli* GALE) is often important (Figure 9C). SDR epimerase substrate binding occurs in the C-terminal region, which is the main structural determinant of substrate specificity and, although substrate specificity is not well understood, the volume of the active site is a main contributor<sup>130</sup>.

The catalytic mechanisms for UDP-galactose-4-epimerase and CDP-tyvelose-epimerase (Scheme 9)<sup>41, 48, 56, 129</sup> involve three fundamental steps: i) NAD(P)<sup>+</sup>-dependent oxidation of the C4 hydroxyl; ii) a rotation of the enzyme-bound keto-sugar intermediate by  $\sim 180^\circ$  about the UDP-sugar bond; and iii) NAD(P)H-dependent reduction of the C4 carbonyl on the opposite face of the sugar. The conserved lysine (K153 in *E. coli* GALE) and serine/threonine (S124 in *E. coli* GALE) are believed to modulate the general acid/base function of tyrosine (Y149 in *E. coli* GALE) in this mechanism.

**4.2.3. GT-B fold**—Bacterial UDP-*N*-acetylglucosamine-2-epimerase catalyzes the reversible epimerization at C2 of UDP-*N*-acetylglucosamine (UDP-GlcNAc) to UDP-*N*-acetylmannosamine (UDP-ManNAc)<sup>131, 132</sup>. This enzyme is a homodimer wherein dimerization is mediated by N-terminal helices  $\alpha_3$ ,  $\alpha_4$  and  $\alpha_5$  (Figures 10A & 10B) with each monomer having two  $\alpha/\beta/\alpha$  domains that form a deep cleft at the domain interface (Figure 10B). The N and C-terminal domain both contain a Rossmann fold with a seven-



stranded and six stranded  $\beta$ -sheet, respectively (Figure 10B). Structures revealed both ‘open’ and ‘closed’ monomer conformations in dimer assembly and an observed  $10^\circ$  interdomain rotation, induced upon binding to substrate, activates one monomer for catalysis and may play a role in regulation<sup>131</sup>.

The mechanism of UDP-*N*-acetylglucosamine-2-epimerase proceeds via the elimination of UDP from UDP-GlcNAc (Scheme 10). The *anti* elimination of UDP, which ultimately affords 2-acetamidoglucal and UDP, invokes a transient oxocarbenium species. The subsequent *syn* addition of UDP and C2 reprotonation provides UDP-ManNAc<sup>133, 134</sup>. Based on the UDP-bound structure of UDP-GlcNAc epimerase, H213, which interacts with the NDP moiety in the active site, is assumed to act as a general acid (Figure 9D) while other conserved acidic residues (D95, E117 and E131) may be involved in stabilizing the oxocarbenium intermediate and/or as a general acid/base to promote C2 epimerization<sup>62</sup>.

### 4.3. Sugar ketoreductases

In the context of sugar nucleotide biosynthesis, ketoreductases (E.C. 1.1.1.-) are NAD(P)H-dependent enzymes which transfer hydride from NAD(P)H to a NDP-sugar carbonyl. NDP-sugar ketoreductases adopt two structural folds – the SDR fold (a fold discussed previously in sections 4.1.1. and 4.2.2.) and the glucose-fructose oxidoreductase (GFOR) fold. Representative examples include the *Salmonella enterica serovar Typhimurium* dTDP-6-deoxy-L-lyxo-4-hexulose C4-reductase (EC 1.1.1.133) RmlD [involved in dTDP-L-rhamnose biosynthesis<sup>135</sup>] and the *Actinomadura kijaniata* dTDP-3,4-diketo-2,6-dideoxy-D-glucose C3-ketoreductase KijD10 [involved in dTDP-L-digitoxose biosynthesis<sup>136</sup>], which adopt a SDR fold and GFOR fold, respectively.

**4.3.1. RmlD**—RmlD belongs to the previously discussed SDR structural family (sections 4.1.1. and 4.2.2.)<sup>137</sup> and contains the requisite Y128-X-X-X-K132 and glycine rich, G7-X-X-G10-X-X-G13 conserved motifs. RmlD is a homodimer and the dimerization is mediated by an intrafacial  $Mg^{2+}$  ion, also believed to be important for the proper orientation of cofactor and substrate binding domains. RmlD has several unusual features compared to other SDR enzymes: i) the dimer interface of RmlD is on the opposite face of Rossmann fold (involving helices  $\alpha_1$ ,  $\alpha_6$  and  $\alpha_C$  compared to enzymes such as RmlB (colored purple in Figure 11A); ii) dimerization is mediated mainly by hydrophilic interactions whereas hydrophobic interactions dominate the monomer/monomer interface in other SDR enzymes; iii) RmlD requires  $Mg^{2+}$  for full activity and does not discriminate between the cofactors NAD(H) and NADP(H), iv) large N-terminal cofactor binding domain lacks the structural elements  $\beta_2$ -loop- $\alpha_2$ , (colored orange in Figure 11B) resulting in a  $\beta$ -sheet with six  $\beta$ -strands in the Rossmann fold in the order 213456 (Figure 11A); and v) the small C-terminal substrate binding domain lacks C-terminal helices (colored in red Figure 11B)<sup>135</sup>. The structural difference due to lack of extra structural elements found in RmlB (Figure 11B) is believed to contribute to the cofactor specificity distinction among RmlD and RmlB. Substrate binding by RmlD is accommodated by a solvent-exposed groove that extends from the cofactor binding site into the C-terminal domain to provide a deep cleft formed by two domains and the dimer interface. The hexa-hydrated  $Mg^{2+}$  is hydrogen bonded by three glutamic acid residues, Glu15, Glu190, and Glu292, from each monomer to mediate dimer interaction. Key residues involved in RmlD-catalyzed hydride transfer (Scheme 11) include the conserved Y128 (general acid) and K132 (which modulates the  $pK_a$  of Y128) of the Y-X-X-X-K motif, as well as T104 (Figure 11C). The conserved serine/threonine has been implicated as part of a catalytic triad to enhance proton transfer via the formation of low-barrier hydrogen bonds<sup>135</sup>.

**4.3.2. KijD10**—KijD10 belongs to the GFOR superfamily and crystallized as a tetramer or dimer of dimers<sup>136</sup>. Each monomer comprises a N-terminal NADPH-binding Rossmann fold and a C-terminal mixed “open-faced”  $\beta$ -sheet-based substrate-binding domain. The N-terminal region contains a six-stranded parallel  $\beta$ -sheet flanked by two and three  $\alpha$ -helices (one of which is contributed by the C-terminal domain), respectively (Figure 12A). The C-terminal domain, is dominated by a nine-stranded, mostly antiparallel  $\beta$ -sheet flanked on one side by three  $\alpha$ -helices. The substrate binding site lies at the interface of the two domains and a characteristic of this fold is the presence *cis*-proline in the conserved E101-K102-P103 motif (Figure 12B) found within the active site cleft<sup>138, 139</sup>. The enzyme adopts both “closed” and “open” states when bound to the same substrates (NADP<sup>+</sup> and dTDP-benzene). In KijD10, the nicotinamide ring of the dinucleotide adopts a typical *anti* conformation, whereas the adenine ring is bound in a *syn* orientation (an unusual feature for this family of enzyme). Modeling dTDP-3-keto-6-deoxy-D-galactose within the active site implicated the Asp182 carboxylate to be important for substrate binding (Figure 12B). This aspartate belongs to the active site consensus sequence G177-G178-X-X-X-D182-X-X-X-(Y186/H) observed in related dehydrogenases and reductases. Based on site-directed mutagenesis studies, a mechanism for KijD10 been proposed whereby K102 (Figure 12B) participates as a general acid in protonation of the C3 oxygen. KijD10 is known to tolerate some sugar C2/C4 structural modification and is indiscriminate to C4 stereochemistry.

#### 4.4. Sugar aminotransferases

Sugar aminotransferases (SATs) belong to the aspartate aminotransferase (AAT) type I family (EC 2.6.1.-) previously discussed in **section 4.1.2**. These enzymes catalyze the PLP-mediated transfer of an amino group from an amino acid donor (typically L-Glu, L-Gln, or L-Asp) to a NDP-ketosugar<sup>46</sup>. Key signatures of these enzymes include a conserved lysine (K200, K193 and K183 in DesI, DesV and PseC, respectively), which binds PLP as an imine, and an invariant aspartate (D171, D164 and D154 DesI, DesV and PseC, respectively) which participates as a general acid to facilitate the transamination reaction<sup>140, 141</sup>. SATs of this type are generally involved in the reversible formation of C3 or C4-amino-containing NDP-sugars and the overall reaction often favors the corresponding C3 or C4-keto NDP-sugar<sup>46</sup>. Aminosugars deriving from the action of these enzymes are found in a wide variety of natural products and representatives include: macrolides (*e.g.*, desosamine in erythromycin); anthracyclines (*e.g.*, daunosamine in daunomycin); glycopeptides (*e.g.*, vancosamine in vancomycin), polyenes (*e.g.*, mycosamine in amphotericin B) to name just a few. During the last decade, crystal structures of several bacterial SATs have been solved: GDP-4-amino-6-deoxy-D-mannose transaminase (GDP-perosamine synthase) Per from *Caulobacter crescentus*<sup>142</sup>; dTDP-4-amino-6-deoxy-D-glucose transaminase (EC 2.6.1.33) DesI from *Streptomyces venezuelae*<sup>143</sup>; UDP-4-amino-L-arabinose transaminase (EC 2.6.1.87) ArnB from *Salmonella typhimurium*<sup>144</sup>; UDP-2-acetamido-4-amino-6-deoxy- $\beta$ -L-AltNAc transaminase PseC from *Helicobacter pylori*<sup>145</sup>; UDP-4-amino-6-deoxy-D-GlcNAc transaminase PglE from *Campylobacter jejuni*<sup>146</sup>; UDP-3-amino-2-acetamido-glucuronic acid transaminase WbpE from *Pseudomonas aeruginosa*<sup>147</sup>; dTDP-3-amino-6-deoxy-D-glucose transaminase QdtB from *Thermoanaerobacterium thermosaccharolyticum*<sup>148</sup>; and dTDP-3-amino-4,6-dideoxy- $\alpha$ -D-glucose transaminase (EC 2.6.1.89) DesV from *Streptomyces venezuelae*<sup>149</sup>. In general, the enzymes share many common structural features<sup>45, 46</sup>.

These enzymes function as homodimers (Figure 13B) or higher-order oligomers (Figure 13A), with extensive dimer interfaces, featuring a large mixed  $\beta$ -sheet surrounded by  $\alpha$ -helices (Figure 13C) involving two active sites per dimer. The active sites lie in the clefts formed by the dimer interface, and while each monomer contributes essential residues to both active sites, the active sites are generally independent. The overall architecture of the

subunit comprises two  $\alpha/\beta/\alpha$  domains, a large N-terminal cofactor binding domain, and a small C-terminal domain. The large N-terminal domain features a central seven-stranded mixed  $\beta$ -sheet, with strand order 3245671, flanked on each side by  $\alpha$ -helices. The smaller C-terminal domain consists of an antiparallel, three-stranded  $\beta$ -sheet surrounded by helices (Figure 13C). Additional protein-dependent secondary elements are generally inserted between the helices  $7\alpha_b$  and  $7\alpha_g$  of the C-terminal domain (Figure 13C). The N-terminus of the protein often contributes to the small domain as well (Figure 13C). In some enzymes these domains have been observed or predicted to move considerably upon substrate binding to create a “closed” conformation<sup>45, 150</sup>. The inter-subunit interactions are extensive and involve several elements of secondary structure predominantly from the N-terminal domains of adjoined monomers.

The active sites of SATs are composed of residues from both halves of the dimer and contain several conserved residues. In the resting state, the PLP cofactor is covalently bound to the -amino group of a conserved lysine (K200, K193 and K183 in DesI, DesV and PseC, respectively) as an internal aldimine (Scheme 12)<sup>46</sup>. The major interaction of the NDP-sugar substrate in ligand-bound SATs is through a hydrogen-bonding network centered upon the nucleotide diphosphate with few apparent contacts to the pyranose<sup>143, 145, 148</sup>. Consistent with this perceived relaxed sugar specificity, the C3-aminotransferases DesV and QdtB were demonstrated to accept C4 epimers as substrates<sup>148, 151</sup>. Comparison of ligand-bound structures of the C4-aminotransferases PseC and DesI revealed their respective pyranose binding modes to differ by a 180° rotation<sup>142, 143</sup>. This observed binding mode distinction is consistent with the axial versus equatorial amine installation displayed by PseC and DesI, respectively<sup>148</sup>.

The reaction mechanism of SATs have been well characterized<sup>140, 152–155</sup> where the PLP cofactor serves as an electron sink throughout the reaction, and an invariant aspartate (D171, D164 and D154 DesI, DesV and PseC, respectively) helps to maintain this role by stabilizing the positively charged pyridinium ring of the cofactor<sup>140</sup>. The catalysis is known to proceed via a ping pong mechanism in which the enzyme oscillates between PLP- and PMP-bound forms (Scheme 12A). The  $\alpha$ -amino group of an amino donor (typically L-Glu, L-Gln, or L-Asp) attacks the C4' atom of PLP internal aldimine (I), displacing lysine (K183 in PseC) to yield an external aldimine (II). SATs specifically orient the  $\alpha$ -carbon/hydrogen bond perpendicular to the PLP  $\pi$ -system, thereby favoring this bond for cleavage. Proton extraction is mediated by the active site Lys (to provide quinoid intermediate III), which subsequently serves as general acid in C4' protonation, yielding ketimine intermediate (IV). Subsequent attack by an activated water molecule leads to the formation of PMP and an oxoacid<sup>153, 156, 157</sup>. At this stage, PMP is poised to react with a NDP-ketosugar and the process is reversed resulting in regeneration of the PLP-lysine internal aldimine and the transaminated product (Scheme 12B).

## 5. Additional sugar modification

### 5.1. Sugar *O/N/C*-methyltransferases

Sugar methyltransferases (SMTs) (EC 2.1.1.-) catalyze the transfer of a methyl group from *S*-adenosylmethionine (SAM) to a sugar-based nucleophile (typically *N*-, *C*-, *S*- and *O*). In natural product biosynthesis, sugar *C*- and *N*-methylation typically occurs at the NDP-sugar stage (*i.e.*, prior to glycosyltransferase-catalyzed transfer of the modified sugar to an aglycon) while sugar *O*-methylation commonly occurs post glycosylation (*i.e.*, upon the final natural product glycoside)<sup>14</sup>. The structures for a range of SMTs have emerged in recent years including: RebM [the rebeccamycin sugar 4-*O*-methyltransferase from *Lechevalieria aerocolonigenes*<sup>158</sup>]; NovP [the novobiocin sugar 4-*O*-methyltransferase from *Streptomyces spheroids*<sup>159</sup>]; MycE [the mycinamicin 2-*O*-methyltransferase from

*Micromonospora griseorubida*<sup>160</sup>]; TylM1 [the tylosin associated C3-*N,N*-dimethyltransferase involved in the production of dTDP-D-mycaminose from *Streptomyces fradiae*<sup>161</sup>]; DesVI [the pikromycin C3-*N,N*-dimethyltransferase involved in the production of dTDP-D-desosamine from *Streptomyces venezuelae*<sup>162</sup>]; and TcaB9 [the kijanimicin affiliated C3-methyltransferase involved in the biosynthesis of D-tetronitrose from *Micromonospora chalcea*<sup>163</sup>].

Structurally, SMTs are classified as Class I MTs<sup>164, 165</sup> with some of them being monomeric (Figure 14C)<sup>159, 161, 163</sup>, dimeric (Figure 14A and 14B)<sup>158, 162</sup> and even tetrameric<sup>160</sup>. Despite their low sequence homology, all members of this family share a core Rossmann fold domain responsible for binding SAM comprising a seven stranded mixed  $\beta$ -sheet in the order 3214576 flanked by  $\alpha$ -helices (Figure 14). Most SMTs contain auxiliary domains that are inserted throughout the core MT fold (Figure 14), the size and nature of which varies among the MTs and often contributes to substrate recognition. Examples include insertions at the N-terminus prior to the core domain, after strand  $\beta$ 5, and between the strands  $\beta$ 6 and  $\beta$ 7. Sometimes, these auxiliary domains can form a 'lid' structure responsible for substrate binding and can also contribute to subunit-subunit interaction. The monomer/monomer interface among SMTs can vary. For example, the dimer interface in RebM (Figure 14A) is formed by reciprocal interactions between strands  $\beta$ 6 from both monomers while in DesVI and TylM1 the dimerization is mediated by interaction of all four  $\beta$ -strands of the auxiliary domain (Figure 14B).

Most SMT structures to date are cofactor-bound with SAH or SAM occupying the core Rossmann fold<sup>164, 165</sup>. The SAM binding site is generally populated by hydrophobic residues. The amino acid portion of SAM interacts through the glycine-rich sequence E/D-X-G-X-G-X-G (D67-V-G69-C-G71-I-G73 in RebM). A loop following the strand  $\beta$ 1, and an acidic loop after the strand  $\beta$ 2, interacts with the ribose hydroxyls. The catalytic base histidine/aspartate (H140/H225 in RebM/MycE and D198 in NovP) responsible for deprotonation is located on a  $3_{10}$ -helix (or sometimes just a loop) just after the  $\beta$ 4-strand (Figures 14D & 14E). Substrate interaction is mediated by the secondary structure insertions after strand  $\beta$ 5 and between the strands  $\beta$ 6 and  $\beta$ 7 - the size and type of which dictates the substrate specificity. Though structures bound to the substrate/product are available<sup>160, 163</sup>, comparatively little is known about the structural factors which determine the regio-, stereo- and/or nucleophile-specificity of SMT-catalyzed reactions.

Typically, MTs catalyze  $S_N2$ -like reactions, with inversion of methyl stereochemistry, involving oxygen-, nitrogen-, and carbon-based nucleophiles that require at least one proton transfer step prior to, in concert with, or after methyl group transfer. In *O*-SMTs, general acid/base catalysis contributes to rate acceleration wherein a catalytic base (H140/H278 in RebM/MycE and D198 in NovP) is involved in nucleophile deprotonation. With the exception of RebM, the SMTs discussed above require a divalent metal ion for catalysis with  $Mg^{+2}$  as the preferred metal. Molecular simulation and pH-rate studies suggest the divalent metal to function primarily to organize the substrate-binding site in SMTs, not as a general base<sup>166, 167</sup>. The catalytic mechanism for *C*-MT (TcaB9) is thought to proceed via a similar mechanism wherein an active-site base, H225 (Figure 14E), abstracts the sugar C3 proton to initiate C3 C-methylation<sup>168</sup>. Interestingly, this residue is strictly conserved among all C3 *C*-SMTs but less so in the corresponding *N*-SMTs<sup>161, 163</sup>. Hence in *N*-SMTs TylM1 and DesVI, it is speculated that the proton on the C3 amino group is transferred to the water molecules lining the active site pocket and that the catalysis proceeds via approximation<sup>161</sup> (Figure 14F).

## 5.2. Sugar *N*-Acyltransferases

*N*-acyltransferases (EC 2.3.1.-) catalyze *N*-acylation using acetyl-CoA as an acyl donor. The structures of NDP-aminosugar acyltransferases that have emerged thus far fall into two major structural families - the GCN5-related *N*-acetyltransferase (GNAT) and left-handed  $\beta$ -helix motif (L $\beta$ H) superfamilies. The current L $\beta$ H family members include: a bifunctional UDP-*N*-acetylglucosamine pyrophosphorylase/glucosamine-1-phosphate *N*-acetyltransferase (EC 2.3.1.157) GlmU from *Escherichia coli*, *Mycobacterium tuberculosis* and *Yersinia pestis*<sup>73, 74, 80, 84</sup>; QdtC (responsible for *N*-acetylation of dTDP-Quip3N in *Thermoanaerobacterium thermosaccharolyticum*)<sup>169</sup>; PglD (responsible for *N*-acetylation of UDP-QuiNAc4N in *Campylobacter jejuni*)<sup>170, 171</sup>; WlbB (EC 2.3.1.B6, an enzyme that catalyzes *N*-acetylation of UDP-GlcNAcNA in *Bordetella petrii*)<sup>172</sup>; and the *N*-acyltransferase AntD (which catalyzes the acylation of the C4 amino group of dTDP-4-amino-4,6-dideoxyglucose using 3-hydroxy-3-methylbutyryl-CoA in *Bacillus cereus* en route to dTDP-D-Antrose<sup>173</sup>). The sole GNAT example from NDP-sugar biosynthesis reported to date is the dTDP-fucosamine acetyltransferase WecD from *Escherichia coli*<sup>174</sup>.

**5.2.1. L $\beta$ H fold**—This L $\beta$ H superfamily structural signature is dominated by a stunning parallel  $\beta$ -helix with repeating isoleucine-rich hexapeptide motifs and rare left-handed crossover connections<sup>175</sup> (Figure 15A). Members are typically trimeric wherein each monomer is dominated by a left-handed  $\beta$ -helix motif with a variable number of turns. This motif is characterized by imperfect, tandem repeated copies of the six-residue sequence, [L-I-V]-[G-A-E-D]-X-X-[S-T-A-V]-X which directs folding of a structural domain. Normally the L $\beta$ H domain is accompanied by a second domain – e.g., a N-terminal pyrophosphorylase domain in GlmU<sup>74, 80</sup> (Figure 2D), a N-terminal  $\alpha/\beta$ -domain for PglD (Figure 15D), a C-terminal  $\alpha/\beta$ -domain (Figure 15C) for QdtC and a C-terminal  $\beta$ -sheet for WlbB (Figure 15B). The flat faces of the L $\beta$ H are parallel  $\beta$ -sheets formed by stacks of short untwisted parallel  $\beta$ -strands. These strands participate in classical hydrogen bonding interactions with adjacent  $\beta$ -strands and the active site is located within a cleft between the L $\beta$ H domains of two adjacent subunits (Figure 15B). Within the active site, acetyl-CoA is bound such that the pantetheine arm of the cofactor is arranged in an extended conformation and directed parallel to the 3-fold axis of the trimer (Figure 15A). The phosphoryl groups of the coenzyme project outward toward solvent, whereas the adenine ring, the pantothenate, and  $\beta$ -mercaptoethylamine units are buried within the trimer. The NDP-sugar is found deeper within the active where it adopts an extended conformation and abuts turns of the  $\beta$ -helix which orient the aminosugar for a direct S<sub>N</sub>2 attack upon the acetyl-CoA thioester. QdtC and WlbB bind their sugar nucleotides in a similar manner, primarily through interactions with the NDP and very minimal sugar contacts with the exception of a hydrogen bond between the sugar C3 amino group and a structurally conserved asparagine (N159 and N84 in QdtC and WlbB, respectively)<sup>169, 172</sup>. Consistent with this, QdtC has been found to acetylate both dTDP-3-amino-3,6-dideoxy-D-glucose (dTDP-D-Quip3N) and dTDP-3-amino-3,6-dideoxy- $\alpha$ -D-galactose (dTDP-D-Fucp3N)<sup>169</sup>. As expected, PglD binds its NDP-sugar substrate in a manner distinct from that observed for QdtC/WlbB to afford distinct regiospecificity.

Ternary structures of the ‘pre’ and ‘post’ reaction states of WlbB (namely, with acetyl-CoA and CoA) have revealed the cofactor sulfur atom to move 2 Å upon departure of the acetyl group. A mechanism consistent with this structural information, mutagenesis studies and kinetics has been proposed (Scheme 13)<sup>169, 172</sup>. In this proposed mechanism, the C3’ amino nitrogen of the substrate attacks the si face of the acetyl moiety of acetyl-CoA to produce a tetrahedral oxyanion intermediate which is stabilized and properly oriented by interaction with the structurally conserved N84 (N159 in QdtC and D94 in AntD). As the oxyanion intermediate collapses, the bond between the carbon and the sulfur of acetyl-CoA breaks

allowing the sulfur to shift position and function as a general base by accepting a proton from the C3' amino group. The proposed mechanism assumes that the amino group of the sugar binds in the active site in the unprotonated state and that the pKa of the CoA sulfur is sufficiently perturbed such that the sulfur can function as a general base<sup>173</sup>. Though H125 has been suggested to play a role in PglD catalysis<sup>170</sup>, the functional role of H125 remains open to debate.

**5.2.2. GNAT fold**—Members of the GNAT superfamily adopt an overall fold built around a central mixed  $\beta$ -sheet, which is typically composed of six  $\beta$ -strands<sup>176, 177</sup>. WecD is a dimeric enzyme (Figure 16A) with each monomer adopting the GNAT *N*-acetyltransferase fold (Figure 16B) composed of two  $\alpha/\beta$  domains. The N-terminal domain has a central, five-stranded mixed  $\beta$ -sheet structure and two  $\alpha$ -helices. The C-terminal domain is comprised of a seven-stranded mixed  $\beta$ -sheet and four  $\alpha$ -helices. The strands  $\beta$ 4 and  $\beta$ 10 are long and extend across both domains and, although the two domains are distinct in the WecD structure, the  $\beta$ -strands of both domains extend toward each other to form a continuous, highly concave 10-stranded  $\beta$ -sheet with all  $\alpha$ -helices, except  $\alpha$ 4 and  $\alpha$ 5, lining the outside of this sheet (Figure 16B). The dimer interface is created through interactions on the convex side of the  $\beta$ -sheet along strands  $\beta$ 4 and  $\beta$ 10 and also includes interactions from helices  $\alpha$ 3 and  $\alpha$ 7 (Figure 16A).

The acetyl-CoA binding site is formed by residues from strands  $\beta$ 8 and  $\beta$ 9, the loop linking  $\alpha$ 4 and  $\alpha$ 5, as well as contributions from the loop  $\beta$ 8- $\alpha$ 6 and from helix  $\alpha$ 7. The loop between  $\alpha$ 4 and  $\alpha$ 5 is longer in WecD than in other GNAT proteins and affects the size and shape of the substrate-binding site<sup>174</sup>. The NDP-sugar pyrophosphate forms hydrogen bonds to the backbone amide NH groups of G172 and G174 and to the side chain of R207. Interestingly, these two glycine residues, located in the loop region linking strand  $\beta$ 8 and helix  $\alpha$ 6, are also part of the R/Q-X-X-G-X-G/A segment of the GNAT motif. Stacking interactions between the thymine base of nucleotides and aromatic side chains within WecD also contribute to substrate binding<sup>174</sup>.

From the structure, WecD does not possess a residue that directly functions as a catalytic base. Various proposals have been put forward for how the substrate amino group is deprotonated in members of the GNAT family, including direct proton abstraction by an aspartate or glutamate residue via an activated water molecule or through a series of hydrogen-bonded water molecules that together form a “proton wire”. A structurally conserved active site tyrosine in many GNAT enzymes (Y208 in WecD) is positioned 3.0 Å from the sulfur atom of CoA. This residue has been suggested to stabilize and protonate the departing CoA thiolate anion and to assist in correctly orienting the acetyl group for transfer<sup>174</sup>.

### 5.3. Oxidoreductases (*N*-oxidases)

Nitro, nitroso, and hydroxyamino sugars are found appended to a variety of natural products and generally derive from *N*-oxidase catalyzed oxidation of an appropriate NDP-aminosugar precursor<sup>44</sup>. While hydroxylaminosugar formation en route to calicheamicin in *Micromonospora echinospora* is catalyzed by the P450 *N*-oxidase CalE10<sup>44, 178, 179</sup>, *N*-oxidation in most other comparators are believed to be mediated by flavin containing monooxygenases (FMOs). FMO examples include: EvcC from *Micromonospora carbonacea* var. *africana*<sup>180, 181</sup> that mediates oxidation of dTDP-L-epi-vancosamine to the corresponding nitroso sugar<sup>179, 182</sup>; RubN8 from *Streptomyces achromogenes*, involved in the biosynthesis of dTDP-D-rubranitrose and ultimately the antibiotic rubradirin<sup>183</sup>; and KijD3 from *Actinomadura kijaniata* involved in the biosynthesis of dTDP-D-kijanose en route to the antibiotic kijanimycin<sup>184</sup>. Recently, the structures of EvcC and KijD3 have

been determined to place these *N*-oxidases into the class D FMOs, a family structurally related to the acyl-CoA dehydrogenases<sup>182, 185</sup>. Studies to date suggest a stepwise process of successive oxidation [hydroxylaminosugar – nitrososugar – nitrosugar<sup>44, 185</sup>] and it has been speculated that the formation of a nitroso/nitro product also depends upon the presence of an activating sugar carbonyl group adjacent to the targeted amine.

EvdC and KijD3 are tetrameric with each monomer adopting a three-domain fold (Figure 17A) comprising a N-terminal five  $\alpha$ -helical bundle, an eight-stranded  $\beta$ -sheet, and a second five  $\alpha$ -helical bundle at the C-terminus (Figure 17C) - a fold reminiscent of the acyl-CoA dehydrogenase superfamily. Interactions between adjacent C-terminal helical domains contribute the predominate contacts mediating tetramerization while the active site cleft is located at the junction of three domains. Residues contributing to the presumed active site emanate from all three domains but predominantly from the N-terminal helical domain and loops of the  $\beta$ -sheet domain. The active site is built along the length of helix N $\alpha$ 4, with contributions from N $\alpha$ 1- N $\alpha$ 3, C $\alpha$ 4, and the loops L1, L3 and L5 (Figure 17C)<sup>182</sup>. The dTDP moiety is anchored to the protein via the side chains of E113, Q254, and R330 (Figure 17B). Based upon the structural differences observed between the KijD3-dTDP complex and apo-EvdC, it is speculated that the active site loops undergo some rearrangement upon substrate and/or cofactor binding<sup>182</sup>.

In the proposed reaction mechanism of EvdC<sup>179, 182</sup>, the four-electron flavin-mediated oxidation of an amine to a nitroso functional group involves a two-step process (A and B in Scheme 14) in which the amino sugar is first oxidized to the corresponding hydroxylamine via a classical flavin monooxygenase mechanism. In the second step, oxidation of the hydroxylamino sugar to nitroso sugar may be affected by an iterative oxidative process in which FAD would again be reduced by flavin reductase/NADPH prior to an additional round of oxidation.

#### 5.4. ArnA (Dehydrogenase/Decarboxylase and Formyltransferase)

ArnA is one of the enzymes involved in the biosynthetic pathway of lipid A-4-amino-4-deoxy-L-arabinose (Ara4N), a main component of the outer membrane of gram-negative bacteria lipopolysaccharide (LPS). Modification of lipid A with Ara4N allows Gram-negative bacteria to resist the antimicrobial activity of cationic antimicrobial peptides and antibiotics such as polymyxin<sup>186, 187</sup>. ArnA has two functionally independent domains - a dehydrogenase domain responsible for the NAD<sup>+</sup>-dependent oxidative decarboxylation of UDP-Glucuronic acid (UDP-GlcA)<sup>188</sup>, and a transformylase domain that formylates UDP-4-amino-4-deoxy-L-arabinose (UDP-Ara4N)<sup>189, 190</sup>. ArnA homologs - such CalS9 (calicheamicin biosynthesis, *Micromonospora echinospora*)<sup>191, 192</sup> and AtS9 (indolocarbazole AT2433 biosynthesis, *Actinomadura melliaura*)<sup>193</sup> - have been implicated in the biosynthesis of novel pentoses appended to natural products and recently have been confirmed to display similar dehydrogenase activity<sup>194</sup>. Structures of full length *Escherichia coli* ArnA and that of each individual domain, both which are also independently functional, have been elucidated<sup>188, 189, 195, 196</sup>.

ArnA is a hexamer or dimer of trimers when crystallized as the intact full two domain protein (Figure 18A). The hexamer adopts the shape of a three-blade propeller with the C-terminal decarboxylase domains forming the central core of the propeller and the N-terminal formyltransferase domains arranged on the periphery. The decarboxylase domains mediate most of the contacts between monomers within the hexamer and little interaction is observed between the dehydrogenase and formyltransferase domains. The intersubunit contacts are dominated by hydrogen bonding and ionic interactions.

**5.4.1. Dehydrogenase domain**—The dehydrogenase domain of ArnA belongs to the SDR superfamily (previously described in **sections 4.1.1, 4.2.2 & 4.3.1.**). Like other enzymes belonging to this family that oxidize the C4-OH of sugar nucleotides (**section 4.1.1.**), it has a bilobal structure consisting of a large N-terminus subdomain and a small C-terminus subdomain (Figure 18B). The N-terminus subdomain has a Rossmann fold with a classical glycine-rich NAD<sup>+</sup>-binding motif G-X-(X)-G-X-X-G represented by amino acids G322-V-(N)-G325-F-I-G328 and the characteristic signature sequence Y463-X-X-X-K467 that, along with S433, form a catalytic triad. The active site is located in the cleft formed between two domains and a comparison of structures of the ArnA dehydrogenase domain in the presence and absence of ligand reveals a striking conformational change. Specifically, the loop that blocks the NAD<sup>+</sup>-binding site in the absence of UDP-GlcA moves ~17 Å to ultimately trap UDP-GlcA and allow NAD<sup>+</sup> binding in the presence of the substrate <sup>195</sup>.

The first step of the ArnA-catalyzed reaction is expected to mirror other SDR enzymes where the structurally conserved catalytic triad (T432-Y463 and K467, Figure 18C) facilitates UDP-GlcA C4 oxidation. The resulting 4-keto intermediate is believed to be unstable and the equilibrium for C4 oxidation favors the starting material. While decarboxylation is not spontaneous <sup>189</sup>, irreversible decarboxylation helps drive the overall reaction (Scheme 15A). The identity and role of residues involved in the decarboxylation remains somewhat controversial with one study supporting conserved residues S433 and R619 (Figure 18C) to be important <sup>188</sup> and another noting conserved residues S433 and E434 (Figure 18C) as essential for oxidative decarboxylation of UDP-GlcUA (where E434 is proposed to function as a general base to deprotonate the sugar carboxylic acid) <sup>189</sup>.

**5.4.2. Formyltransferase domain**—The structure of formyltransferase domain consists of two subdomains - a N-terminal subdomain with a Rossmann fold and a C-terminal subdomain resembling an oligonucleotide/oligosaccharide binding (OB) fold (Figure 19A) <sup>196</sup>. The N-terminal subdomain, which contains the *N*-10-formyltetrahydrofolate binding site, is folded into a seven-stranded β-sheet (in the order 3214567 with strand-β6 oriented antiparallel) flanked by α-helices. The C-terminal subdomain has three large β-sheets (β8, β9 and β12) and two small β-sheets (β10 and β11) flanked by two α-helices. The enzyme has a conserved H104-X-S106-L107-L108-P109-X-X-X-G113 motif (reminiscent of *N*-10-formyltetrahydrofolate dehydrogenase FDH), where histidine, proline, and glycine are strictly conserved. This motif is located at the C-terminus of strand-β5 and in the loop following the strand-β5. A N102 H104 D140 triad has been identified as important for catalysis <sup>196</sup> wherein N102 is located on the strand-β5, His104 on the loop L5 (the loop after the strand β5) within the conserved motif at the beginning of the loop, and D140 is found on loop L6 (the loop after strand-β6). The side chain of the conserved H104 is stabilized by two hydrogen bonds (one by the side chain of S106 and another by a water molecule) that bring together all three catalytic residues (Figure 19B).

A mechanism has been put forth which invokes all three residues of the catalytic triad in catalysis <sup>196, 197</sup>. In the proposed mechanism (Scheme 15B), H104 and N102 activate the carbonyl carbon of the formyl group, which undergoes nucleophilic attack by the primary amine of UDP-Ara4N. The oxyanion of the putative tetrahedral intermediate is stabilized by H104 and N102. A water molecule, properly positioned in the active site by hydrogen bonding with the side chain of D140 (Figure 19B), works as a proton shuttle in mediating proton transfer from UDP-Ara4N to the nitrogen of the folate <sup>198</sup>. The proton transfer is followed by decomposition of the tetrahedral intermediate and release of the products UDP-Ara4FN and tetrahydrofolate.



## 5.5. Enolpyruvyltransferase (MurA)

Enolpyruvyltransferase (MurA, also called MurZ, EC 2.5.1.7) catalyzes the transfer of an enolpyruvyl group from phosphoenolpyruvate (PEP) to UDP-*N*-acetylglucosamine (UDP-GlcNAc) to form UDP-*N*-acetylglucosamine enolpyruvate (UNAGEP). Both of these sugars form part of peptidoglycan essential for the integrity of the bacterial cell wall<sup>199</sup>. One of the important naturally-occurring inhibitor of MurA is fosfomycin, which inhibits the enzyme by alkylating a catalytic cysteine residue<sup>200</sup>. To date, structures of apo-MurA and ligand-bound form have been reported (Figure 20)<sup>200–203</sup>.

The structure of MurA enzyme is folded into two globular domains (N- and C-terminal domains) linked together by two linker regions [L1 and L22 consisting of residues S19-G20-A21 and L229-P230-D231 respectively (Figure 20)] with the interface cleft composing the active site. The two domains have similar secondary structure, consisting of an  $\alpha\beta\alpha$  motif arranged in such a way that the  $\alpha$ -helices are surrounded by three mixed  $\beta$ -sheets exposed to the solvent. Several motifs containing conserved sequence L-X-X-L-G-A-Y-Z-Y (where Y= a polar residue and Z= a hydrophobic residue) have been found throughout the length of the enzyme<sup>200</sup> and a flexible ten amino acid-length loop, referred to as loop La (for active site), from the N-terminal domain (P112-P121 in *E. coli* and *E. cloacae*), moves toward the active site and closes the interdomain cleft like a lid upon ligand binding in the *E. coli* and *E. cloacae* enzymes<sup>200, 202</sup> (Figures 20A, 20B and 20C). This loop contains the conserved cysteine (Cys115 in MurA from *E. cloacae* and *E. coli*) critical for MurA activity and modified by fosfomycin<sup>200, 203–206</sup>, a residue which is replaced by aspartate in fosfomycin-resistant strains<sup>199, 204</sup>. In contrast, the active site loop conformational change has not been observed in *H. influenzae* MurA complex structures as both the ternary complex (with UDP-GlcNAc and fosfomycin) and binary complex (with UDP-GlcNAc) of *H. influenzae* MurA remain in a half open conformation [Figure 20B, <sup>203</sup>].

UDP-GlcNAc interacts with the enzyme mainly through the NDP pyrophosphate via hydrogen-bonding with residues from both domains. Conserved interactions include participation of K22 in the formation of covalent adducts with PEP and fosfomycin<sup>207, 208</sup>; C115 (Figure 20D) in the participation of catalysis and product release<sup>205</sup>; D305 in the final deprotonation from the C3 atom of the tetrahedral intermediate<sup>209</sup>; and D369 and L370 for specific interactions with fosfomycin (both residues mutated in resistant strains)<sup>210</sup>.

In the proposed mechanism based upon biochemical and structural studies (Scheme 16), MurA binds UDP-GlcNAc and PEP sequentially. Substrate binding facilitates a large conformational change (open to closed form) wherein the closed form brings C115 into close proximity of PEP to enable the reaction. C115 has been proposed to function as a general acid/base catalyst in protonating/deprotonating C3<sup>201, 204</sup> and/or participate in product release<sup>205</sup>. After the substrates bind to the active site, a proton is transferred to PEP resulting in the formation of an oxocarbenium ion, which is then attacked by the sugar C3 oxygen nucleophile. Subsequent elimination of phosphate from the tetrahedral intermediate leads to the final desired product<sup>199</sup>.

## 5.6. UDP-galactopyranose mutase

UDP-galactopyranose mutase (UGM, EC 5.4.99.9) is a flavoenzyme that catalyzes the reversible interconversion of UDP-galactopyranose (UDP-Galp) to UDP-galactofuranose (UDP-Galf). UDP-Galf is one of the main building blocks of the cell wall and extracellular matrix of many pathogenic bacteria, fungi, and protozoa and the absence of UGM in humans offers potential as a drug target<sup>211, 212, 213</sup>. Unlike flavin dependent oxidoreductases, the redox state of the flavin in UGM is unchanged upon product formation. Although it is known that flavin must be in a reduced state, the precise role of the flavin in catalysis

remains controversial. Studies determined that the reduced form of the flavin in UGM is anionic FADH (FAD<sub>red</sub>) rather than neutral FADH<sub>2</sub><sup>214, 215</sup>. Recently structures of *Escherichia coli*<sup>216</sup>, *Klebsiella pneumoniae*<sup>217, 218</sup>, *Mycobacterium tuberculosis*<sup>217</sup>, and *Deinococcus radiodurans*<sup>219, 220</sup> UGMs have emerged.

Bacterial UGMs are homodimeric enzymes belonging to a mixed  $\alpha/\beta$  class of proteins. Each monomer contains three domains. Domain 1 includes a Rossmann fold that binds flavin adenine dinucleotide (FAD), domain 2 is a bundle of  $\alpha$ -helices that forms a major interface for the monomer-monomer interactions, and domain 3 features a twisted six-stranded  $\beta$ -sheet, situated between domain 1 and domain 2 which participates in substrate binding (Figure 21A). The FAD isoalloxazine binds in the crevice between domains 1 and 3 and the substrate binding site is located in a cleft adjacent to the isoalloxazine ring of FAD (Figure 21A). Superposition of monomers indicates domain 2 moves with respect to domain 1 using domain 3 as a molecular hinge to close on the substrate during the catalytic cycle<sup>216</sup>. A mobile loop region is located at the entrance of substrate binding cleft, shown to close upon substrate binding. Structures of oxidized and reduced UGM in the presence of substrate show differential modes of substrate binding<sup>218, 221</sup>. Comparison of the oxidized and reduced UGM substrate complexes reveals that flavin reduction results in a translocation of the mobile loop approximately 4 Å toward substrate (Figures 21B) and a shift in the relative orientation of the flavin and the UDP-Galp substrate positions the C1 of the galactose moiety directly adjacent to the nucleophilic N5 of the flavin. A comparison of UGM complexes with UDP-Galp and UDP-Glc implicates the C4-OH of galactose to engage in a hydrogen bond with the C4 carbonyl of the reduced flavin within this complex. It has been speculated that this H-bond may play an important role in enzyme's ability to discriminate against UDP-Glc and may also provide a means to shuttle the proton from C4-OH to the nascent C5-OH after ring opening<sup>218</sup>.

The substrate and FAD are bound in the active site by a network of H-bonding and stacking interactions. Several conserved interactions have been demonstrated as essential for UGM activity<sup>219, 222</sup>. These include stacking interactions of uridine with an aromatic residue (Y155 in *K. pneumoniae*) and stabilization of the negatively charged diphosphate backbone of the sugar nucleotide substrate by H-bonding interactions from the side-chains of the conserved arginines and tyrosines (R174, R280 Y185, Y314 and Y349 in *K. pneumoniae*). Similarly, the sugar moiety of the substrate is stabilized through hydrogen bonding, some of which are mediated through water as shown in Figures 21C & D.

Studies indicate that the reaction mechanism of UGM proceeds via the formation of a covalent iminium intermediate (Scheme 17)<sup>215, 218, 223, 224</sup>. This enables opening of the sugar ring and subsequent ring contraction to the furanose form. In the final reaction step, UDP serves as a nucleophile to displace the flavin, releasing product (Scheme 17). Three mechanistic hypotheses have been proposed to explain the formation of the iminium adduct: *i*) S<sub>N</sub>2 attack by N5 of FAD<sub>red</sub> upon the anomeric carbon position of the substrate concerted with cleavage of the C1-OP $\beta$  bond (Scheme 17, **path A**); *ii*) a stepwise S<sub>N</sub>1-type substitution where elimination of UDP to produce an oxocarbenium intermediate precedes the nucleophilic attack by N5 of FAD (Scheme 17, **path B**); or *iii*) via covalent bond formation facilitated by single-electron transfer between a substrate and FAD<sub>red</sub> radical pair (Scheme 17, **path C**)<sup>215</sup>. Evidence for the participation of N5 in nucleophilic attack at C1 of the substrate to form the covalent adduct<sup>218, 224</sup> is provided by the close proximity of the anomeric carbon of the substrate and N5 of FAD<sub>red</sub> in the ligand-bound UGM structure<sup>218</sup>. Inability of 5-deaza-FAD to catalyze the reaction is also consistent with the proposed nucleophilic role of the flavin N5<sup>215, 225</sup>. The intermediacy of the covalent iminium intermediate is supported by the ability to trap a covalent adduct during turnover via hydride reduction<sup>218, 223</sup>. Indirect evidence for an oxocarbenium intermediate derives from a

significant rate reduction observed with UDP-[2-F] Galf<sup>226</sup> and the inability of UGM to displace UDP from the linear substrate analog UDP-galactitol<sup>227</sup>.

## 6. Perspectives

As highlighted within this review, a wide range of structures have emerged and the new available structural information has contributed, in many cases, to better understanding of mechanism and/or specificity. Consistent with the divergent nature of sugar nucleotide biosynthetic pathways that derive from a series common core transformations, the structural scaffolds that support these core chemistries also utilize a relatively small set of common structural folds. As a result, dramatic shifts in chemical mechanism can be accomplished through very subtle residue substitutions or small secondary structure alterations – highlighting the potential of structure-based engineering and/or directed evolution of existing scaffolds to access new chemistries. Despite the considerable increase in available structures relevant to sugar nucleotide biosynthesis in recent years, there remains a deficiency of structural and/or biochemical understanding regarding the putative protein-protein interactions within these pathways. Such interactions have been invoked to explain how highly unstable sugar nucleotides may be ‘tunneled’ from one enzyme to the next to avoid degradation and also put forth as a mechanism of ‘co-localization’ to avoid cross-talk among potential competing sugar nucleotide pathways within a cell. As the structural biology of individual catalysts within these pathways matures, the pursuit of such putative enzyme complexes may serve as one of the next frontiers in this exciting interdisciplinary field.

## Acknowledgments

The authors would like to thank all past and present members of the Thorson and Phillips laboratories for contributions relevant to this work. Support for this work was provided in part by NIH grants AI52218 (JST), CA84374 (JST), and PSI GM098248 (GNP and JST).

## References

1. Kilcoyne M, Joshi L. *Cardiovasc Hematol Agents, Med Chem.* 2007; 5:186–197. [PubMed: 17630944]
2. Kren V, Martinkova L. *Curr Med Chem.* 2001; 8:1303–1328. [PubMed: 11562268]
3. Kren V, Rezanka T. *FEMS Microbiol Rev.* 2008; 32:858–889. [PubMed: 18647177]
4. Wrodnigg TM, Sprenger FK. *Mini Rev Med Chem.* 2004; 4:437–459. [PubMed: 15134545]
5. Thorson JS, Hosted TJ, Jiang JQ, Biggins JB, Ahlert J. *Curr Org Chem.* 2001; 5:139–167.
6. Le GT, Abbenante G, Becker B, Grathwohl M, Halliday J, Tometzki G, Zuegg J, Meuterms W. *Drug Discov Today.* 2003; 8:701–709. [PubMed: 12927513]
7. Nicotra F, Cipolla L, La Ferla B, Airoidi C, Zona C, Orsato A, Shaikh N, Russo L. *J Biotechnol.* 2009; 144:234–241. [PubMed: 19539677]
8. Cipolla L, Araujo AC, Bini D, Gabrielli L, Russo L, Shaikh N. *Expert Opin Drug Discov.* 2010; 5:721–737. [PubMed: 22827796]
9. Cipolla L, La Ferla B, Airoidi C, Zona C, Orsato A, Shaikh N, Russo L, Nicotra F. *Future Med Chem.* 2010; 2:587–599. [PubMed: 21426009]
10. Butler MS. *J Nat Prod.* 2004; 67:2141–2153. [PubMed: 15620274]
11. Weymouth-Wilson AC. *Nat Prod Rep.* 1997; 14:99–110. [PubMed: 9149408]
12. Gantt RW, Peltier-Pain P, Thorson JS. *Nat Prod Rep.* 2011; 28:1811–1853. [PubMed: 21901218]
13. Thibodeaux CJ, Melancon CE, Liu HW. *Nature.* 2007; 446:1008–1016. [PubMed: 17460661]
14. Thibodeaux CJ, Melancon CE 3rd, Liu HW. *Angew Chem Int Ed Engl.* 2008; 47:9814–9859. [PubMed: 19058170]

15. Oh TJ, Mo SJ, Yoon YJ, Sohng JK. *J Microbiol Biotechnol.* 2007; 17:1909–1921. [PubMed: 18167436]
16. Blanchard S, Thorson JS. *Curr Opin Chem Biol.* 2006; 10:263–271. [PubMed: 16675288]
17. Volchegursky Y, Hu Z, Katz L, McDaniel R. *Mol Microbiol.* 2000; 37:752–762. [PubMed: 10972798]
18. Hutchinson CR. *Curr Opin Microbiol.* 1998; 1:319–329. [PubMed: 10066498]
19. Butler AR, Bate N, Kiehl DE, Kirst HA, Cundliffe E. *Nat Biotechnol.* 2002; 20:713–716. [PubMed: 12089557]
20. Hutchinson E, Murphy B, Dunne T, Breen C, Rawlings B, Caffrey P. *Chem Biol.* 2010; 17:174–182. [PubMed: 20189107]
21. Pageni BB, Oh TJ, Liou K, Yoon YJ, Sohng JK. *J Microbiol Biotechnol.* 2008; 18:88–94. [PubMed: 18239422]
22. Salas JA, Mendez C. *Trends Microbiol.* 2007; 15:219–232. [PubMed: 17412593]
23. Yang J, Hoffmeister D, Liu L, Fu X, Thorson JS. *Bioorg Med Chem.* 2004; 12:1577–1584. [PubMed: 15112655]
24. Monica MP. *Curr Opin Chem Biol.* 2011; 15:226–233. [PubMed: 21334964]
25. Erb A, Weiss H, Harle J, Bechthold A. *Phytochem.* 2009; 70:1812–1821.
26. Williams GJ, Gantt RW, Thorson JS. *Curr Opin Chem Biol.* 2008; 12:556–564. [PubMed: 18678278]
27. Luzhetskyy A, Mendez C, Salas JA, Bechthold A. *Curr Top Med Chem.* 2008; 8:680–709. [PubMed: 18473892]
28. Williams GJ, Zhang C, Thorson JS. *Nat Chem Biol.* 2007; 3:657–662. [PubMed: 17828251]
29. Fu X, Albermann C, Zhang C, Thorson JS. *Org Lett.* 2005; 7:1513–1515. [PubMed: 15816740]
30. Bowles D, Isayenkova J, Lim EK, Poppenberger B. *Curr Opin Plant Biol.* 2005; 8:254–263. [PubMed: 15860422]
31. Breton C, Šnajdrová L, Jeanneau C, Ko a J, Imberty A. *Glycobiology.* 2006; 16:29R–37R. [PubMed: 16049187]
32. Xiaoqiang W. *FEBS Lett.* 2009; 583:3303–3309. [PubMed: 19796637]
33. Pesnot T, Jorgensen R, Palcic MM, Wagner GK. *Nat Chem Biol.* 2010; 6:321–323. [PubMed: 20364127]
34. Wagner GK, Pesnot T. *Chembiochem.* 2010; 11:1939–1949. [PubMed: 20672277]
35. Chang A, Singh S, Phillips GN Jr, Thorson JS. *Curr Opin Biotechnol.* 2011; 22:800–808. [PubMed: 21592771]
36. Liu HW, Thorson JS. *Annu Rev Microbiol.* 1994; 48:223–256. [PubMed: 7826006]
37. Shackelford GS, Regni CA, Beamer LJ. *Protein Sci.* 2004; 13:2130–2138. [PubMed: 15238632]
38. Mouilleron S, Badet-Denisot MA, Badet B, Golinelli-Pimpaneau B. *Arch Biochem Biophys.* 2011; 505:1–12. [PubMed: 20709015]
39. Sellick CA, Campbell RN, Reece RJ. *Int Rev Cell Mol Biol.* 2008; 269:111–150. [PubMed: 18779058]
40. Holden HM, Thoden JB, Timson DJ, Reece RJ. *Cell Mol Life Sci.* 2004; 61:2471–2484. [PubMed: 15526155]
41. Thoden JB, Holden HM. *J Biol Chem.* 2005; 280:21900–21907. [PubMed: 15795221]
42. Thoden JB, Timson DJ, Reece RJ, Holden HM. *J Biol Chem.* 2005; 280:9662–9670. [PubMed: 15590630]
43. Tanner ME. *Curr Org Chem.* 2001; 5:169–192.
44. Timmons SC, Thorson JS. *Curr Opin Chem Biol.* 2008; 12:297–305. [PubMed: 18424273]
45. Holden HM, Cook PD, Thoden JB. *Curr Opin Struct Biol.* 2010; 20:543–550. [PubMed: 20832292]
46. Romo AJ, Liu HW. *Biochim Biophys Acta.* 2011; 1814:1534–1547. [PubMed: 21315852]
47. He XM, Liu HW. *Annu Rev Biochem.* 2002; 71:701–754. [PubMed: 12045109]
48. Field RA, Naismith JH. *Biochemistry.* 2003; 42:7637–7647. [PubMed: 12820872]

49. Naismith JH. *Chem Soc Rev.* 2006; 35:763–770. [PubMed: 16936924]
50. Naismith JH. *Biochem Soc Trans.* 2004; 32:647–654. [PubMed: 15493979]
51. Beyer N, Alam J, Hallis TM, Guo Z, Liu HW. *J Am Chem Soc.* 2003; 125:5584–5585. [PubMed: 12733868]
52. Cook PD, Thoden JB, Holden HM. *Protein Sci.* 2006; 15:2093–2106. [PubMed: 16943443]
53. Smith P, Szu PH, Bui C, Liu HW, Tsai SC. *Biochemistry.* 2008; 47:6329–6341. [PubMed: 18491919]
54. Blankenfeldt W, Asuncion M, Lam JS, Naismith JH. *EMBO J.* 2000; 19:6652–6663. [PubMed: 11118200]
55. Allard ST, Beis K, Giraud MF, Hegeman AD, Gross JW, Wilmouth RC, Whitfield C, Graninger M, Messner P, Allen AG, Maskell DJ, Naismith JH. *Structure.* 2002; 10:81–92. [PubMed: 11796113]
56. Liu Y, Thoden JB, Kim J, Berger E, Gulick AM, Ruzicka FJ, Holden HM, Frey PA. *Biochemistry.* 1997; 36:10675–10684. [PubMed: 9271498]
57. Gerratana B, Cleland WW, Frey PA. *Biochemistry.* 2001; 40:9187–9195. [PubMed: 11478886]
58. Cook PD, Kubiak RL, Toomey DP, Holden HM. *Biochemistry.* 2009; 48:5246–5253. [PubMed: 19402712]
59. Barton WA, Lesniak J, Biggins JB, Jeffrey PD, Jiang J, Rajashankar KR, Thorson JS, Nikolov DB. *Nat Struct Biol.* 2001; 8:545–551. [PubMed: 11373625]
60. Hoffmeister D, Thorson JS. *Chembiochem.* 2004; 5:989–992. [PubMed: 15239057]
61. Moretti R, Thorson JS. *J Biol Chem.* 2007; 282:16942–16947. [PubMed: 17434871]
62. Allard ST, Giraud MF, Naismith JH. *Cell Mol Life Sci.* 2001; 58:1650–1665. [PubMed: 11706991]
63. Samuel J, Tanner ME. *Nat Prod Rep.* 2002; 19:261–277. [PubMed: 12137277]
64. Dong C, Beis K, Giraud MF, Blanchard S, Allard ST, Major LL, Kerr ID, Whitfield C, Naismith JH. *Biochem Soc Trans.* 2003; 31:532–536. [PubMed: 12773151]
65. Zuccotti S, Zanardi D, Rosano C, Sturla L, Tonetti M, Bolognesi M. *J Mol Biol.* 2001; 313:831–843. [PubMed: 11697907]
66. Thoden JB, Holden HM. *Protein Sci.* 2007; 16:1379–1388. [PubMed: 17567737]
67. Thoden JB, Holden HM. *Protein Sci.* 2007; 16:432–440. [PubMed: 17322528]
68. Thoden JB, Ruzicka FJ, Frey I, Rayment PA, Holden HM. *Biochemistry.* 1997; 36:1212–1222. [PubMed: 9063869]
69. Kim H, Choi J, Kim T, Lokanath N, Ha S, Suh S, Hwang HY, Kim K. *Mol Cells.* 2010; 29:397–405. [PubMed: 20238176]
70. Kim H, Wu CA, Kim DY, Han YH, Ha S, Suh S, Kim K. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:1447–1449. [PubMed: 15272173]
71. Brown K, Pompeo F, Dixon S, Mengin-Lecreux D, Cambillau C, Bourne Y. *EMBO J.* 1999; 18:4096–4107. [PubMed: 10428949]
72. Kostrewa D, D'Arcy A, Takacs B, Kamber M. *J Mol Biol.* 2001; 305:279–289. [PubMed: 11124906]
73. Verma SK, Jaiswal M, Kumar N, Parikh A, Nandicoori VK, Prakash B. *Acta Crystallogr F.* 2009; 65:435–439.
74. Zhang Z, Bulloch EM, Bunker RD, Baker EN, Squire CJ. *Acta Crystallogr D Biol Crystallogr.* 2009; 65:275–283. [PubMed: 19237750]
75. Mochalkin I, Lightle S, Zhu Y, Ohren JF, Spessard C, Chirgadze NY, Banotai C, Melnick M, McDowell L. *Protein Sci.* 2007; 16:2657–2666. [PubMed: 18029420]
76. Koropatkin NM, Holden HM. *J Biol Chem.* 2004; 279:44023–44029. [PubMed: 15292268]
77. Koropatkin NM, Cleland WW, Holden HM. *J Biol Chem.* 2005; 280:10774–10780. [PubMed: 15634670]
78. Pelissier MC, Lesley SA, Kuhn P, Bourne Y. *J Biol Chem.* 2010; 285:27468–27476. [PubMed: 20573954]

79. Cupp-Vickery JR, Igarashi RY, Perez M, Poland M, Meyer CR. *Biochemistry*. 2008; 47:4439–4451. [PubMed: 18355040]
80. Olsen LR, Vetting MW, Roderick SL. *Protein Sci*. 2007; 16:1230–1235. [PubMed: 17473010]
81. McCoy JG, Bitto E, Bingman CA, Wesenberg GE, Bannen RM, Kondrashov DA, Phillips GN Jr. *J Mol Biol*. 2007; 366:830–841. [PubMed: 17178129]
82. Jin X, Ballicora MA, Preiss J, Geiger JH. *EMBO J*. 2005; 24:694–704. [PubMed: 15692569]
83. Giraud MF, Naismith JH. *Curr Opin Struct Biol*. 2000; 10:687–696. [PubMed: 11114506]
84. Sivaraman J, Sauve V, Matte A, Cygler M. *J Biol Chem*. 2002; 277:44214–44219. [PubMed: 12171937]
85. Mizanur RM, Zea CJ, Pohl NL. *J Am Chem Soc*. 2004; 126:15993–15998. [PubMed: 15584733]
86. Allard ST, Giraud MF, Whitfield C, Graninger M, Messner P, Naismith JH. *J Mol Biol*. 2001; 307:283–295. [PubMed: 11243820]
87. Allard ST, Cleland WW, Holden HM. *J Biol Chem*. 2004; 279:2211–2220. [PubMed: 14570895]
88. Vogan EM, Bellamacina C, He X, Liu HW, Ringe D, Petsko GA. *Biochemistry*. 2004; 43:3057–3067. [PubMed: 15023057]
89. Koropatkin NM, Holden HM. *Acta Crystallogr D Biol Crystallogr*. 2005; 61:365–373. [PubMed: 15805590]
90. Fruscione F, Sturla L, Duncan G, Van Etten JL, Valbuzzi P, De Flora A, Di Zanni E, Tonetti M. *J Biol Chem*. 2008; 283:184–193. [PubMed: 17974560]
91. Webb NA, Mulichak AM, Lam JS, Rocchetta HL, Garavito RM. *Protein Sci*. 2004; 13:529–539. [PubMed: 14739333]
92. Somoza JR, Menon S, Schmidt H, Joseph-McCarthy D, Dessen A, Stahl ML, Somers WS, Sullivan FX. *Structure*. 2000; 8:123–135. [PubMed: 10673432]
93. Ishiyama N, Creuzenet C, Miller WL, Demendi M, Anderson EM, Harauz G, Lam JS, Berghuis AM. *J Biol Chem*. 2006; 281:24489–24495. [PubMed: 16651261]
94. Cook PD, Holden HM. *Biochemistry*. 2007; 46:14215–14224. [PubMed: 17997582]
95. Cook PD, Holden HM. *J Biol Chem*. 2008; 283:4295–4303. [PubMed: 18045869]
96. Gross JW, Hegeman AD, Gerratana B, Frey PA. *Biochemistry*. 2001; 40:12497–12504. [PubMed: 11601973]
97. Gross JW, Hegeman AD, Vestling MM, Frey PA. *Biochemistry*. 2000; 39:13633–13640. [PubMed: 11076501]
98. Melo A, Elliott WH, Glaser L. *J Biol Chem*. 1968; 243:1467. [PubMed: 4869560]
99. Gabriel O, Lindquis Lc. *J Biol Chem*. 1968; 243:1479–1484. [PubMed: 4869561]
100. Agnihotri G, Liu YN, Paschal BM, Liu HW. *Biochemistry*. 2004; 43:14265–14274. [PubMed: 15518577]
101. Johnson DA, Gassner GT, Bandarian V, Ruzicka FJ, Ballou DP, Reed GH, Liu HW. *Biochemistry*. 1996; 35:15846–15856. [PubMed: 8961949]
102. Miller VP, Thorson JS, Ploux O, Lo SF, Liu HW. *Biochemistry*. 1993; 32:11934–11942. [PubMed: 8218267]
103. Lo SF, Miller VP, Lei Y, Thorson JS, Liu HW, Schottel JL. *J Bacteriol*. 1994; 176:460–468. [PubMed: 8288541]
104. Thorson JS, Liu HW. *J Am Chem Soc*. 1993; 115:12177–12178.
105. Thorson JS, Liu HW. *J Am Chem Soc*. 1993; 115:7539–7540.
106. Thorson JS, Lo SF, Liu HW. *J Am Chem Soc*. 1993; 115:5827–5828.
107. Agarwal G, Rajavel M, Gopal B, Srinivasan N. *Plos One*. 2009; 4:e5736. [PubMed: 19478949]
108. Dunwell JM, Culham A, Carter CE, Sosa-Aguirre CR, Goodenough PW. *Trends Biochem Sci*. 2001; 26:740–746. [PubMed: 11738598]
109. Dong C, Major L, Allen A, Blankenfeldt W, Maskell D, Naismith J. *Structure*. 2003; 11:715–723. [PubMed: 12791259]
110. Giraud MF, Leonard GA, Field RA, Berling C, Naismith JH. *Nat Struct Biol*. 2000; 7:398–402. [PubMed: 10802738]

111. Kantardjieff KA, Kim CY, Naranjo C, Waldo GS, Lakin T, Segelke BW, Zemla A, Park MS, Terwilliger TC, Rupp B. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:895–902. [PubMed: 15103135]
112. Dong C, Major LL, Srikannathasan V, Errey JC, Giraud MF, Lam JS, Graninger M, Messner P, McNeil MR, Field RA, Whitfield C, Naismith JH. *J Mol Biol*. 2007; 365:146–159. [PubMed: 17046787]
113. Merkel AB, Major LL, Errey JC, Burkart MD, Field RA, Walsh CT, Naismith JN. *J Biol Chem*. 2004; 279:32684–32691. [PubMed: 15159413]
114. Merkel AB, Temple GK, Burkart MD, Losey HC, Beis K, Walsh CT, Naismith JH. *Acta Crystallogr D Biol Crystallogr*. 2002; 58:1226–1228. [PubMed: 12077451]
115. Jakimowicz P, Tello M, Meyers CLF, Walsh CT, Buttner MJ, Field RA, Lawson DM. *Proteins*. 2006; 63:261–265. [PubMed: 16411240]
116. Davis ML, Thoden JB, Holden HM. *J Biol Chem*. 2007; 282:19227–19236. [PubMed: 17459872]
117. Tello M, Jakimowicz P, Errey JC, Freel Meyers CL, Walsh CT, Buttner MJ, Lawson DM, Field RA. *Chem Commun*. 2006:1079–1081.
118. Rosano C, Bisso A, Izzo G, Tonetti M, Sturla L, De Flora A, Bolognesi M. *J Mol Biol*. 2000; 303:77–91. [PubMed: 11021971]
119. Rizzi M, Tonetti M, Vigevani P, Sturla L, Bisso A, Flora AD, Bordo D, Bolognesi M. *Structure*. 1998; 6:1453–1465. [PubMed: 9817848]
120. Somers WS, Stahl ML, Sullivan FX. *Structure*. 1998; 6:1601–1612. [PubMed: 9862812]
121. Bauer AJ, Rayment I, Frey PA, Holden HM. *Proteins*. 1992; 12:372–381. [PubMed: 1579570]
122. Thoden JB, Holden HM. *Biochemistry*. 1998; 37:11469–11477. [PubMed: 9708982]
123. Holden HM, Rayment I, Thoden JB. *J Biol Chem*. 2003; 278:43885–43888. [PubMed: 12923184]
124. Thoden JB, Heheman HD, Wesenberg G, Chapeau MC, Frey PA, Holden HM. *Biochemistry*. 1997; 199736:6294–6304. [PubMed: 9174344]
125. Read JA, Ahmed RA, Morrison JP, Coleman WG, Tanner ME. *J Am Chem Soc*. 2004; 126:8878–8879. [PubMed: 15264802]
126. Deacon AM, Ni YS, Coleman WG Jr, Ealick SE. *Structure*. 2000; 8:453–462. [PubMed: 10896473]
127. Shaik MM, Zanotti G, Cendron L. *Biochim Biophys Acta*. 2011; 1814:1641–1647. [PubMed: 21979583]
128. Kowatz T, Morrison JP, Tanner ME, Naismith JH. *Protein Sci*. 2010; 19:1337–1343. [PubMed: 20506248]
129. Koropatkin NM, Liu HW, Holden HM. *J Biol Chem*. 2003; 278:20874–20881. [PubMed: 12642575]
130. Bhatt VS, Guo CY, Guan W, Zhao G, Yi W, Liu ZJ, Wang PG. *Protein Sci*. 2011; 20:856–866. [PubMed: 21384454]
131. Campbell RE, Mosimann SC, Tanner ME, Strynadka NCJ. *Biochemistry*. 2000; 39:14993–15001. [PubMed: 11106477]
132. Velloso LM, Bhaskaran SS, Schuch R, Fischetti VA, Stebbins CE. *EMBO Rep*. 2008; 9:199–205. [PubMed: 18188181]
133. Glaser L. *Biochim Biophys Acta*. 1960; 41:534–536. [PubMed: 13828349]
134. Sommar KM, Ellis DB. *Biochim Biophys Acta*. 1972; 268:590–595. [PubMed: 5063602]
135. Blankenfeldt W, Kerr ID, Giraud MF, McMiken HJ, Leonard G, Whitfield C, Messner P, Graninger M, Naismith JH. *Structure*. 2002; 10:773–786. [PubMed: 12057193]
136. Kubiak RL, Holden HM. *Biochemistry*. 2011; 50:5905–5917. [PubMed: 21598943]
137. Jornvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D. *Biochemistry*. 1995; 34:6003–6013. [PubMed: 7742302]
138. Kingston RL, Scopes RK, Baker EN. *Structure*. 1996; 4:1413–1428. [PubMed: 8994968]
139. Thoden JB, Holden HM. *Biochemistry*. 2011; 50:1483–1491. [PubMed: 21241053]
140. Eliot AC, Kirsch JF. *Annu Rev Biochem*. 2004; 73:383–415. [PubMed: 15189147]
141. Paiardini A, Bossa F, Pascarella S. *Protein Sci*. 2004; 13:2992–3005. [PubMed: 15498941]

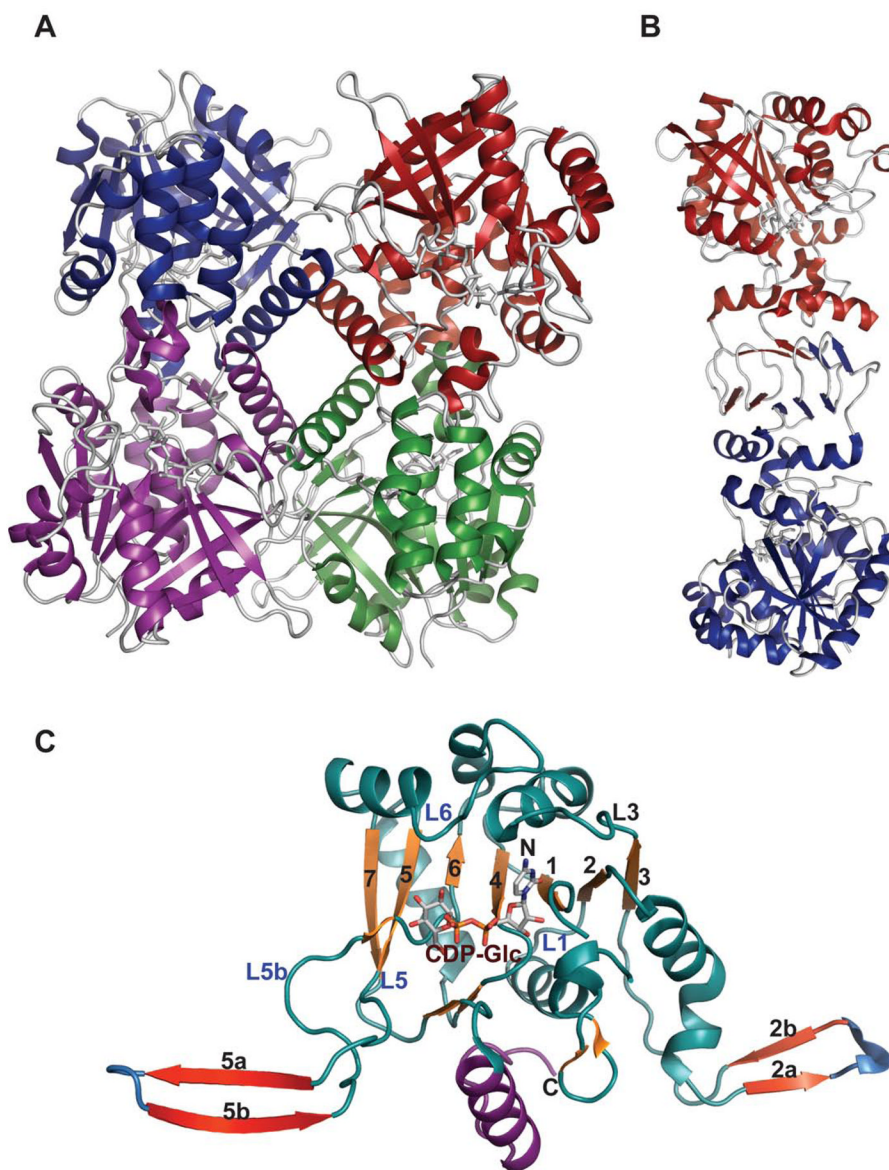
142. Cook PD, Holden HM. *Biochemistry*. 2008; 47:2833–2840. [PubMed: 18247575]
143. Burgie ES, Holden HM. *Biochemistry*. 2007; 46:8999–9006. [PubMed: 17630700]
144. Noland BW, Newman JM, Hendle J, Badger J, Christopher JA, Tresser J, Buchanan MD, Wright TA, Rutter ME, Sanderson WE, Muller-Dieckmann HJ, Gajiwala KS, Buchanan SG. *Structure*. 2002; 10:1569–1580. [PubMed: 12429098]
145. Schoenhofen IC, Lunin VV, Julien JP, Li Y, Ajamian E, Matte A, Cygler M, Brisson JR, Aubry A, Logan SM, Bhatia S, Wakarchuk WW, Young NM. *J Biol Chem*. 2006; 281:8907–8916. [PubMed: 16421095]
146. Badger J, Sauder JM, Adams JM, Antonysamy S, Bain K, Bergseid MG, Buchanan SG, Buchanan MD, Batiyenko Y, Christopher JA, Emtage S, Eroshkina A, Feil I, Furlong EB, Gajiwala KS, Gao X, He D, Hendle J, Huber A, Hoda K, Kearins P, Kissinger C, Laubert B, Lewis HA, Lin J, Loomis K, Lorimer D, Louie G, Maletic M, Marsh CD, Miller I, Molinari J, Muller-Dieckmann HJ, Newman JM, Noland BW, Pagarigan B, Park F, Peat TS, Post KW, Radojicic S, Ramos A, Romero R, Rutter ME, Sanderson WE, Schwinn KD, Tresser J, Winhoven J, Wright TA, Wu L, Xu J, Harris TJR. *Proteins: Struct, Funct, Bioinf*. 2005; 60:787–796.
147. Larkin A, Olivier NB, Imperiali B. *Biochemistry*. 2010; 49:7227–7237. [PubMed: 20604544]
148. Thoden JB, Schaffer C, Messner P, Holden HM. *Biochemistry*. 2009; 48:1553–1561. [PubMed: 19178182]
149. Burgie ES, Thoden JB, Holden HM. *Protein Sci*. 2007; 16:887–896. [PubMed: 17456741]
150. Cook PD, Carney AE, Holden HM. *Biochemistry*. 2008; 47:10685–10693. [PubMed: 18795799]
151. Schoenhofen IC, McNally DJ, Vinogradov E, Whitfield D, Young NM, Dick S, Wakarchuk WW, Brisson JR, Logan SM. *J Biol Chem*. 2006; 281:723–732. [PubMed: 16286454]
152. Jansonius JN. *Curr Opin Struct Biol*. 1998; 8:759–769. [PubMed: 9914259]
153. Soda K, Yoshimura T, Esaki N. *Chem Rec*. 2001; 1:373–384. [PubMed: 11933244]
154. Percudani R, Peracchi A. *EMBO Rep*. 2003; 4:850–854. [PubMed: 12949584]
155. Michael DT. *Arch Biochem Biophys*. 2005; 433:279–287. [PubMed: 15581583]
156. Johan NJ. *Curr Opin Struct Biol*. 1998; 8:759–769. [PubMed: 9914259]
157. Hayashi H, Kagamiyama H. *Biochemistry*. 1995; 34:9413–9423. [PubMed: 7626611]
158. Singh S, McCoy JG, Zhang C, Bingman CA, Phillips GN, Thorson JS. *J Biol Chem*. 2008; 283:22628–22636. [PubMed: 18502766]
159. Garcia IG, Stevenson CEM, Uson I, Meyers CLF, Walsh CT, Lawson DM. *J Mol Biol*. 2010; 395:390–407. [PubMed: 19857499]
160. Akey DL, Li S, Konwerski JR, Confer LA, Bernard SM, Anzai Y, Kato F, Sherman DH, Smith JL. *J Mol Biol*. 2011; 413:438–450. [PubMed: 21884704]
161. Carney AE, Holden HM. *Biochemistry*. 2011; 50:780–787. [PubMed: 21142177]
162. Burgie ES, Holden HM. *Biochemistry*. 2008; 47:3982–3988. [PubMed: 18327916]
163. Bruender NA, Thoden JB, Kaur M, Avey MK, Holden HM. *Biochemistry*. 2010; 49:5891–5898. [PubMed: 20527922]
164. Martin JL, McMillan FM. *Curr Opin Struct Biol*. 2002; 12:783–793. [PubMed: 12504684]
165. Schubert HL, Blumenthal RM, Cheng X. *Trends Biochem Sci*. 2003; 28:329–335. [PubMed: 12826405]
166. Thakker DR, Boehlert C, Kirk KL, Antkowiak R, Creveling CR. *J Biol Chem*. 1986; 261:178–184. [PubMed: 3753600]
167. Zheng YJ, Bruice TC. *J Am Chem Soc*. 1997; 119:8137–8145.
168. Chen H, Zhao Z, Hallis TM, Guo Z, Liu H-w. *Angew Chem Int Ed Engl*. 2001; 40:607–610. [PubMed: 11180386]
169. Thoden JB, Cook PD, Schaffer C, Messner P, Holden HM. *Biochemistry*. 2009; 48:2699–2709. [PubMed: 19191736]
170. Olivier NB, Imperiali B. *J Biol Chem*. 2008; 283:27937–27946. [PubMed: 18667421]
171. Rangarajan ES, Ruane KM, Sulea T, Watson DC, Proteau A, Leclerc S, Cygler M, Matte A, Young NM. *Biochemistry*. 2008; 47:1827–1836. [PubMed: 18198901]



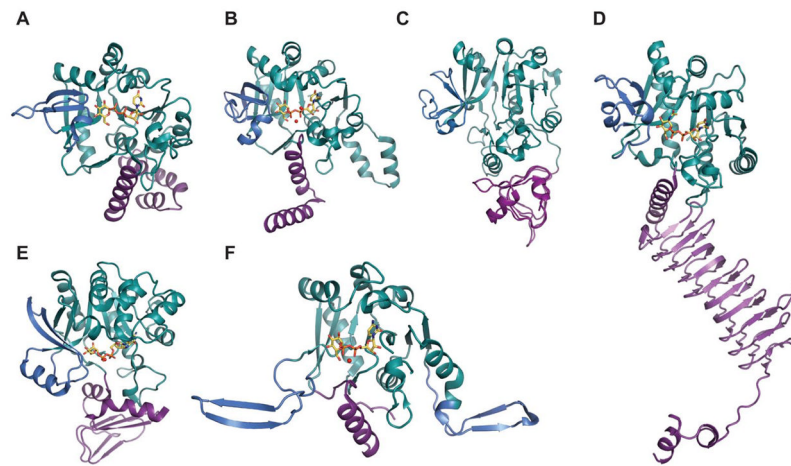
172. Thoden JB, Holden HM. *Biochemistry*. 2010; 49:4644–4653. [PubMed: 20433200]
173. Kubiak RL, Holden HM. *Biochemistry*. 2012; 51:867–878. [PubMed: 22220494]
174. Hung M-N, Rangarajan E, Munger C, Nadeau G, Sulea T, Matte A. *J Bacteriol*. 2006; 188:5606–5617. [PubMed: 16855251]
175. Raetz CR, Roderick SL. *Science*. 1995; 270:997–1000. [PubMed: 7481807]
176. Dyda F, Klein DC, Hickman AB. *Annu Rev Biophys Biomol Struct*. 2000; 29:81–103. [PubMed: 10940244]
177. Vetting MW, SdCLP, Yu M, Hegde SS, Magnet S, Roderick SL, Blanchard JS. *Arch Biochem Biophys*. 2005; 433:212–226. [PubMed: 15581578]
178. Johnson HD, Thorson JS. *J Am Chem Soc*. 2008; 130:17662–17663. [PubMed: 19055330]
179. Hu Y, Al-Mestarihi A, Grimes CL, Kahne D, Bachmann BO. *J Am Chem Soc*. 2008; 130:15756–15757. [PubMed: 18983146]
180. Cooper R, Horan AC, Gentile F, Gullo V, Loebenberg D, Marquez J, Patel M, Puar MS, Truumees I. *J Antibiot*. 1988; 41:13–19. [PubMed: 3346184]
181. Ganguly AK, Girjavallabhan VM, Miller GH, Sarre OZ. *J Antibiot*. 1982; 35:561–570. [PubMed: 7107521]
182. Vey JL, Al-Mestarihi A, Hu Y, Funk MA, Bachmann BO, Iverson TM. *Biochemistry*. 2010; 49:9306–9317. [PubMed: 20866105]
183. Kim CG, Lamichhane J, Song KI, Nguyen VD, Kim DH, Jeong TS, Kang SH, Kim KW, Maharjan J, Hong YS, Kang JS, Yoo JC, Lee JJ, Oh TJ, Liou K, Sohng JK. *Arch Microbiol*. 2008; 189:463–473. [PubMed: 18080113]
184. Zhang H, White-Phillip JA, Melancon CE 3rd, Kwon HJ, Yu WL, Liu HW. *J Am Chem Soc*. 2007; 129:14670–14683. [PubMed: 17985890]
185. Bruender NA, Thoden JB, Holden HM. *Biochemistry*. 2010; 49:3517–3524. [PubMed: 20334431]
186. Gunn JS, Ryan SS, Van Velkinburgh JC, Ernst RK, Miller SI. *Infect Immun*. 2000; 68:6139–6146. [PubMed: 11035717]
187. Gunn JS. *J Endotoxin Res*. 2001; 7:57–62. [PubMed: 11521084]
188. Gatzeva-Topalova PZ, May AP, Sousa MC. *Biochemistry*. 2004; 43:13370–13379. [PubMed: 15491143]
189. Williams GJ, Breazeale SD, Raetz CR, Naismith JH. *J Biol Chem*. 2005; 280:23000–23008. [PubMed: 15809294]
190. Breazeale SD, Ribeiro AA, McClerren AL, Raetz CR. *J Biol Chem*. 2005; 280:14154–14167. [PubMed: 15695810]
191. Bililign T, Shepard EM, Ahlert J, Thorson JS. *Chembiochem*. 2002; 3:1143–1146. [PubMed: 12404643]
192. Ahlert J, Shepard E, Lomovskaya N, Zazopoulos E, Staffa A, Bachmann BO, Huang K, Fonstein L, Czisny A, Whitwam RE, Farnet CM, Thorson JS. *Science*. 2002; 297:1173–1176. [PubMed: 12183629]
193. Gao Q, Zhang C, Blanchard S, Thorson JS. *Chem Biol*. 2006; 13:733–743. [PubMed: 16873021]
194. Simkhada D, Oh T-J, Pageni B, Lee H, Liou K, Sohng J. *Appl Microbiol Biotechnol*. 2009; 83:885–895. [PubMed: 19290519]
195. Gatzeva-Topalova PZ, May AP, Sousa MC. *Structure*. 2005; 13:929–942. [PubMed: 15939024]
196. Gatzeva-Topalova PZ, May AP, Sousa NC. *Biochemistry*. 2005; 44:5328–5338. [PubMed: 15807526]
197. Warren MS, Marolewski AE, Benkovic SJ. *Biochemistry*. 1996; 35:8855–8862. [PubMed: 8688421]
198. Klein C, Chen P, Arevalo JH, Stura EA, Marolewski A, Warren MS, Benkovic SJ, Wilson IA. *J Mol Biol*. 1995; 249:153–175. [PubMed: 7776369]
199. Gautam A, Rishi P, Tewari R. *Appl Microbiol Biotechnol*. 2011; 92:211–225. [PubMed: 21822642]

200. Skarzynski T, Mistry A, Wonacott A, Hutchinson SE, Kelly VA, Duncan K. *Structure*. 1996; 4:1465–1474. [PubMed: 8994972]
201. Skarzynski T, Kim DH, Lees WJ, Walsh CT, Duncan K. *Biochemistry*. 1998; 37:2572–2577. [PubMed: 9485407]
202. Schönbrunn E, Eschenburg S, Krekel F, Luger K, Amrhein N. *Biochemistry*. 2000; 39:2164–2173. [PubMed: 10694381]
203. Yoon H-J, Lee SJ, Mikami B, Park H-J, Yoo J, Suh SW. *Proteins: Struct, Funct, Bioinf*. 2008; 71:1032–1037.
204. Kim DH, Lees WJ, Kempell KE, Lane WS, Duncan K, Walsh CT. *Biochemistry*. 1996; 35:4923–4928. [PubMed: 8664284]
205. Eschenburg S, Priestman M, Schonbrunn E. *J Biol Chem*. 2005; 280:3757–3763. [PubMed: 15531591]
206. Jackson SG, Zhang F, Chindemi P, Junop MS, Berti PJ. *Biochemistry*. 2009; 48:11715–11723. [PubMed: 19899805]
207. Samland AK, Amrhein N, Macheroux P. *Biochemistry*. 1999; 38:13162–13169. [PubMed: 10529188]
208. Samland AK, Etezady-Esfarjani T, Amrhein N, Macheroux P. *Biochemistry*. 2001; 40:1550–1559. [PubMed: 11327813]
209. Eschenburg S, Kabsch W, Healy ML, Schönbrunn E. *J Biol Chem*. 2003; 278:49215–49222. [PubMed: 13129913]
210. Takahata S, Ida T, Hiraishi T, Sakakibara S, Maebashi K, Terada S, Muratani T, Matsumoto T, Nakahama C, Tomono K. *Int J Antimicrob Agents*. 2010; 35:333–337. [PubMed: 20071153]
211. Richards MR, Lowary TL. *Chembiochem*. 2009; 10:1920–1938. [PubMed: 19591187]
212. Peltier P, Euzen R, Daniellou R, Nugier-Chauvin C, Ferrières V. *Carbohydr Res*. 2008; 343:1897–1923. [PubMed: 18440497]
213. Pan F, Jackson M, Ma Y, McNeil M. *J Bacteriol*. 2001; 183:3991–3998. [PubMed: 11395463]
214. Zhang Q, Liu H-w. *J Am Chem Soc*. 2000; 122:9065–9070.
215. Sun HG, Ruzsyczky MW, Chang W-c, Thibodeaux CJ, Liu H-w. *J Biol Chem*. 2012; 287:4602–4608. [PubMed: 22187430]
216. Sanders DAR, Staines AG, McMahon SA, McNeil MR, Whitfield C, Naismith JH. *Nat Struct Mol Biol*. 2001; 8:858–863.
217. Beis K, Srikanthasan V, Liu H, Fullerton SWB, Bamford VA, Sanders DAR, Whitfield C, McNeil MR, Naismith JH. *J Mol Biol*. 2005; 348:971–982. [PubMed: 15843027]
218. Gruber TD, Westler WM, Kiessling LL, Forest KT. *Biochemistry*. 2009; 48:9171–9173. [PubMed: 19719175]
219. Partha SK, Sadeghi-Khomami A, Slowski K, Kotake T, Thomas NR, Jakeman DL, Sanders DAR. *J Mol Biol*. 2010; 403:578–590. [PubMed: 20850454]
220. Partha SK, van Straaten KE, Sanders DAR. *J Mol Biol*. 2009; 394:864–877. [PubMed: 19836401]
221. Gruber TD, Borrok MJ, Westler WM, Forest KT, Kiessling LL. *J Mol Biol*. 2009; 391:327–340. [PubMed: 19500588]
222. Chad JM, Sarathy KP, Gruber TD, Addala E, Kiessling LL, Sanders DAR. *Biochemistry*. 2007; 46:6723–6732. [PubMed: 17511471]
223. Soltero-Higgin M, Carlson EE, Gruber TD, Kiessling LL. *Nat Struct Mol Biol*. 2004; 11:539–543. [PubMed: 15133501]
224. Yuan Y, Bleile DW, Wen X, Sanders DAR, Itoh K, Liu H-w, Pinto BM. *J Am Chem Soc*. 2008; 130:3157–3168. [PubMed: 18278916]
225. Huang Z, Zhang Q, Liu H. *Bioorg Chem*. 2003; 31:494–502. [PubMed: 14613770]
226. Zhang Q, Liu H. *J Am Chem Soc*. 2001; 123:6756–6766. [PubMed: 11448178]
227. Itoh K, Huang Z, Liu HW. *Org Lett*. 2007; 9:879–882. [PubMed: 17266324]
228. Barton WA, Biggins JB, Jiang J, Thorson JS, Nikolov DB. *Proc Natl Acad Sci U S A*. 2002; 99:13397–13402.

229. Moretti R, Chang A, Peltier-Pain P, Bingman CA, Phillips GN Jr, Thorson JS. *J Biol Chem.* 2011; 286:13235–13243. [PubMed: 21317292]
230. Aragão D, Fialho AM, Marques AR, Mitchell EP, Sá-Correia I, Frazão C. *J Bacteriol.* 2007; 189:4520–4528. [PubMed: 17434970]
231. Sulzenbacher G, Gal L, Peneff C, Fassy F, Bourne Y. *J Biol Chem.* 2001; 276:11844–11851. [PubMed: 11118459]
232. Beis K, Allard ST, Hegeman AD, Murshudov G, Philp D, Naismith JH. *J Am Chem Soc.* 2003; 125:11872–11878. [PubMed: 14505409]
233. Morrison JP, Read JA, Coleman WG Jr, Tanner ME. *Biochemistry.* 2005; 44:5907–5915. [PubMed: 15823050]
234. Samuel J, Tanner ME. *Biochim Biophys Acta.* 2004; 1700:85–91.

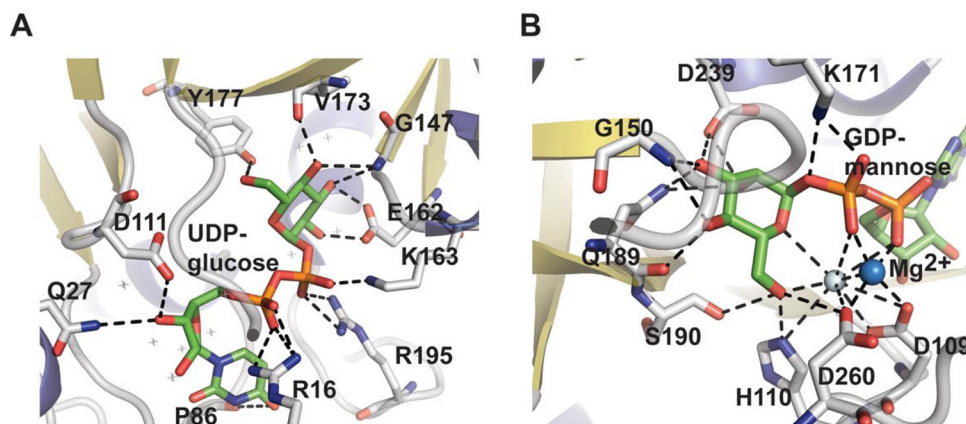


**Figure 1.** Representative global structural features of nucleotidyltransferases **A.** Tetramer of dTDP-glucose pyrophosphorylase (RmlA) from *Salmonella enterica* where each monomer is represented by a distinct color (PDB 1IIN). **B.** Dimer of GDP-mannose pyrophosphorylase (GMPase) from *Thermotoga maritime* where each monomer is represented by a distinct color. (PDB 2X5Z). **C.** Monomer of CDP-glucose pyrophosphorylase (CGPase) from *Salmonella typhi* (PDB 1TZF) with secondary structures of core  $\beta$ -sheet and variable regions numbered. The core N-terminal Rossmann fold is colored in green and orange ( $\beta$ -strands), the variable regions are colored blue and red, CDP-Glucose is colored grey, the C-terminal domain is purple and the letters 'C' and 'N' represent the C- and N-terminus of the enzyme, respectively.

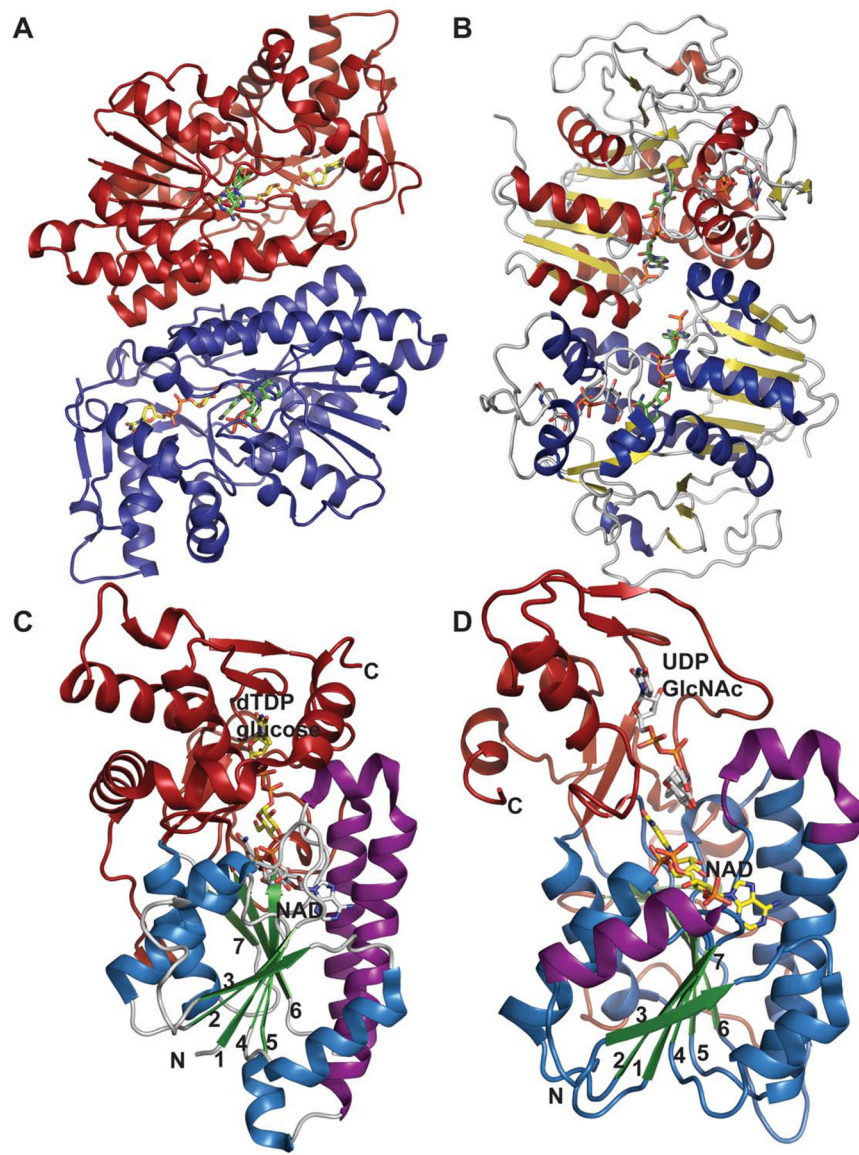


**Figure 2.**

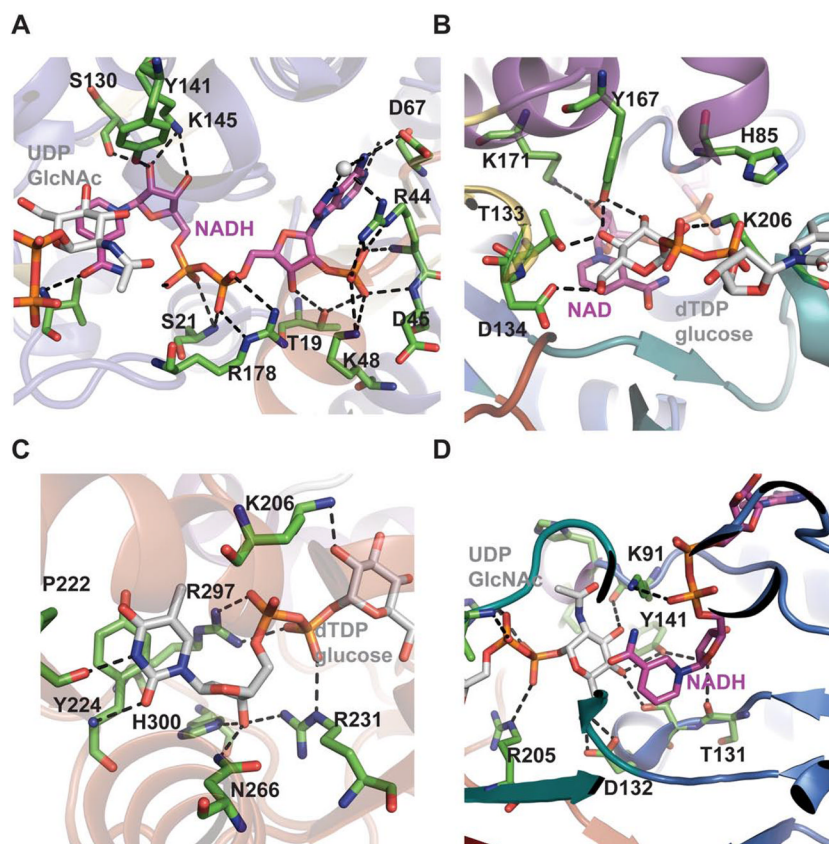
Representative monomeric features of nucleotidyltransferases. **A.** dTDP-glucose pyrophosphorylase RmlA from *Salmonella enterica* (PDB 1IIN). **B.** UDP-glucose pyrophosphorylase UGPase from *Escherichia coli* (PDB 2E3D). **C.** ADP-glucose pyrophosphorylase AGPase from *Agrobacterium tumefaciens* (PDB 3BRK). **D.** *N*-acetylglucosamine diphosphorylase uridylyltransferase GlmU from *Mycobacterium tuberculosis*. **E.** GDP-mannose pyrophosphorylase GMPase from *Thermotoga maritima* (PDB 2X5Z). **F.** CDP-glucose pyrophosphorylase CGPase from *Salmonella typhi* (PDB 1TZF). Rossmann fold is represented in green, variable regions colored blue, C-terminal domain in purple, ligands in yellow and Mg<sup>2+</sup> as red sphere.



**Figure 3.** Nucleotidyltransferase active-site interactions. **A.** dTDP-glucose pyrophosphorylase RmlA from *Salmonella enterica* bound to UDP-glucose (PDB 1IIN). **B.** GDP-mannose pyrophosphorylase GMPase from *Thermotoga maritime* bound to GDP-mannose (PDB 2X5Z). NDP-sugar is colored green, Mg<sup>2+</sup> is represented by a blue sphere, and water by a grey sphere.

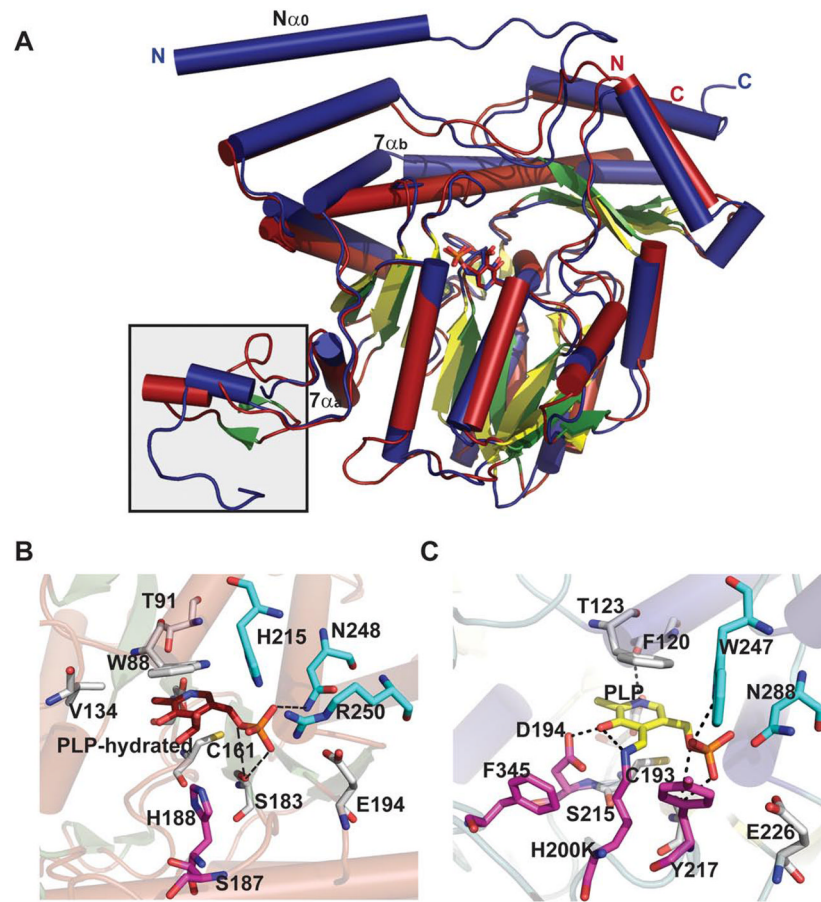


**Figure 4.** Representative global folds and monomeric structures of retaining and inverting dehydratases in the SDR structural family. Dimer (A) and monomer (C) of the retaining dehydratase RmlB bound to dTDP-Glucose and NAD from *Salmonella enterica* (PDB 1KEU). Dimer (B) and monomer (D) of the inverting dehydratase FlaA1 from *Helicobacter pylori* bound to UDP-GlcNAc and NADPH (PDB 2GN6). The core Rossmann fold is colored in blue/green, C-terminal domain highlighted in red, and purple signifies the helices involved in dimerization.

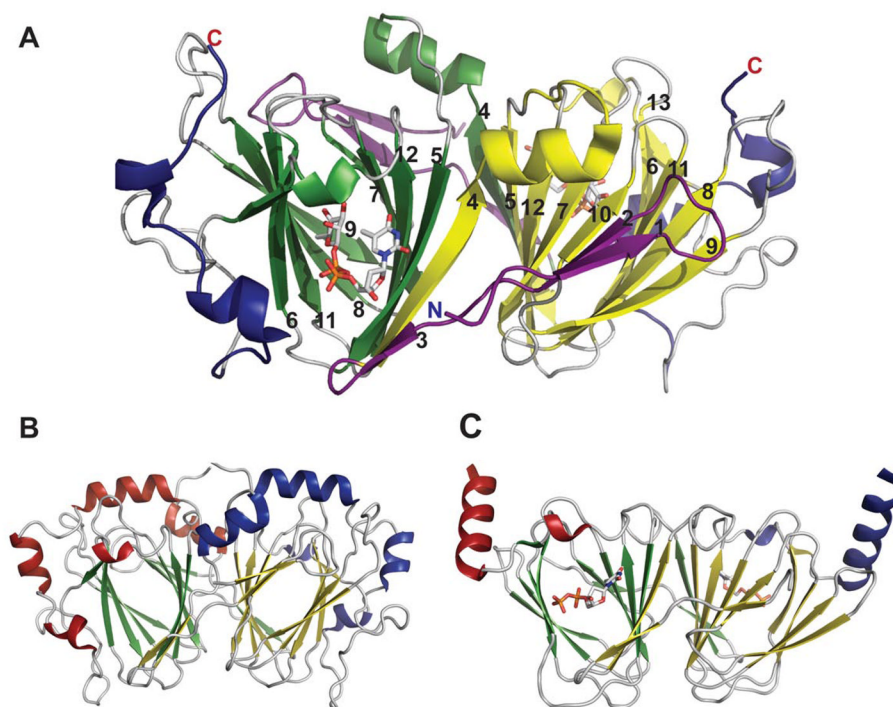


**Figure 5.** SDR structural family dehydratase active-site interactions. The cofactor binding sites of *Helicobacter pylori* FlaA1 bound to UDP-GlcNAc and NADP (PDB 2GN6) and *Salmonella enterica* RmlB bound to NAD<sup>+</sup> and dTDP-glucose (PDB 1KEU) are highlighted in panels (A) and (B), respectively. The corresponding sugar binding sites of FlaA1 and RmlB are highlighted in panels (C) and (D), respectively.

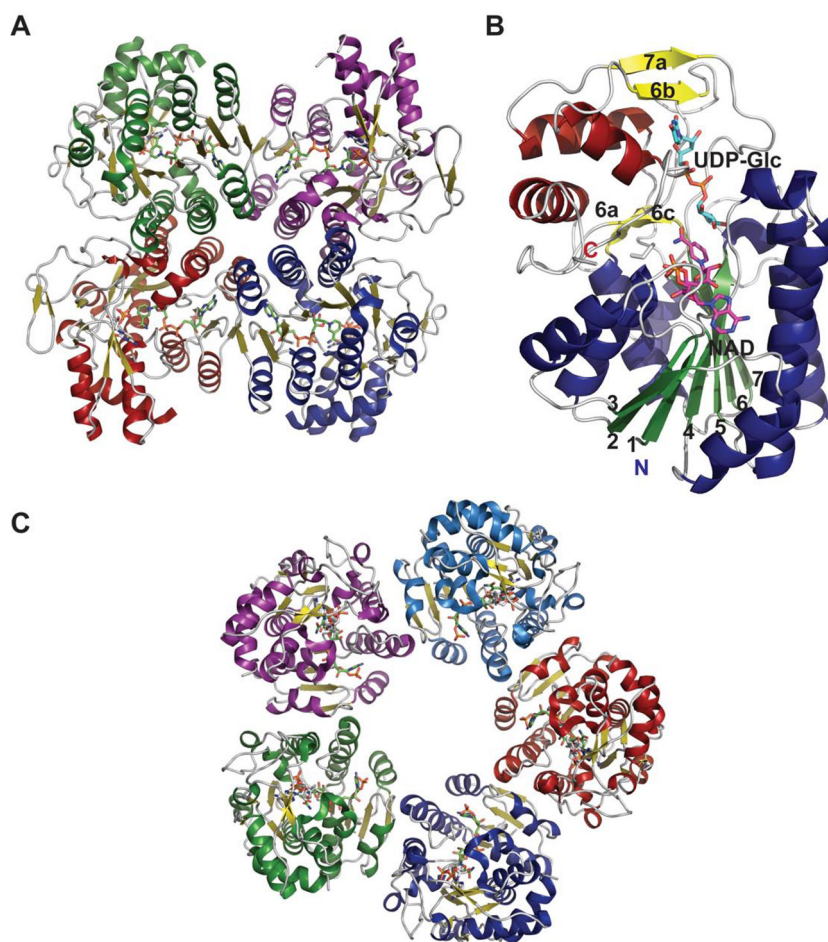




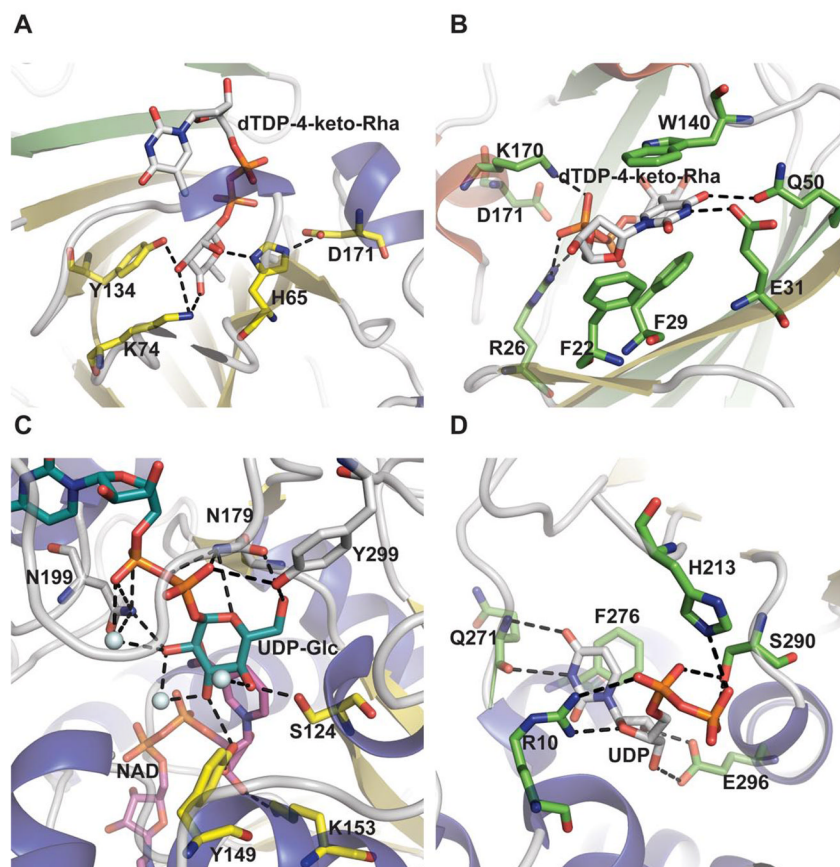
**Figure 6.** AAT structural family dehydrase global fold and active-site interactions. **A.** Overlay of ColD from *Escherichia coli* bound to hydrated PLP (red, PDB 2GMS) and E1 (H220K) from *Yersinia pseudotuberculosis* bound to PLP (blue, PDB 3BB8). The boxed flexible loop indicates the region that binds to the E1 [2Fe-2S] cluster. **B.** The active sites of ColD and E1 are represented in panels **B** and **C**, respectively. Letters ‘C’ and ‘N’ represent C and N-terminus, respectively.



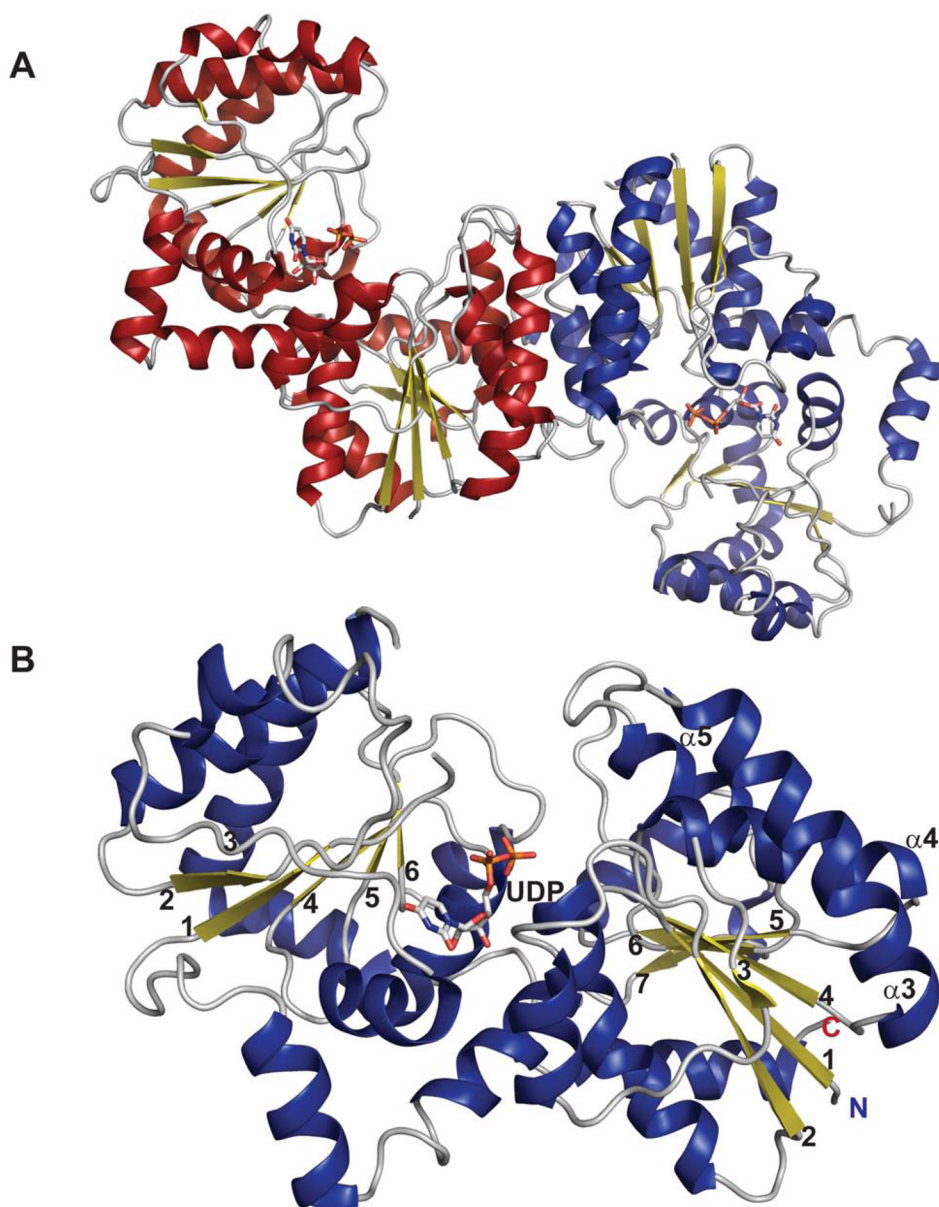
**Figure 7.** Global fold of epimerases belonging to cupin structural family. **A.** dTDP-4-dehydrorhamnose 3,5-epimerase RmlC from *Pseudomonas aeruginosa* bound to dTDP-4-keto-rhamnose (PDB 2IXK ). N-terminal domain colored purple, C-terminal domain colored blue, core-domain in yellow and green **B.** *Amycolatopsis orientalis* dTDP-3-amino-4-keto-2,3,6-trideoxy-3-*C*-methyl-glucose-5-epimerase EvaD from (PDB 1OFN). **C.** *Aneurinibacillus thermoaerophilus* dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase FdtA bound to dTDP (PDB 2PA7). In panels B and C, red/green represents one monomer while blue/yellow represents another.



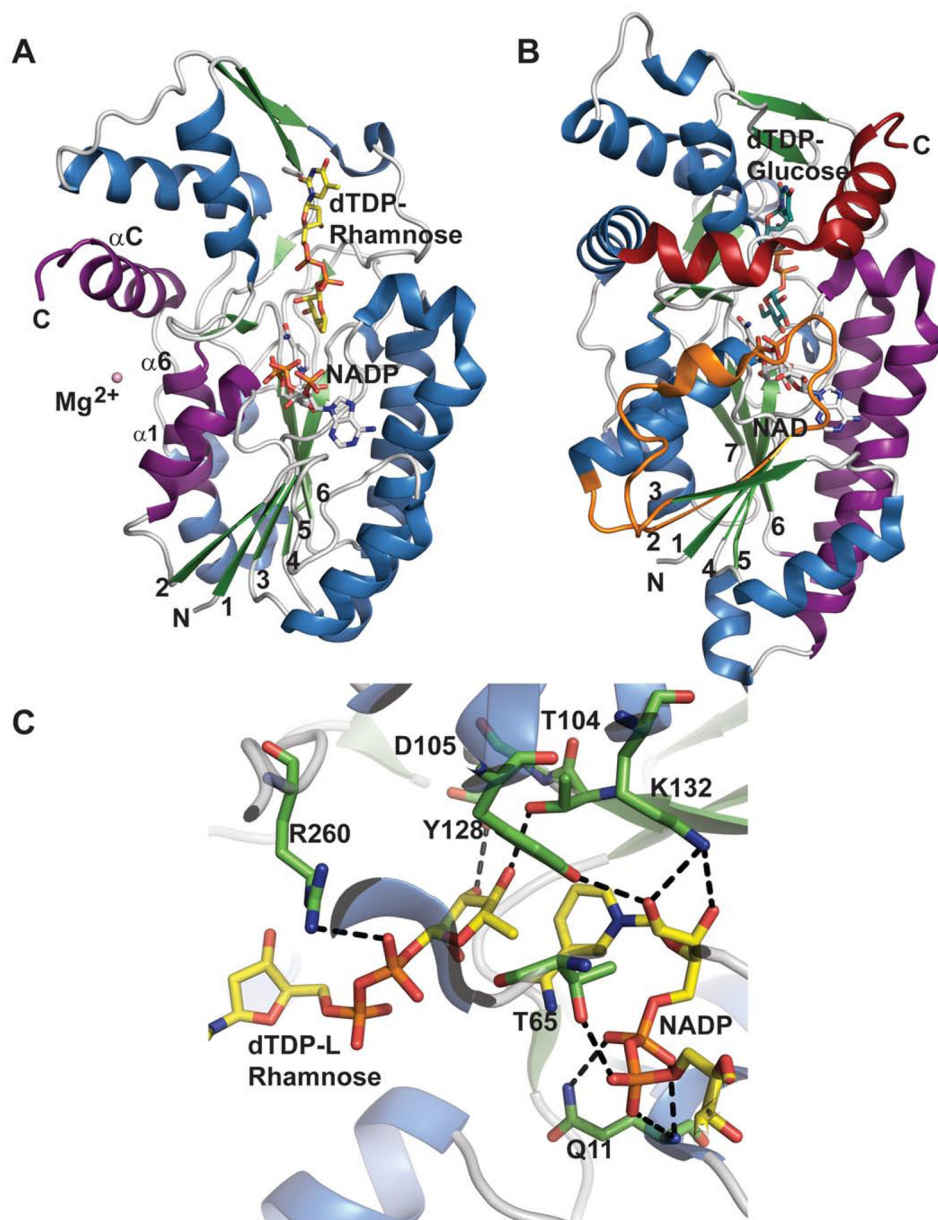
**Figure 8.** Global fold of epimerases belonging to SDR structural family. **A.** Tetramer of CDP-tyvelose-2-epimerase from *Salmonella enterica subsp. enterica serovar Typhi* bound to CDP and NAD (PDB 1ORR). **B.** ADP-L-glycero-D-mannoheptose-6-epimerase from *Escherichia coli* (PDB 1EQ2). **C.** UDP-galactose epimerase from *Escherichia coli* bound to UDP-Glc and NAD (PDB 1XEL). For panels **A** and **C**, each monomer is represented by a distinct color. In panel **B** the Rossmann fold is highlighted in blue/green and the C-terminal domain in red/yellow.



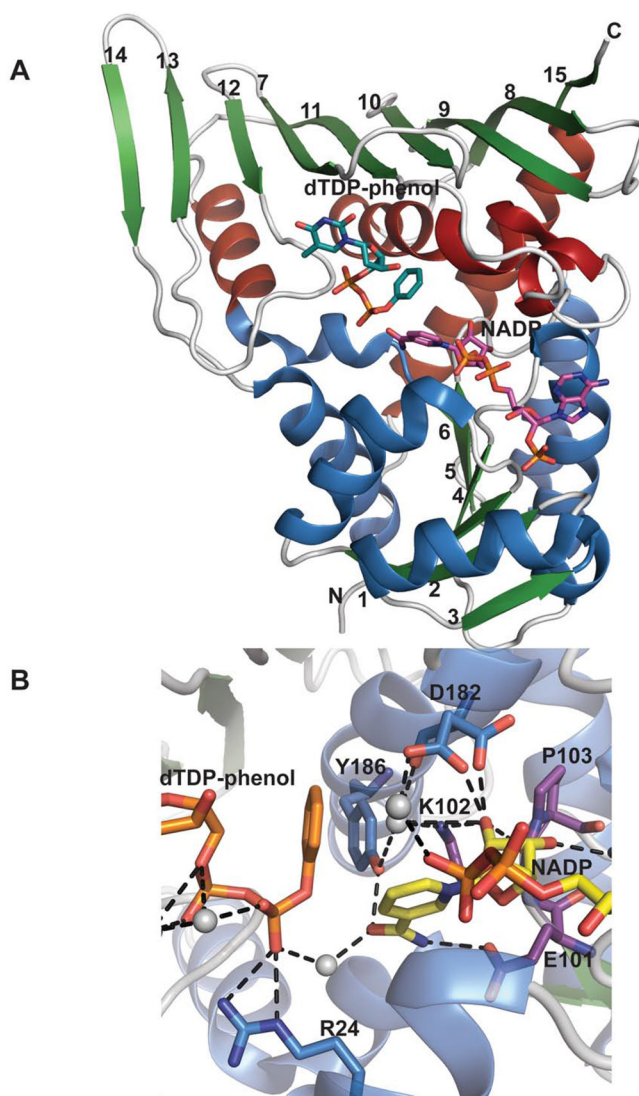
**Figure 9.** Epimerase active site architecture. **A.** Interaction of *Pseudomonas aeruginosa* dTDP-4-dehydrorhamnose 3,5-epimerase RmlC bound to dTDP-4-keto-rhamnose (PDB 2IXK ). **B.** Nucleotide base interactions of the RmlC ligand-bound complex. **C.** Interaction of sugar within *Escherichia coli* UDP-galactose-4-epimerase bound to UDP-Glc and NAD (PDB 1XEL). **D.** *Escherichia coli* UDP-N-acetylglucosamine 2-epimerase bound to UDP (PDB 1F6D).



**Figure 10.** Global fold and monomeric structure of epimerases belonging to the GT-B structural family. Dimeric (**A**) and monomeric (**B**) structure of *Escherichia coli* UDP-N-acetylglucosamine 2-epimerase bound to UDP (PDB 1F6D). The labeled secondary structure elements ( $\alpha 3$ ,  $\alpha 4$ ,  $\alpha 5$ ) signify the helices involved in dimerization.



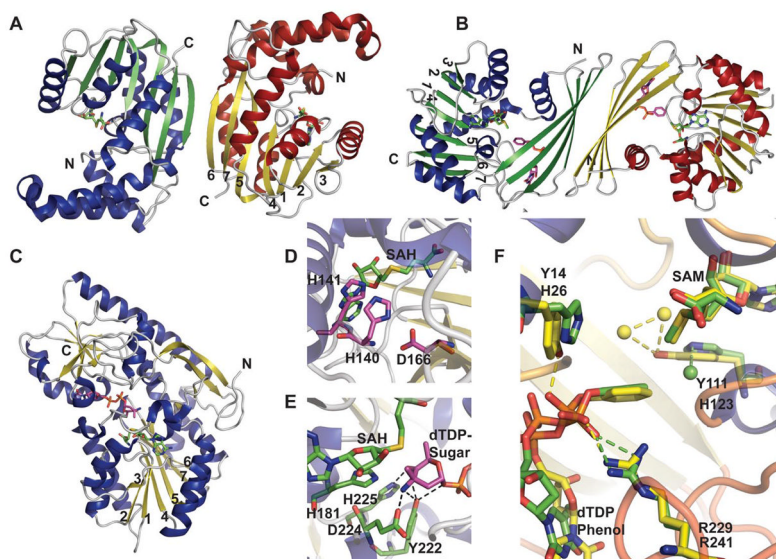
**Figure 11.** Comparison of sugar ketoreductase and sugar 4,6-dehydratase SDR folds. **A.** *Salmonella enterica* RmlD bound to dTDP-Rhamnose and NADPH (PDB 1KC3). **B.** *Salmonella enterica* RmlB bound to dTDP-Glucose and  $NAD^+$  (PDB 1KEU). Orange represents additional RmlB secondary structure ( $\beta 2$ -L2- $\alpha 2$ ), red distinguishes an extra RmlB C-terminus helix and purple signifies secondary structures involved in dimerization **C.** The active site of RmlD.



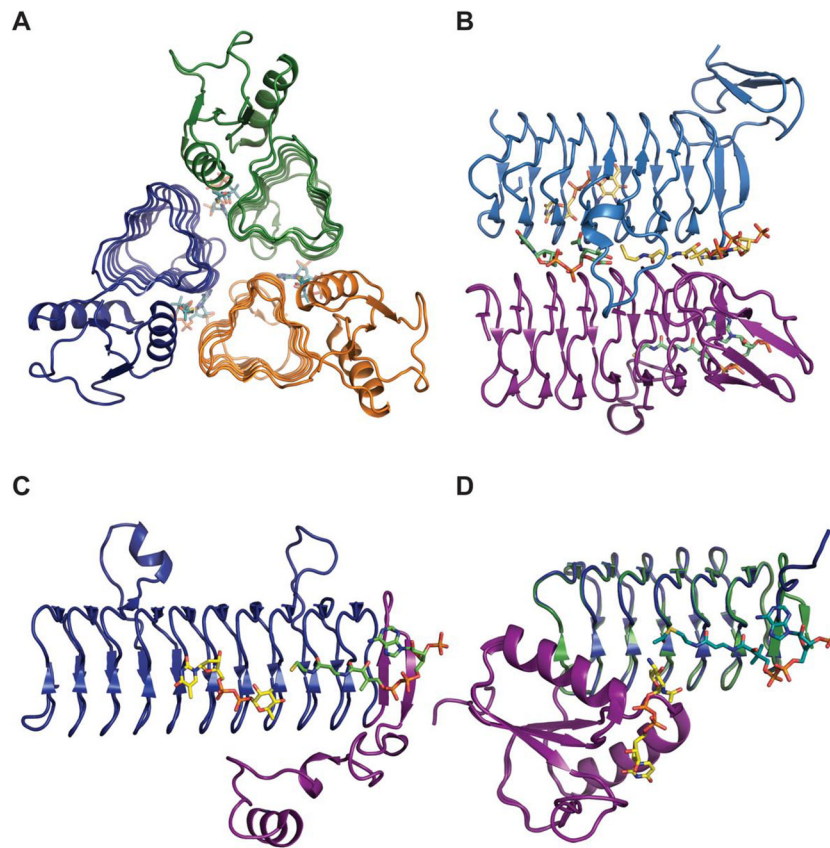
**Figure 12.** Monomeric structure of an epimerase belonging to the GFOR structural family. **A.** Monomeric structure of *Actinomadura kijaniata* KijD10 from (PDB 3RC1) bound to NADP and dTDP-phenol. Helices within the N- and C- terminal domains are represented by distinct colors. **B.** KijD10 active site. Purple signifies residues within the conserved EKP motif, grey spheres highlight water molecules, and the side chain of Asp182 adopts two different configurations.



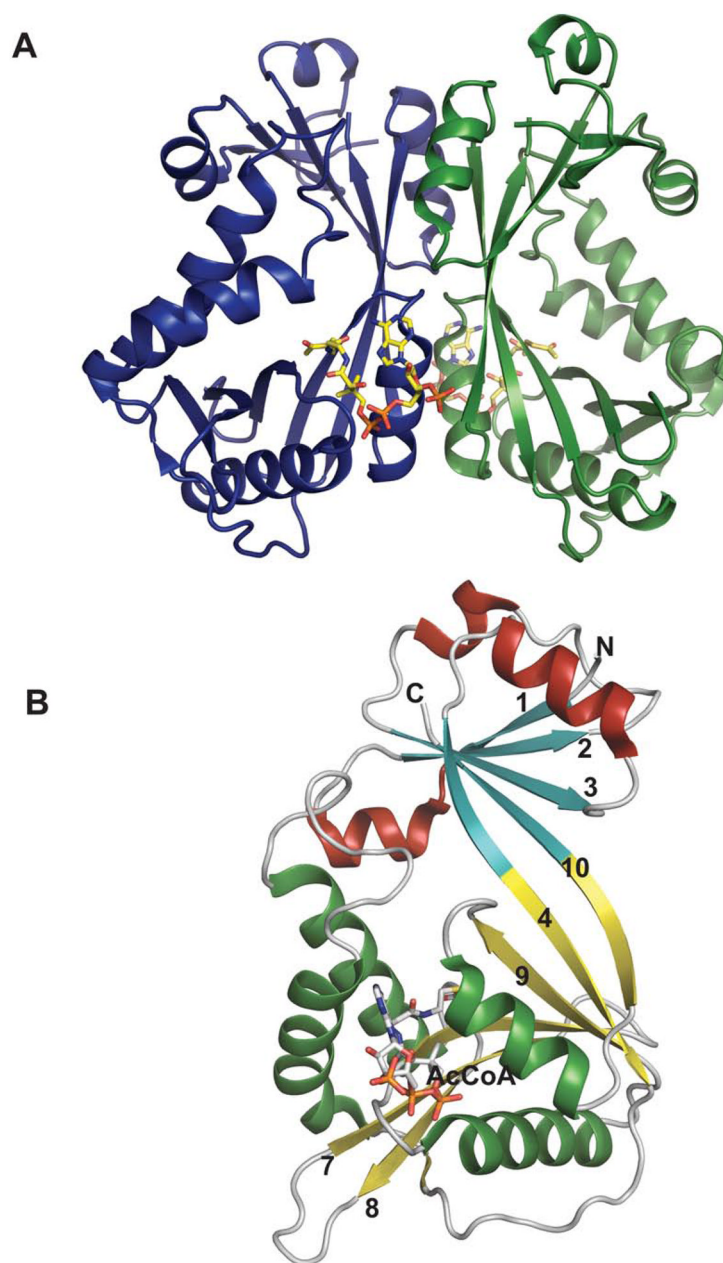




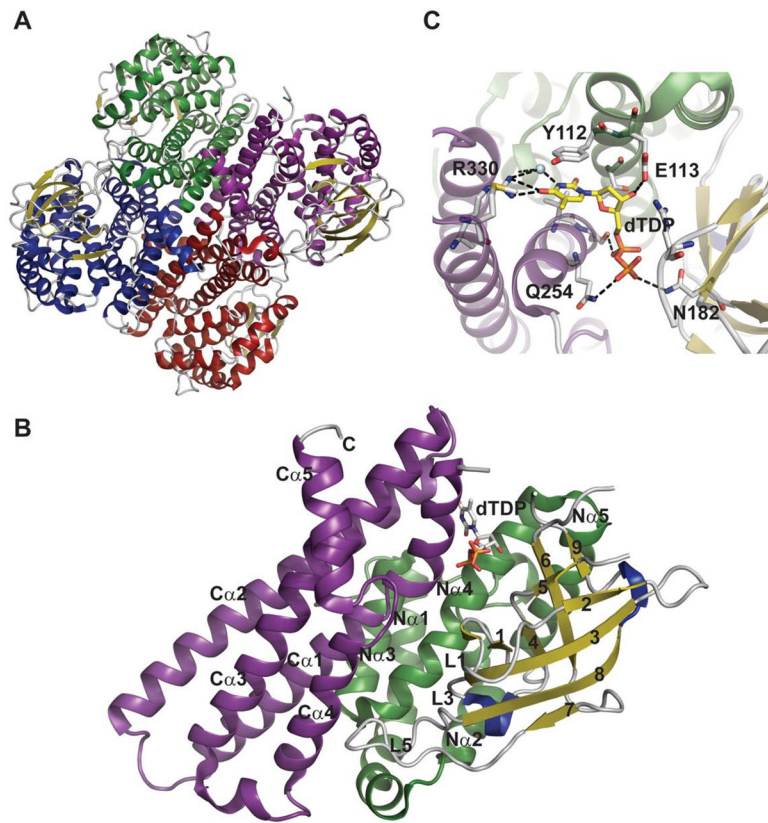
**Figure 14.** Structure and active site architecture of sugar methyltransferases. **A.** Dimer of *Lechevalieria aerocolonigenes* RebM, a rebeccamycin sugar *O*-methyltransferase (PDB 3BUS). **B.** Dimer of *Streptomyces venezuelae* DesVI a *N,N*-dimethyltransferase (PDB 3BXO). **C.** *Micromonospora chalcea* TcaB9, a *C*-methyltransferase (PDB 3NDJ) involved in the biosynthesis of D-tetronitrose. **D.** Active site of RebM. **E.** Active site of TcaB9 bound to dTDP-3-amino-2,3,6-trideoxy-4-keto-3-methyl-D-glucose. **F.** Overlay of the active sites of *N*-methyltransferases TylM1 (green, PDB 3PFG) and DesVI (yellow). Spheres represent water molecules.



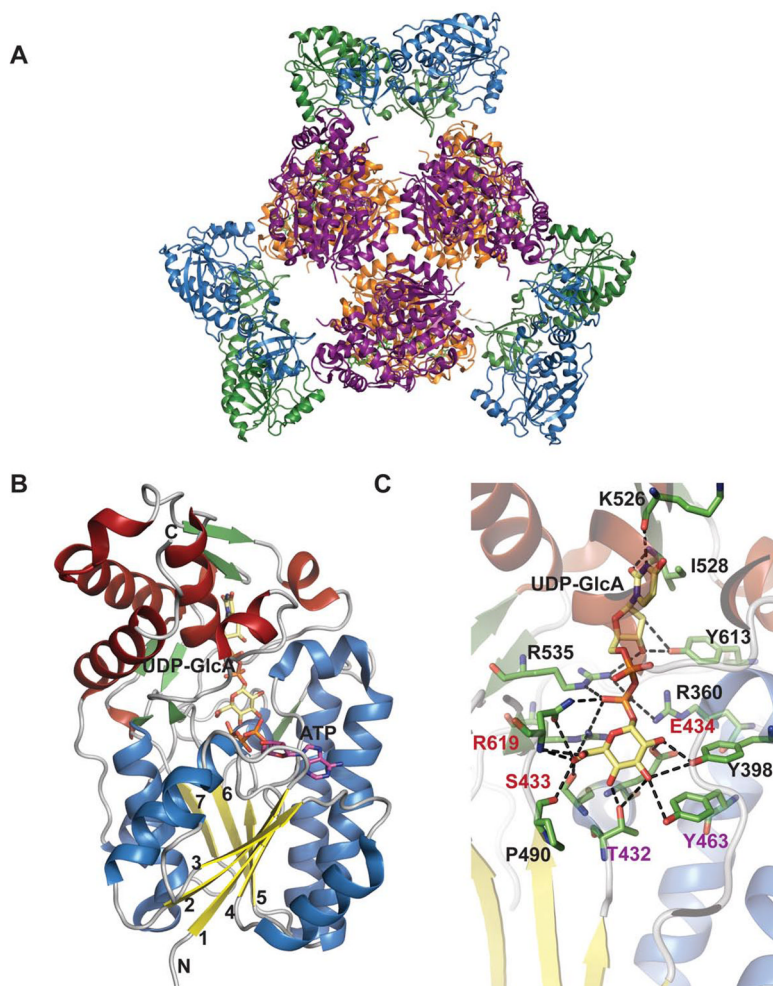
**Figure 15.** Representative global folds and monomeric structures for sugar *N*-acyltransferases belonging to the L $\beta$ H family. **A.** Trimer of *Campylobacter jejuni* PglD (PDB 3BSW). **B.** Dimer interface of *Bordetella petrii* WlbB with subunit A and B substrates colored yellow and green, respectively (PDB 3MQH). **C.** Monomer of *Thermoanaerobacterium thermosaccharolyticum* QdtC bound to CoA and dTDP-3,6-dideoxy-3-amino galactose (PDB 3FSC). **D.** Overlay of monomer of PglD ligand-bound complexes - UDP-2-acetamido-4-amino-6-deoxy-glucose-bound (PDB 3BSS) and AcCoA-bound (PDB 3BSY). In panels **C** and **D**, the N-terminal domains are colored purple.



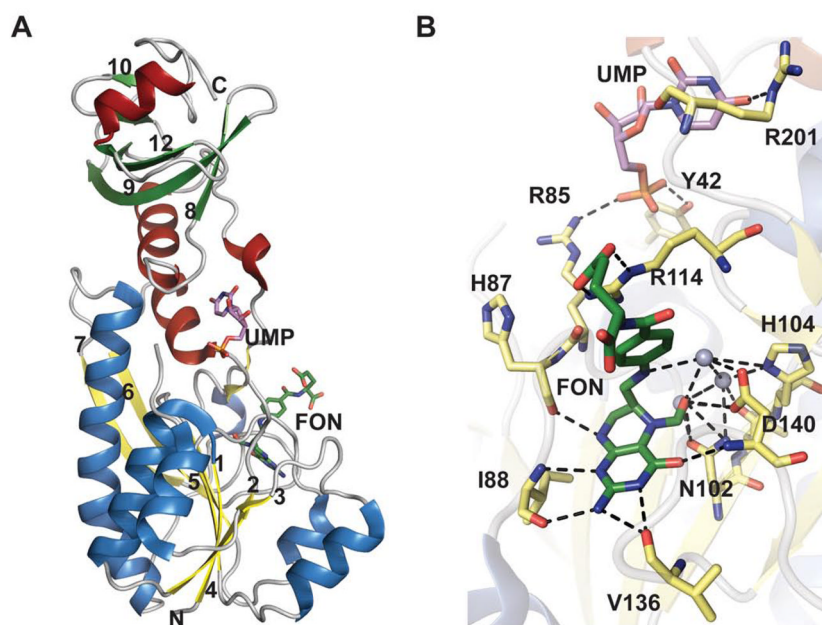
**Figure 16.** Representative global fold and monomeric structure for a sugar *N*-acyltransferase belonging to the GNAT family. Dimeric (A) and monomeric (B) structure of *Escherichia coli* WecD from bound to acetyl-CoA (PDB 2FT0).



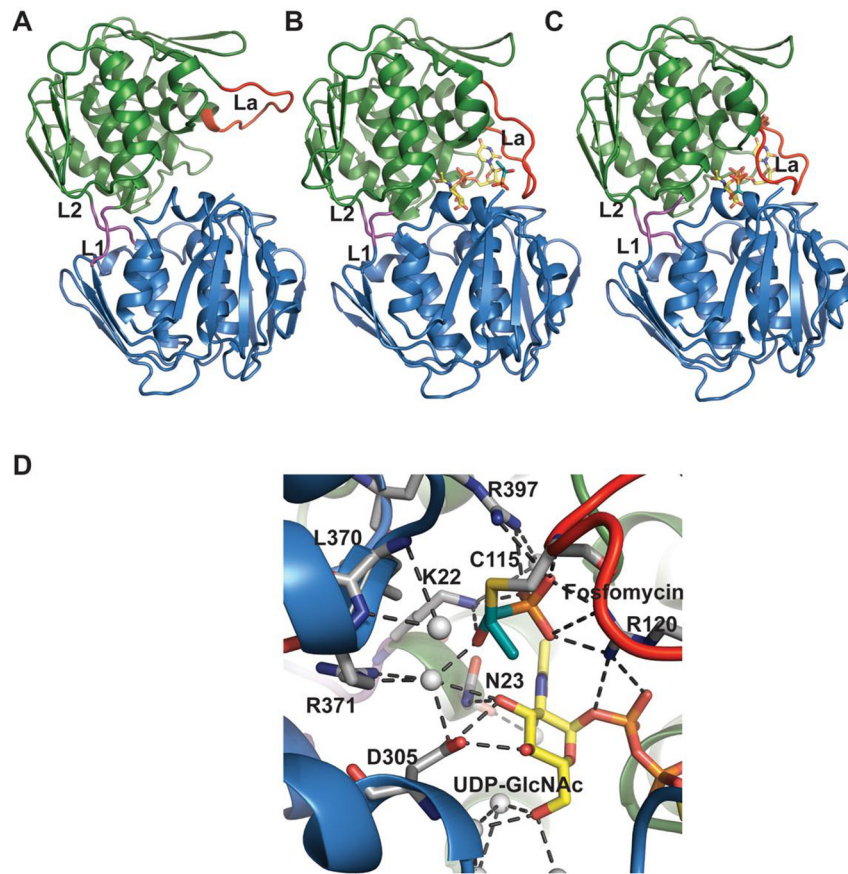
**Figure 17.** Structures of N-oxidases. **A.** Tetramer of *Micromonospora carbonacea* var. *africana* nitrososynthase EvcC (PDB 3MXL). **B.** dTDP binding site of *Actinomadura kijaniata* nitrososynthase KijD3 (PDB 3M9V). **C.** Monomer of KijD3. Green color distinguishes N-terminal  $\alpha$ -helical domain,  $\beta$ -sheet domain is colored yellow, and purple highlights the C-terminal  $\alpha$ -helical domain which is also the tetramerization interface.



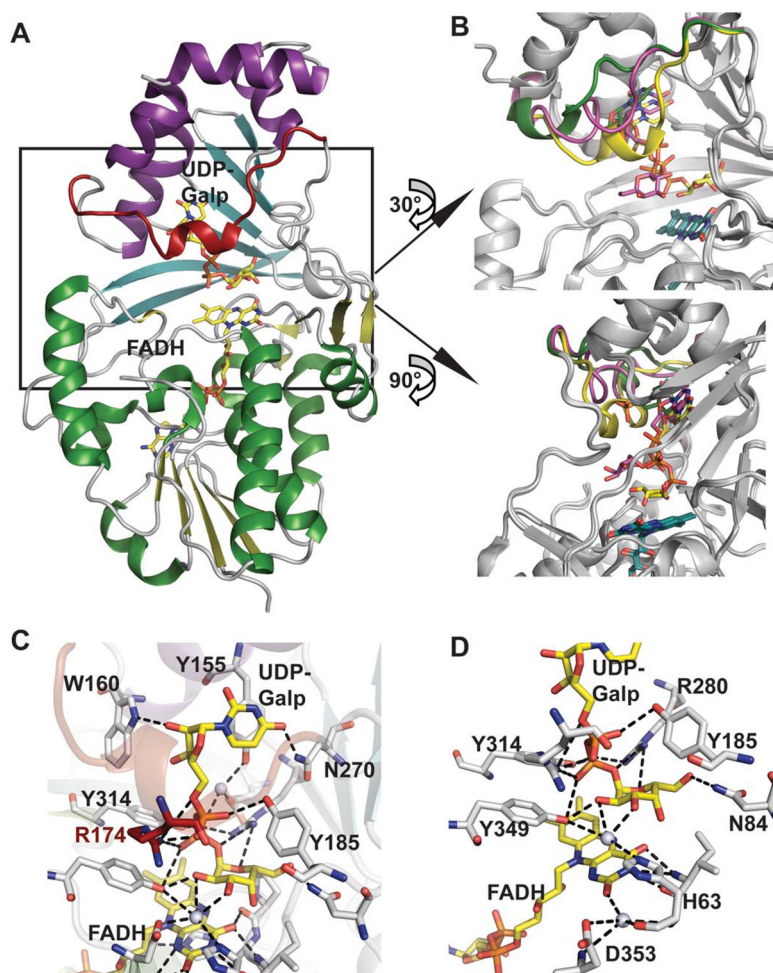
**Figure 18.** Global fold and dehydrogenase domain of *E. coli* ArnA. **A.** Hexameric structure of full-length ArnA bound to ATP and UDP-GlcA (PDB 1Z7E). **B.** Dehydrogenase (DH) domain of ArnA. **C.** Active site of DH domain of ArnA. The residues colored red and purple play a role in catalysis.



**Figure 19.** The formyltransferase domain of *E. coli* ArnA. **A.** ArnA formyltransferase domain bound to UMP and N-5-formyltetrahydrofolate (FON) (PDB 2BLN). Helices in N- and C-terminal domains are colored differently. **B.** Active site of formyltransferase domain bound to UMP and N-5-Formyltetrahydrofolate. Grey spheres are water molecules.



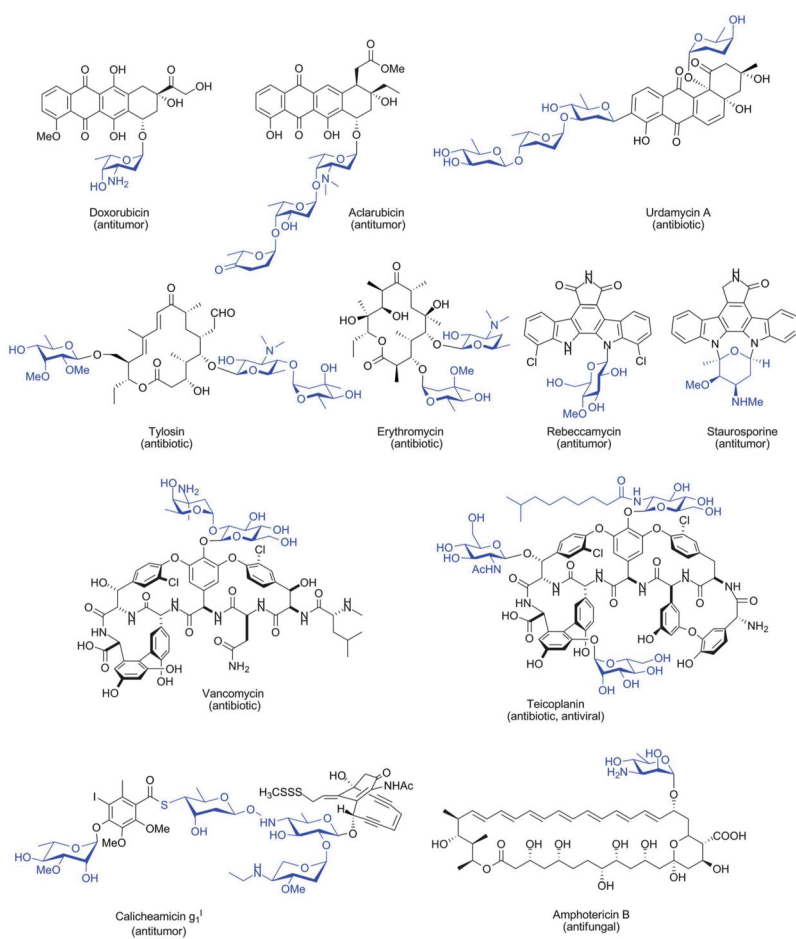
**Figure 20.** Structures of MurA where the active site flexible loop is colored red and domain linkers are colored magenta. **A.** Open, ligand-free form of *E. cloacae* MurA (PDB 1NAW). **B.** Half open conformation of *H. influenzae* MurA complexed with UDP-GlcNAc and fosfomycin (PDB 2RL2). **C.** Closed conformation of *E. coli* MurA complexed with UDP-GlcNAc and fosfomycin (PDB 1UAE). L1 and L2 are the two linker regions, loop La is an active site loop containing residues P111-P121. **D.** Active-site interactions within *E. coli* MurA complexed with UDP-GlcNAc and fosfomycin.



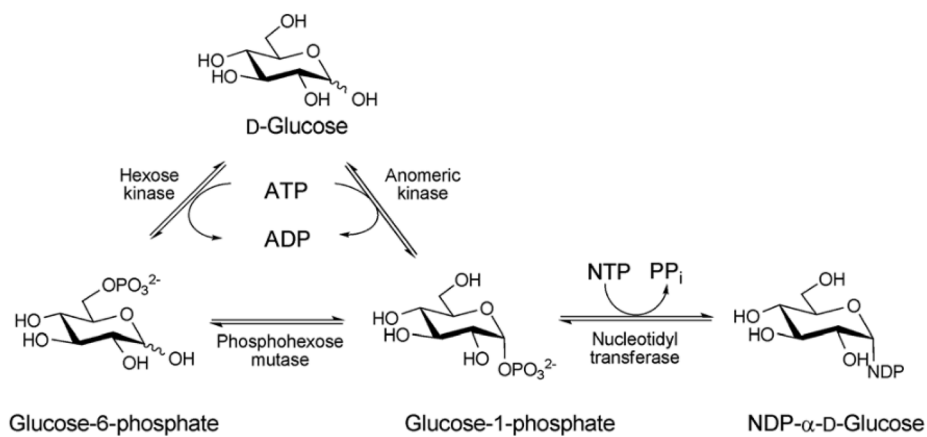
**Figure 21.**

**A.** Structure of UDP-galactopyranose mutase (UGM)-substrate complex from *Klebsiella pneumoniae* (PDB 3INT). Domains 1 (FAD binding domain), 2 (substrate binding domain) and 3 (dimerization domain) are colored green/yellow, cyan and purple, respectively, the mobile loop is colored red, and substrate (UDP-Galp) and cofactor (FAD) are rendered as sticks and colored yellow. **B.** Two different views of the conformation of the mobile loop in the superposed structures of UGMs bound to UDP-Glc (green, 3GF4), UDP-Galp in oxidized state (magenta, PDB 3INR) and UDP-Galp in reduced state (yellow, PDB 3INT). **C.** The substrate binding site highlighting uridine and diphosphate interactions. **D.** The sugar and flavin binding region. Water molecules are illustrated as spheres.

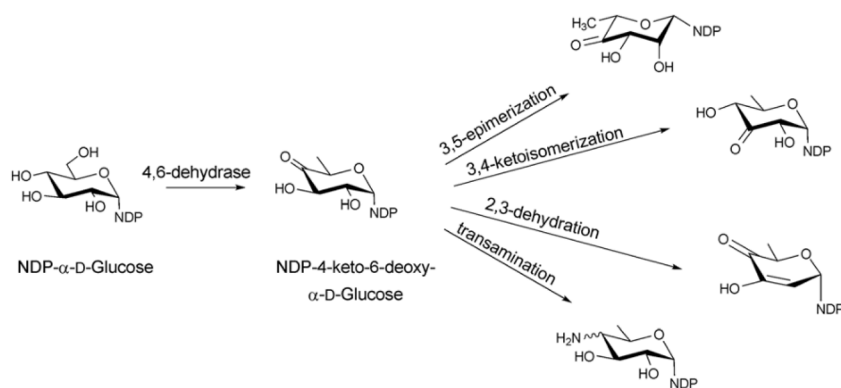




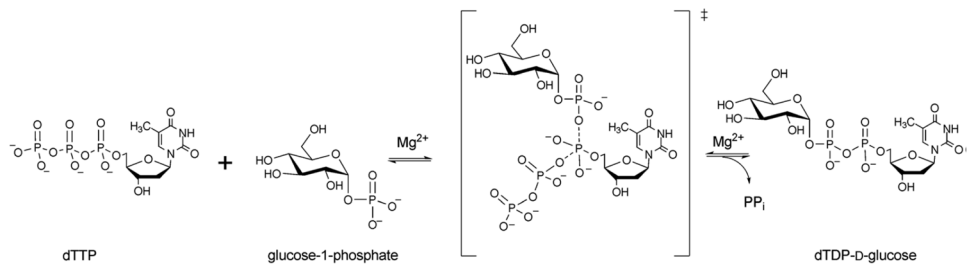
**Scheme 1.**  
Representative glycosylated natural products of microbial origin. Appended sugars are highlighted in blue.

**Scheme 2.**

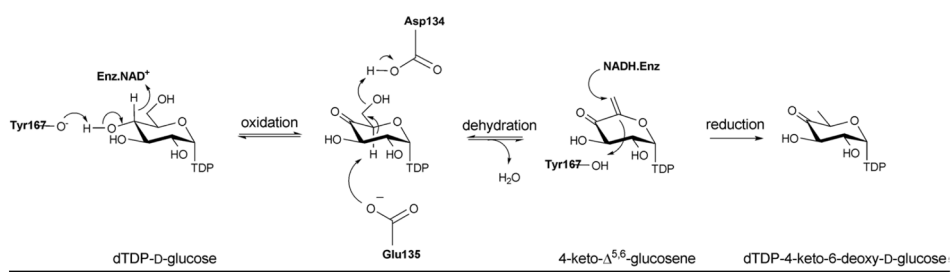
The biosynthesis of NDP-sugars from sugar-1-phosphate. The anomeric kinase reaction presented reflects reactions catalyzed by non-native engineered catalysts.

**Scheme 3.**

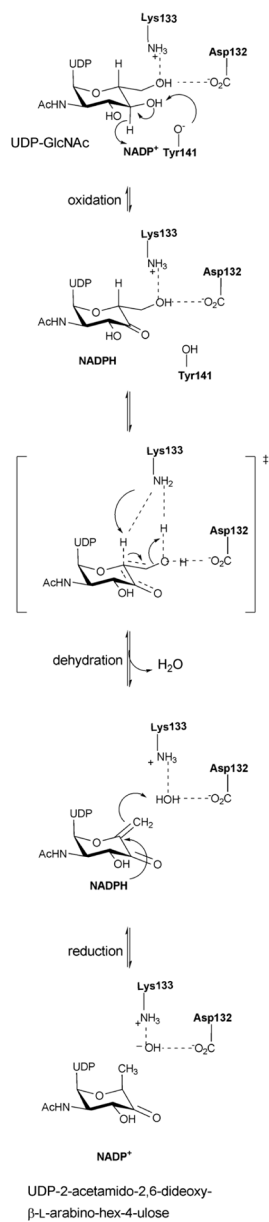
Reactions originating from the common intermediate, NDP-4-keto-6-deoxy- $\alpha$ -D-glucose, which is the product of the 4,6-dehydratase reaction of NDP-D-glucose.

**Scheme 4.**

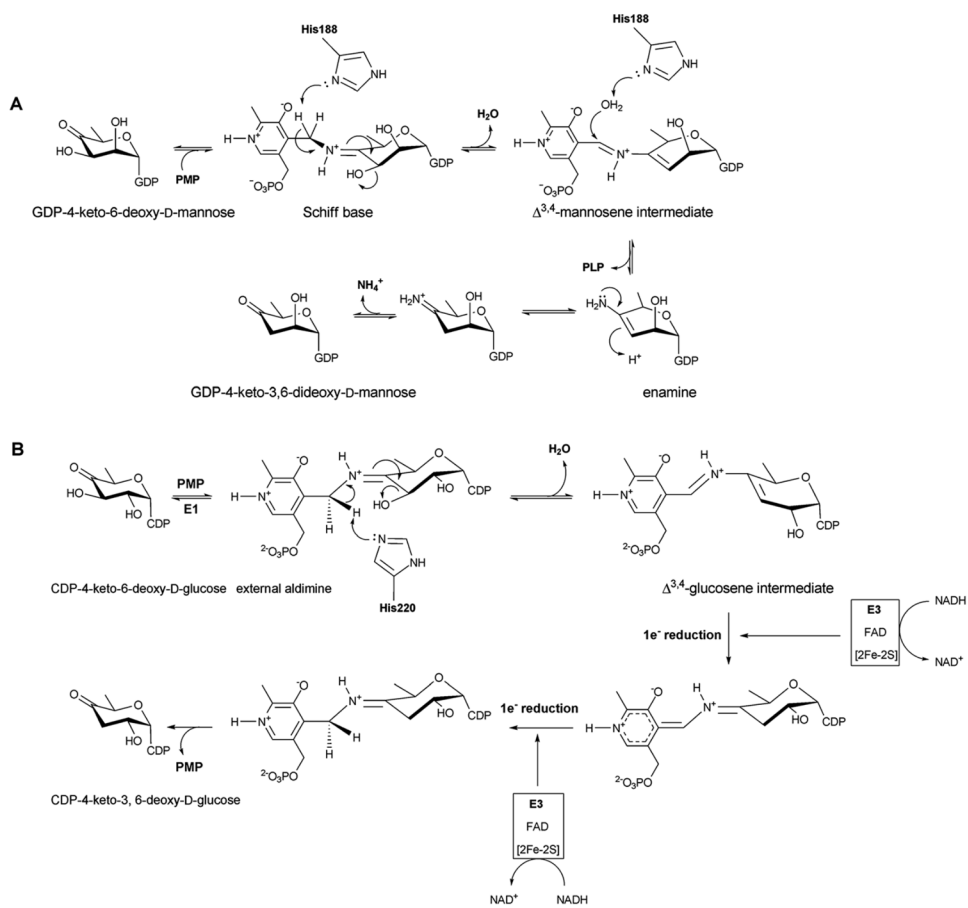
Reaction mechanism of glucose-1-phosphate thymidylytransferase.



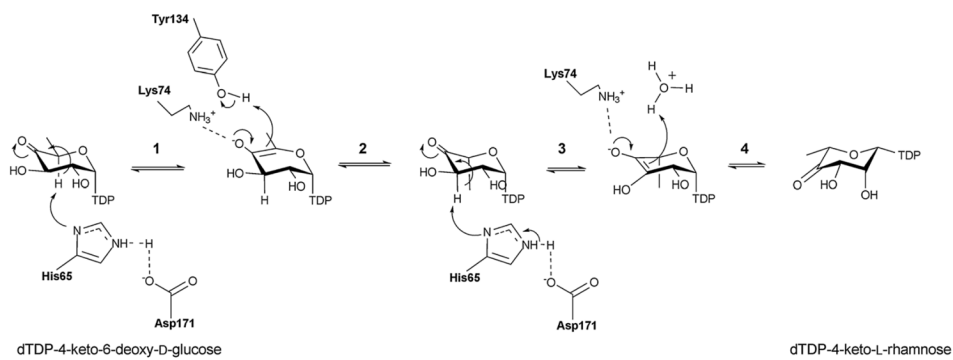
**Scheme 5.**  
Reaction mechanism of dTDP-glucose-4,6-dehydratase (RmlB).



**Scheme 6.**  
Reaction mechanism of an inverting 4,6-dehydratase (FlaA1).

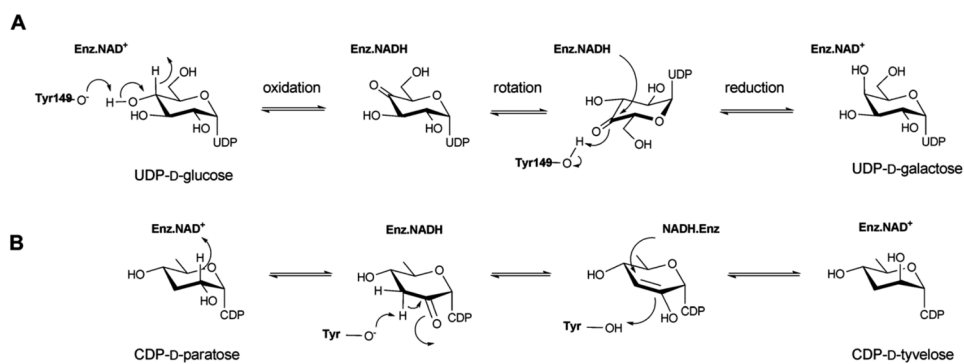
**Scheme 7.**

**A.** Reaction mechanism of GDP-4,6-mannose-dehydratase ColD. **B.** Reaction mechanism of CDP-6-deoxy-L-threo-D-glycero-4-hexulose 3-dehydrase E1.

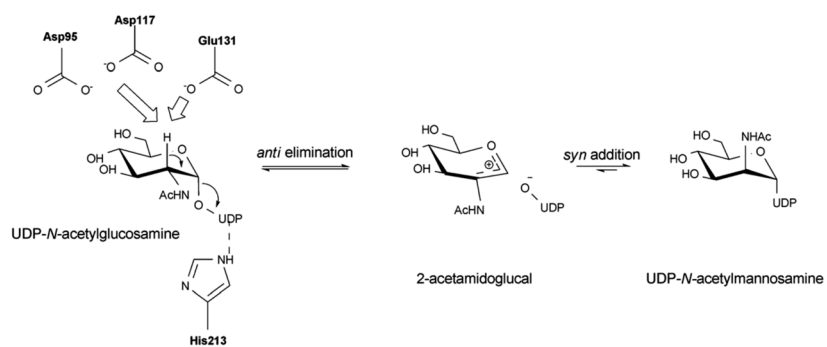
**Scheme 8.**

Reaction mechanism of the 3,5-epimerase RmlC. Step **1**: C5 proton abstraction by the active site His-Asp diad; step **2**: proton addition assisted by active site tyrosine and subsequent inversion; step **3**: C3 proton abstraction by the active site His-Asp diad; step **4**: proton addition and inversion.

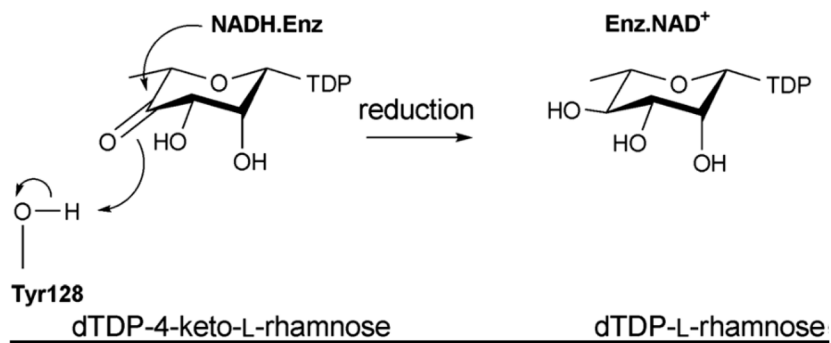


**Scheme 9.**

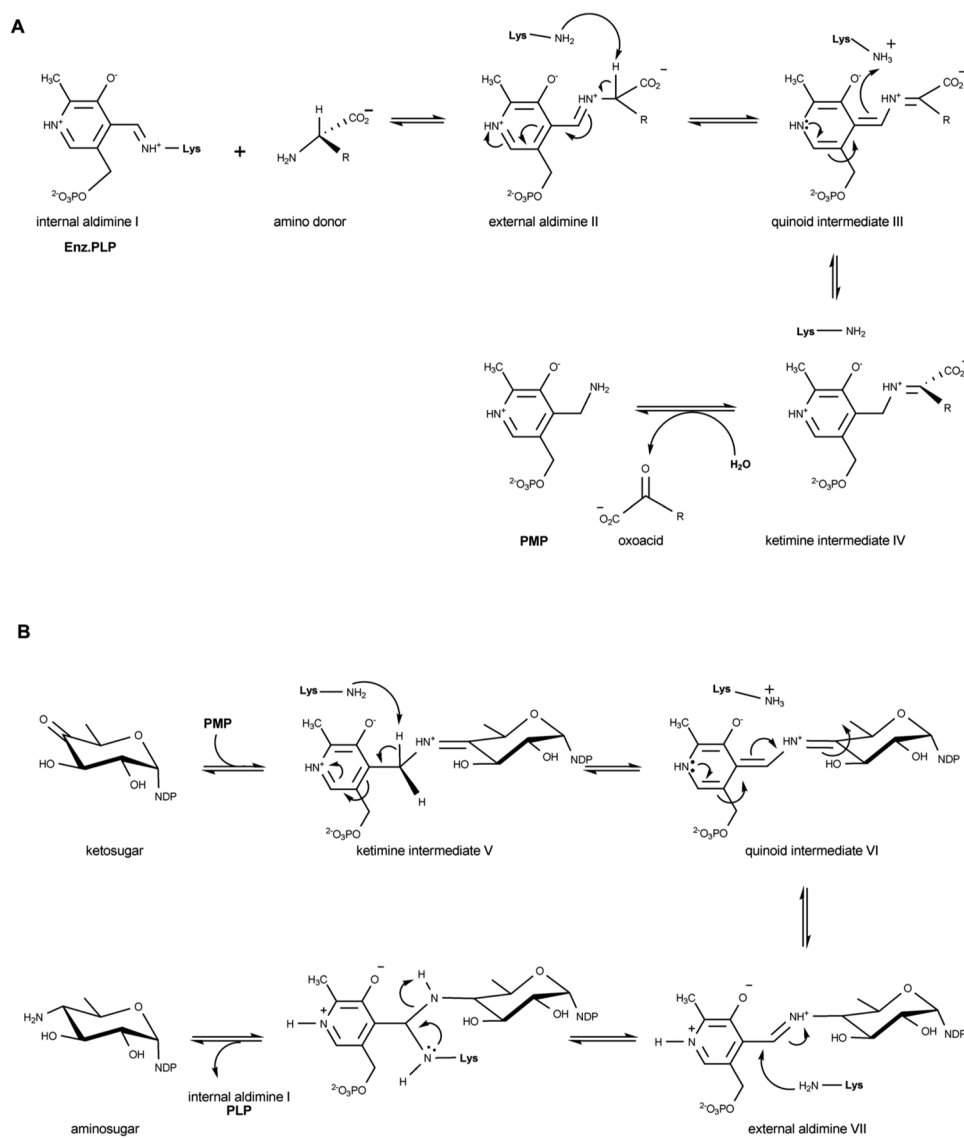
Reaction mechanism employed by UDP-galactose-4-epimerase, which catalyzes the interconversion of UDP-glucose into UDP-galactose (**A**), and CDP-tyvelose-2-epimerase, which interconverts CDP-tyvelose and CDP-paratose (**B**).

**Scheme 10.**

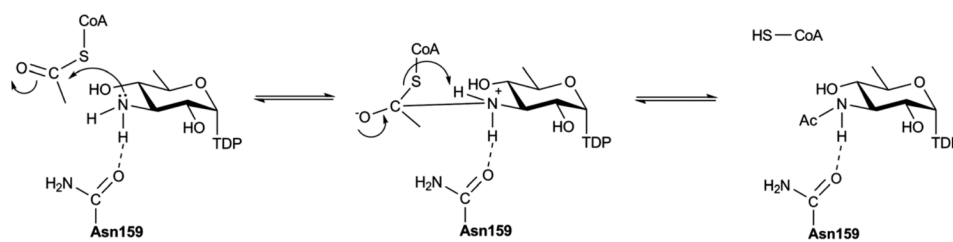
Reaction mechanism of UDP-GlcNAc-2-epimerase, the enzyme that converts UDP-GlcNAc to UDP-ManNAc.



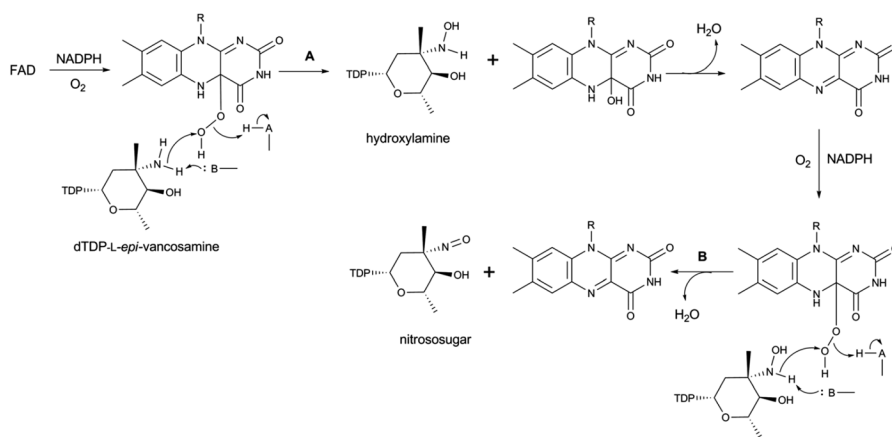
**Scheme 11.**  
The mechanism of C4 reductase RmlD.

**Scheme 12.**

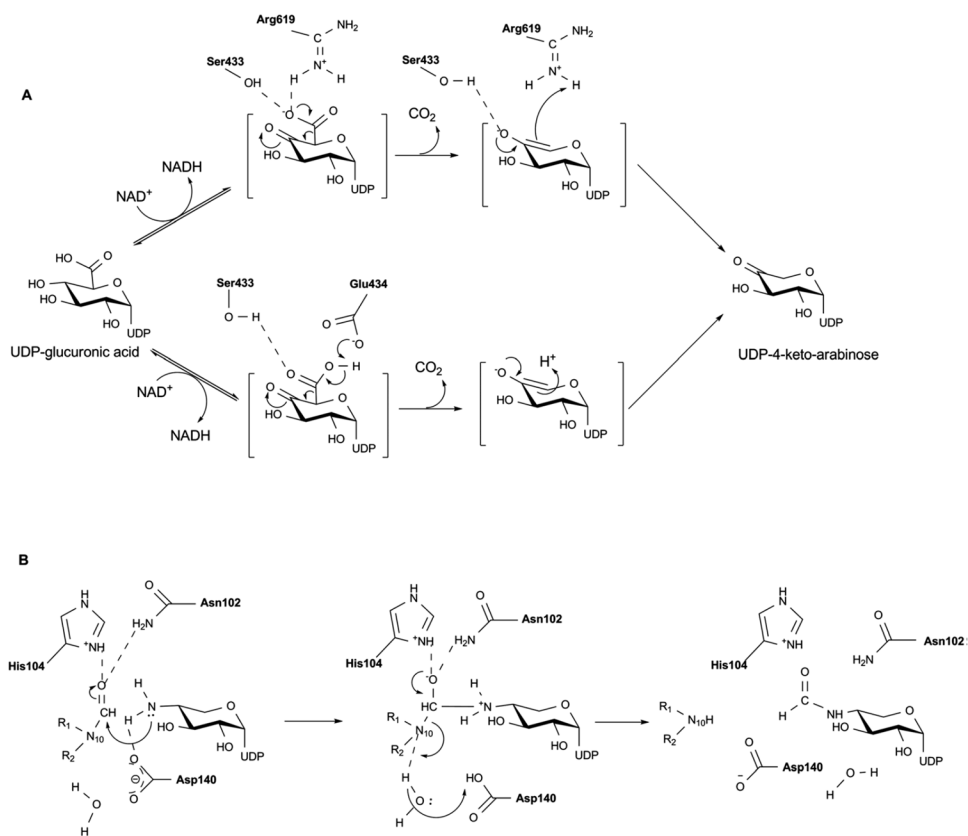
Reaction mechanism employed by aminotransferases. The first half of the reaction enables ‘amine loading’ (namely conversion of PLP to PMP) (A) while the second half of the reaction completes the cycle (B) to ultimately produce an amino sugar and regenerated PLP (bound as internal aldimine).

**Scheme 13.**

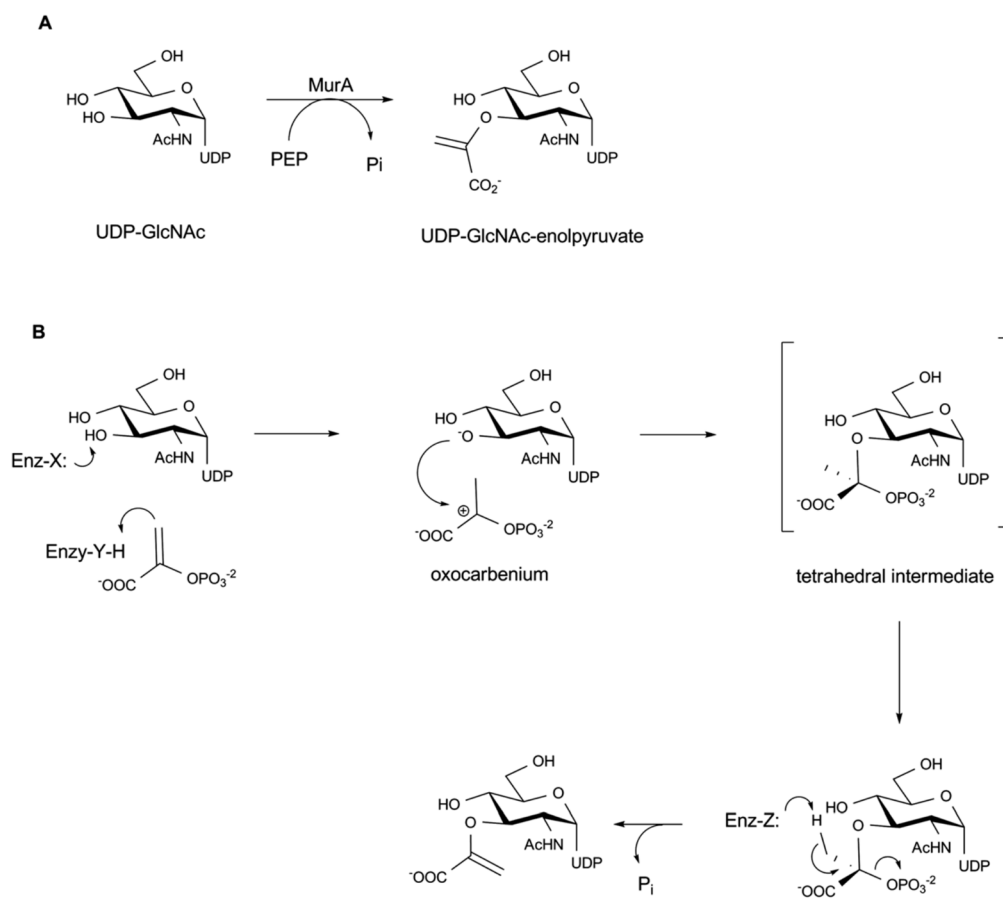
Proposed mechanism for the *N*-acetyltransferase QdtC.

**Scheme 14.**

Proposed mechanism for FMO amine oxidase catalysis which provides for sequential oxidation (hydroxylamino – nitroso – nitro) of a target amino sugar.

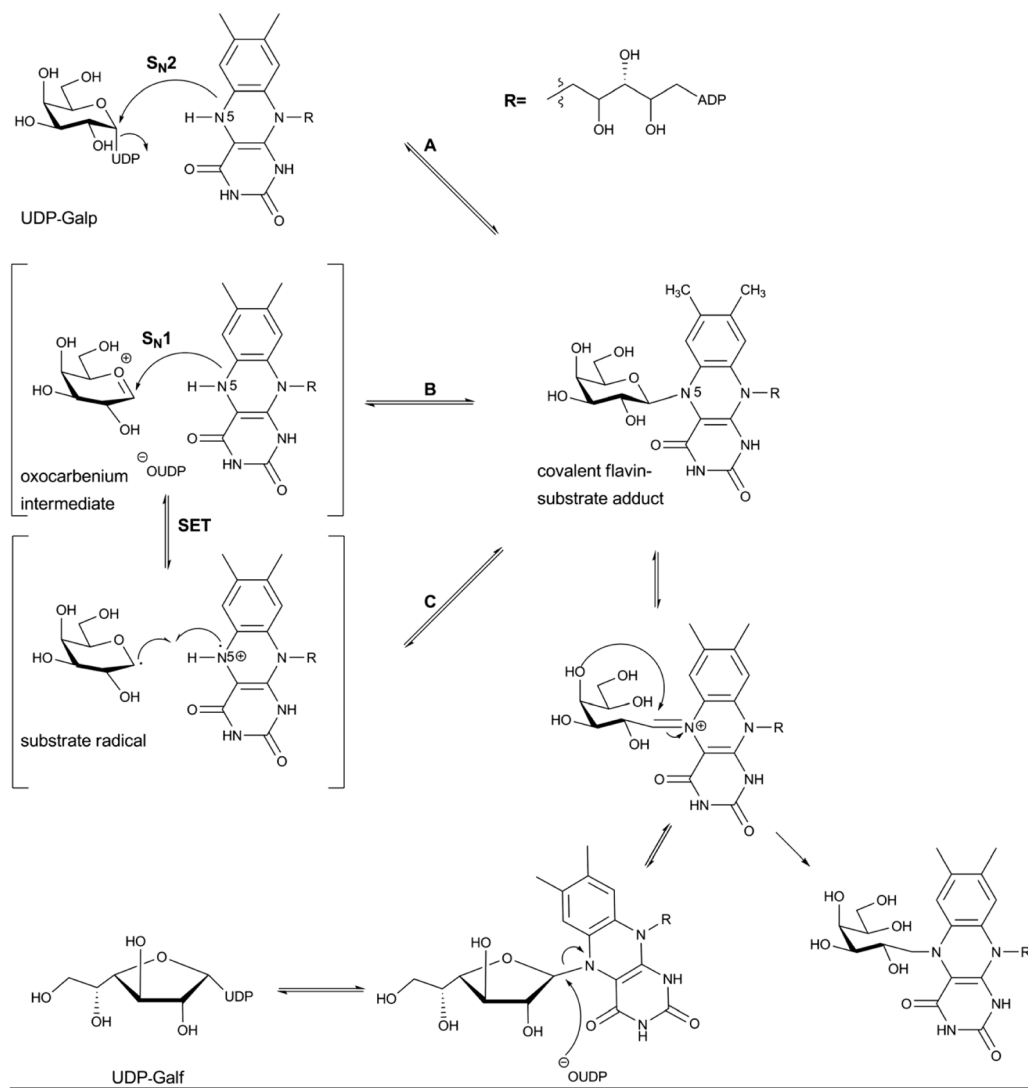


**Scheme 15.**  
Proposed mechanism for the bifunctional activities of ArnA – dehydrogenase (**A**) and formyltransferase (**B**).



**Scheme 16.**  
Reaction (A) and proposed mechanism (B) catalyzed by MurA enzyme.



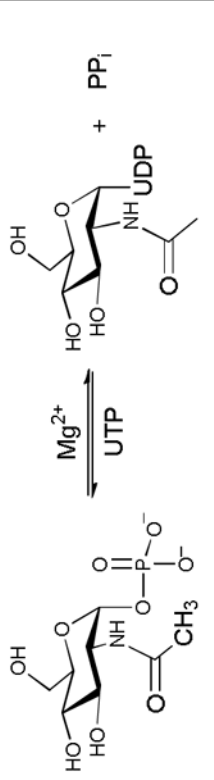
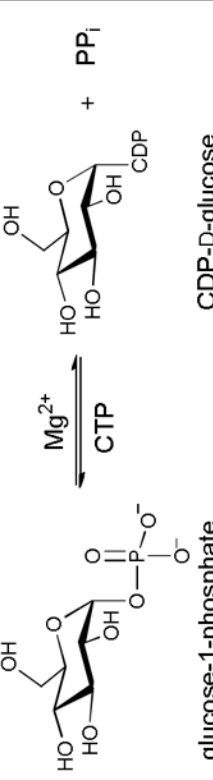
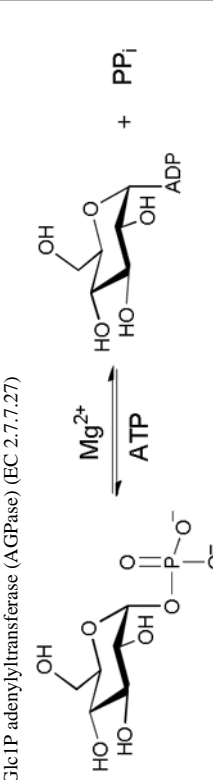



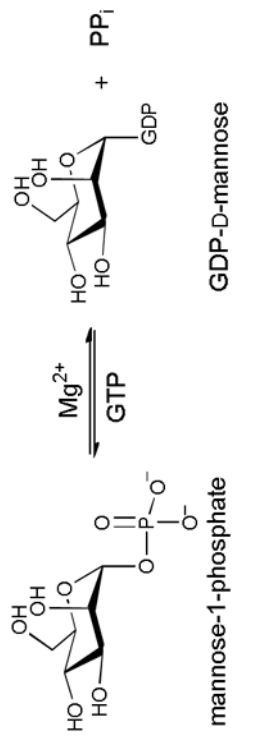
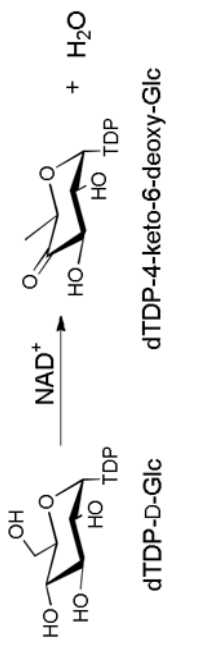
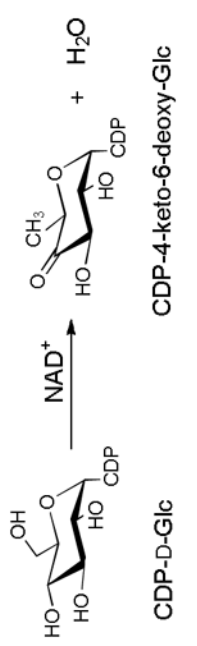
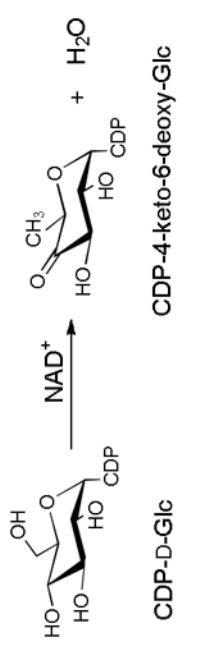
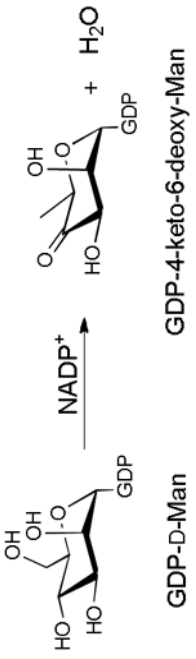
**Scheme 17.**  
Proposed mechanism of UDP-galactopyranose mutase reaction.

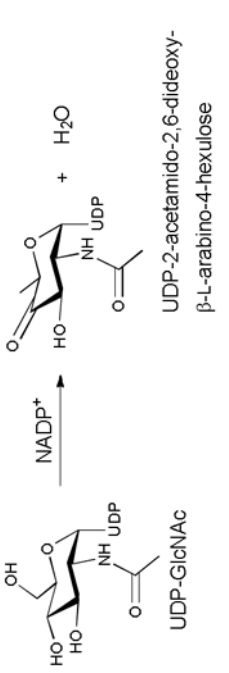
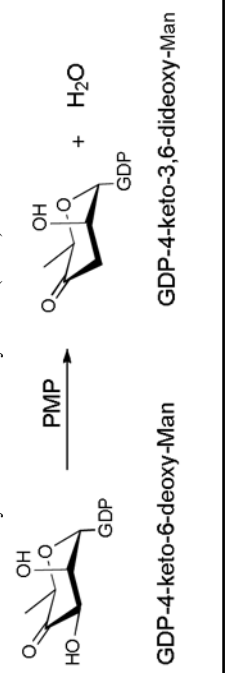
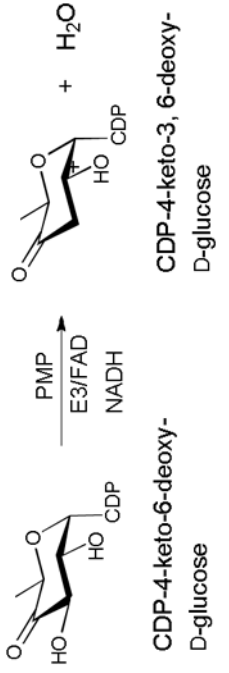
Table 1

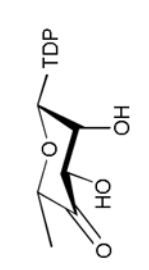
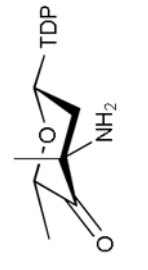
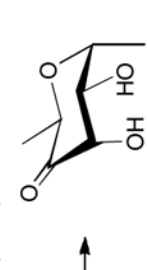
Representative structures, structural classes and enzyme-catalyzed reactions highlighted within this review.

| Organism   | PDB (Variant) | Ligands                         | Ref              | Enzyme & Reaction  |
|--|---------------|---------------------------------|------------------|--|
| <b>Nucleotidyltransferases (EC 2.7.7.-) (Section 3.1 in this review)</b> |               |                                 |                  |  |
| <b>N-ter Rossmann fold and C-ter auxiliary domain</b>                    |               |                                 |                  |  |
| <i>Salmonella enterica</i>   | 1IIN          | UDP-Glc                         | 59, 61, 228, 229 | <p>Glc1P thymidyltransferase (RmlA) (EC 2.7.7.24)</p> <p>glucose-1-phosphate + TTP + Mg<sup>2+</sup> → dTDP-D-glucose + PP<sub>i</sub></p> |
|  | 1IIM          | TTP                             |                  |  |
|  | 1MP3(L89T)    | Apo                             |                  |  |
| <i>Pseudomonas aeruginosa</i>  | 1MP4(W224H)   | Apo                             | 54               |  |
|  | 1MP5(Y177F)   | Apo                             |                  |  |
|  | 3PKP(Q83S)    | dATP                            |                  |  |
|  | 3PKQ(Q83D)    | dGTP                            |                  |  |
|  | 1FZW          | Apo                             |                  |  |
|  | 1FXO          | TMP                             |                  |  |
| <i>Escherichia coli</i>  | 1G0R          | Thymidine/Glc1P                 | 84, 65           |  |
|  | 1G1L          | dTDP-Glc                        |                  |  |
|  | 1G23          | Glc1P                           |                  |  |
|  | 1G2V          | TTP                             |                  |  |
|  | 1G3L          | dTDP-Rha                        |                  |  |
|  | 1MC3          | TTP                             |                  |  |
| <i>Bacillus anthracis</i>  | 1H5T          | dTDP-Glc/dTDP                   |                  |  |
|  | 1H5S          | THM/Glc1P                       |                  |  |
| <i>Helicobacter pylori</i>   | 3HL3          | Sucrose                         | 69               |  |
|  | 3JUJ          | Apo                             |                  |  |
|  | 3IUK          | UDP-Glc                         |                  |  |
| <i>Escherichia coli</i>  | 2E3D          | Apo                             | 66, 67           |  |
|  | 2PA4          | UDP-Glc                         |                  |  |
| <i>Corynebacterium glutamicum</i>  | 2UX8          | Glc1P                           | 230              | <p>Glc1P uridylyltransferase (UGPase) (EC 2.7.7.9)</p> <p>glucose-1-phosphate + UTP + Mg<sup>2+</sup> → UDP-D-glucose + PP<sub>i</sub></p> |
|  |               |                                 |                  |  |
| <i>Streptococcus pneumoniae</i>  | 1HM8          | AcCoA                           | 231              |  |
|  | 1HM0          | Apo                             |                  |  |
|  | 1HM9          | UDP-GlcNAc/AcCoA                |                  |  |
| <i>Escherichia coli</i>  | 1FXJ          | Apo                             | 71, 80           |  |
|  | 1FWY          | UDP-GlcNAc                      |                  |  |
|  | 1HV9          | UDP-GlcNAc/CoA                  |                  |  |
|  | 2015          | UDP-GlcNAc/AcCoA                |                  |  |
|  | 2016          | UDP-GlcNAc/CoA/GlcN1P           |                  |  |
|  | 2017          | UDP-GlcNAc/Desulfo-CoA/GlcNAc1P |                  |  |

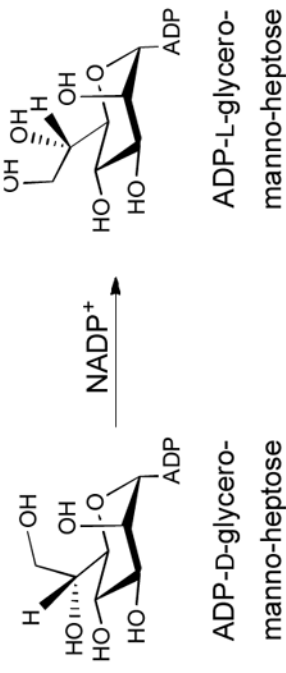
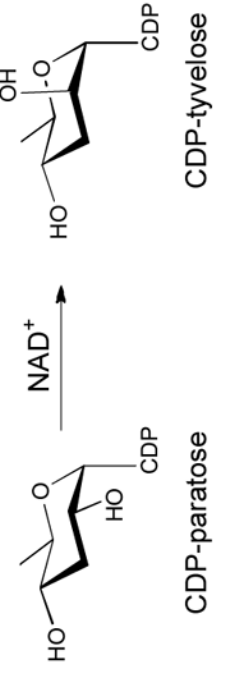
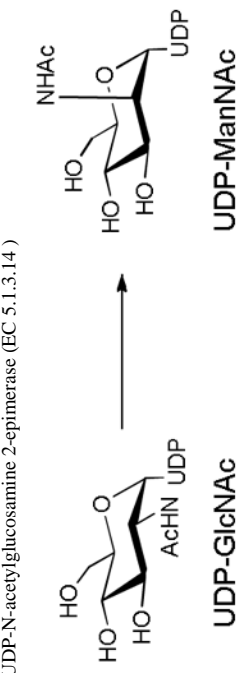
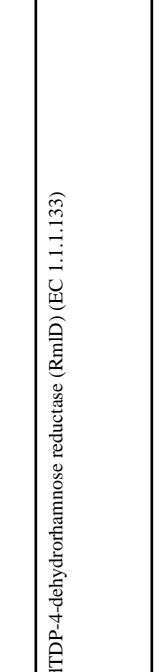
| Organism   | PDB (Variant)                | Ligands                         | Ref    | Enzyme & Reaction   |
|--|------------------------------|---------------------------------|--------|---|
| <i>Mycobacterium tuberculosis</i>                        | 3D8V<br>3D98<br>3DK5<br>3DJ4 | UDP-GlcNAc<br>Apo<br>UDP-GlcNAc | 73, 74 |  <p><b>N-acetylglucosamine-1-phosphate (GlcNAc)</b></p> <p><b>UDP-GlcNAc</b></p> <p><b>PP<sub>i</sub></b></p> <p>Glc1P cytidylyltransferase (CGPase) (EC 2.7.7.33)</p> |
| <i>Yersinia pestis Pestoides F</i>                       | 3FWW                         | Apo                             |        |  <p><b>glucose-1-phosphate</b></p> <p><b>CDP-D-glucose</b></p> <p><b>PP<sub>i</sub></b></p> <p>Glc1P cytidylyltransferase (CGPase) (EC 2.7.7.33)</p>                   |
| <i>Salmonella enterica subsp. enterica serovar Typhi</i> | 1WVC<br>1TZF                 | CTP<br>CDP-Glc                  | 76, 77 |  <p><b>glucose-1-phosphate</b></p> <p><b>ADP-D-glucose</b></p> <p><b>PP<sub>i</sub></b></p> <p>Glc1P adenylyltransferase (AGPase) (EC 2.7.7.27)</p>                   |
| <i>Agrobacterium tumefaciens</i>                         | 3BRK                         | Apo                             | 79     |  <p><b>Man1P guanylyltransferase (GMPase) (EC 2.7.7.22)</b></p>  |
| <i>Thermotoga maritima</i>                               | 2X5S<br>2X5Z<br>2X60<br>2X65 | Apo<br>GDP-Man<br>GTP<br>Man1P  | 78     |   |

| Organism   | PDB (Variant)                | Ligands  | Ref     | Enzyme & Reaction   |
|--|------------------------------|--|---------|---|
| <b>Dehydratases (EC. 4.2.1.-) Section 4.1 in this review</b>           |                              |  |         |   |
| <b>SDR fold (Section 4.1.1. in the review)</b>                         |                              |  |         |   |
| <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>typhi</i> | IG1A<br>IKEU<br>IKEW         | NAD<br>NAD/dTDP-Glc<br>NAD/dTDP                          | 55, 86  |  <p>mannose-1-phosphate + Mg<sup>2+</sup> + GTP → GDP-D-mannose + PP<sub>i</sub></p> |
| <i>Streptococcus suis</i>  | IKET<br>IKER<br>IKEP<br>IOC2 | NAD/dTDP<br>NAD/dTDP-Glc<br>NAD/dTDP-Xyl<br>NAD/dTDP-Xyl | 55, 232 |  <p>dTDP-D-Glc + NAD<sup>+</sup> → dTDP-4-keto-6-deoxy-Glc + H<sub>2</sub>O</p>      |
| <i>Streptomyces venezuelae</i>   | IR66<br>IR6D                 | NAD/dTDP<br>NAD/dTDP-Glc                                 | 87      |  <p>CDP-D-Glc + NAD<sup>+</sup> → CDP-4-keto-6-deoxy-Glc + H<sub>2</sub>O</p>       |
| <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>typhi</i> | 1WVG                         | CDP-Xyl  | 89      | <p>CDP-glucose-4,6-dehydratases (EC 4.2.1.45)</p>   |
| <i>Yersinia pseudotuberculosis</i>                                     | IRKX                         | NAD  | 88      |  <p>CDP-D-Glc + NAD<sup>+</sup> → CDP-4-keto-6-deoxy-Glc + H<sub>2</sub>O</p>       |
| <i>Pseudomonas aeruginosa</i>  | IRPN                         | GDP/NADPH  | 91      | <p>GDP-mannose-4,6-dehydratases (GMD) (EC 4.2.1.47)</p>   |
| <i>Escherichia coli</i>  | IDB3                         | Apo  | 92      |  <p>GDP-D-Man + NAD<sup>+</sup> → GDP-4-keto-6-deoxy-Man + H<sub>2</sub>O</p>      |
| <i>Aquifex aeolicus</i> VF5  | 1Z95<br>2Z1M                 | GDP/NADPH<br>GDP/NADPH                                   |         |   |

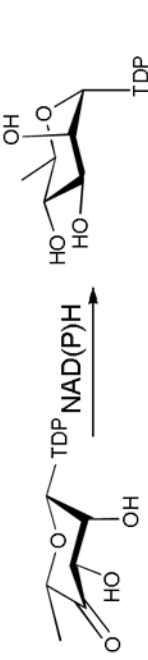
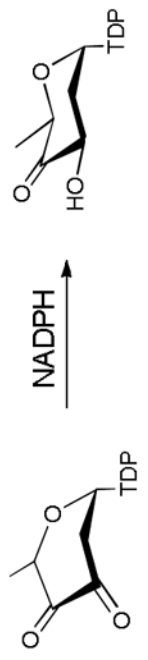
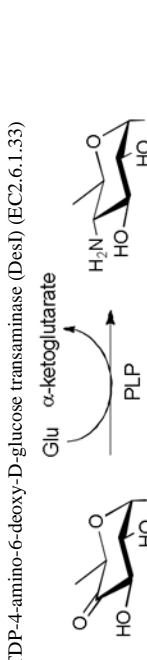
| Organism  | PDB (Variant)  | Ligands   | Ref        | Enzyme & Reaction  |
|---|--|---|------------|--|
| <i>Helicobacter pylori</i>                                  | 2GNA<br>2GN4<br>2GN6<br>2GN8<br>2GN9                           | NADP/UDP-Gal<br>NADPH/UDP-GlcNAc<br>NADP/UDP-GlcNAc<br>NADP/UDP<br>NADP/UDP-Glc                             | 93         |  <p>UDP-GlcNAc-4,6-dehydratases (FlaA1) (Inverting) (EC 4.2.1.115)</p> <p>UDP-2-acetamido-2,6-dideoxy-<math>\beta</math>-L-arabino-4-hexulose</p> |
| <b>AAT fold (Section 4.1.2. in the review)</b>              |  |   |            |  |
| <i>Escherichia coli</i>                                     | 2GMU<br>2GMS<br>2ROT(H188K)<br>3B8X(H188N)<br>3GR9(H188KS187N) | PLP-glutamate ketimine<br>Hydrated PLP<br>PLP-glutamate diamine<br>GDP-Perosamine<br>PLP-glutamate ketimine | 52, 58, 94 |  <p>GDP-4-keto-6-deoxy-D-mannose 3-dehydratase (CoID)</p> <p>GDP-4-keto-6-deoxy-Man      GDP-4-keto-3,6-dideoxy-Man</p>                           |
| <i>Yersinia pseudotuberculosis</i>                          | 3BCX<br>3BB8 (H220K)   | Apo<br>PLP  | 53         |  <p>CDP-6-deoxy-L-threo-D-glycero-4-hexulose 3-dehydratase (E1)</p> <p>CDP-4-keto-6-deoxy-D-glucose      CDP-4-keto-3,6-dideoxy-D-glucose</p>    |
| <b>Epimerases (EC 5.1.3.-) (Section 4.2 in this review)</b> |  |   |            |  |
| <b>Cupin superfamily (Section 4.2.1. in the review)</b>     |  |   |            |  |
| <i>Salmonella enterica subsp. enterica serovar Typhi</i>    | IDZR<br>IDZT   | Apo<br>dTDP-phenol  | 110        | dTDP-4-dehydrothamnose 3,5-epimerase (RmlC) (EC 5.1.3.13)  |
| <i>Streptococcus suis</i>                                   | INXM<br>INYW<br>INZC<br>IIXL                                   | Apo<br>dTDP-Glc<br>dTDP-Xyl<br>dTDP-Rha   | 108, 109   |  |

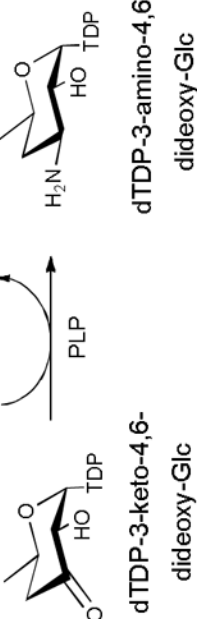
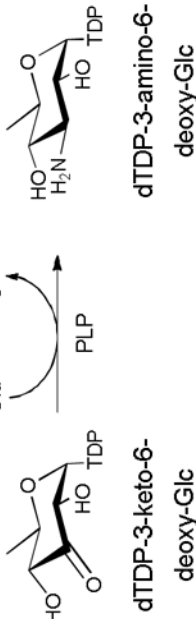
| Organism                            | PDB (Variant)                     | Ligands  | Ref      | Enzyme & Reaction   |
|-------------------------------------|-----------------------------------|--|----------|---|
| <i>Mycobacterium tuberculosis</i>   | 2IXC<br>1UPI<br>1PM7              | dTDP-Rha<br>Apo<br>Apo                         | 111, 112 |  <p>dTDP-4-keto-6-deoxy-D-glucose<br/>D-glucose</p> <p>dTDP-4-keto-L-rhamnose<br/>L-rhamnose</p>             |
| <i>Pseudomonas aeruginosa</i>       | 2IXH<br>2IXI<br>2IXJ<br>2IXK      | dTDP-Rha<br>dTDP-Xyl<br>Apo<br>dTDP-4-Keto Rha | 112      |   |
| <i>Bacillus anthracis str. Ames</i> | 3RYK                              | dTDP   |          |   |
| <i>Amycolatopsis orientalis</i>     | 1WA4(M131E/L135A)<br>1OFN<br>1O16 | Apo<br>Apo<br>TMP                              | 113, 114 |  <p>dTDP-4-Keto-6-deoxy-D-glucose-5-epimerase (EvaD)</p> <p>dTDP-3-amino-3-methyl-4-keto-6-deoxy-D-Glc</p>   |
| <i>Streptomyces caeruleus</i>       | 2C0Z                              | Apo  | 115      |  <p>dTDP-6-deoxy-D-xylo-4-hexulose 3-epimerase (NovW)</p> <p>dTDP-4-keto-6-deoxy-D-glucose<br/>D-glucose</p> |



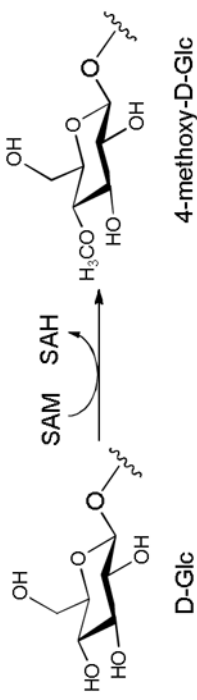
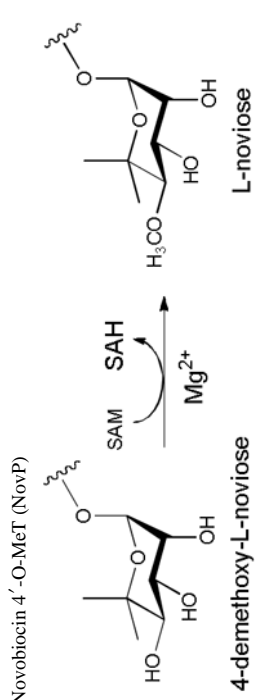
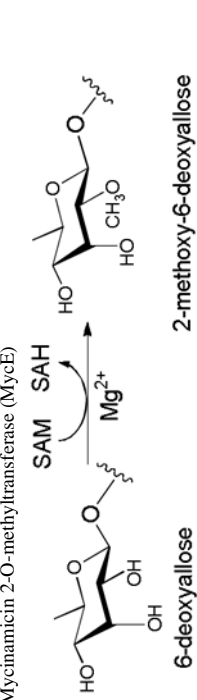
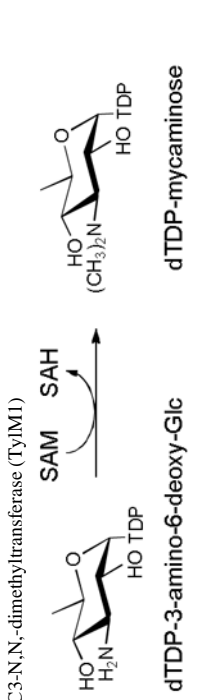
| Organism   | PDB (Variant)                        | Ligands   | Ref      | Enzyme & Reaction   |
|--|--------------------------------------|---|----------|---|
| <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhi</i> | IORR                                 | NAD/CDP   | 129      |  <p>ADP-D-glycero-manno-heptose + NADP<sup>+</sup> → ADP-L-glycero-manno-heptose + NADP<sup>+</sup></p> <p>CDP-paratose 2-epimerase (EC 5.1.3.10)</p>  <p>CDP-paratose + NAD<sup>+</sup> → CDP-tyvelose + NAD<sup>+</sup></p> |
| <b>GT-B fold (Section 4.2.3. in the review)</b>                        |                                      |   |          |   |
| <i>Escherichia coli</i>  | IF6D                                 | UDP   | 131, 234 |  <p>UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14)</p>  <p>UDP-GlcNAc + NAD<sup>+</sup> → UDP-ManNAc + NAD<sup>+</sup></p>  |
| <i>Bacillus anthracis</i>  | 3BEO                                 | UDP/UDP-GlcNAc                                  | 132      |   |
| <i>Vibrio cholerae</i>   | 3DZC                                 | Apo   |          |   |
| <i>Thermus Thermophilus</i> HB8  | 1V4V                                 | Apo   |          |   |
| <b>Ketoreductases (EC 1.1.1.-) (Section 4.3 in this review)</b>        |                                      |   |          |   |
| <b>SDR fold (Section 4.3.1. in the review)</b>                         |                                      |   |          |   |
| <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhi</i> | IKBZ<br>IKC0<br>IKC1<br>IKC3<br>IN2S | Apo<br>NADH<br>NADPH<br>NADPH/dTDP-L-Rha<br>NAD | 135      | dTDP-4-dehydrohamnose reductase (RmID) (EC 1.1.1.133)   |

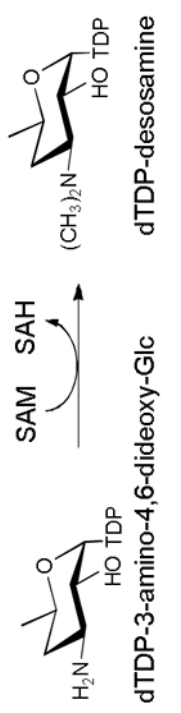
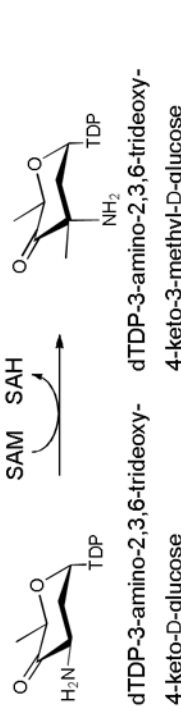
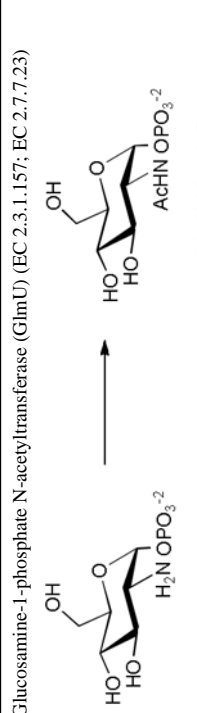
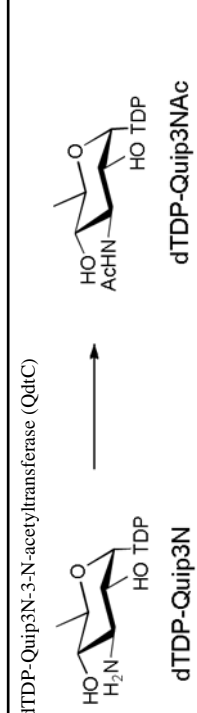
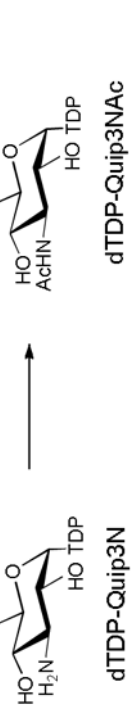


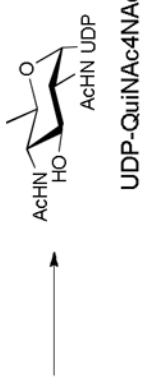
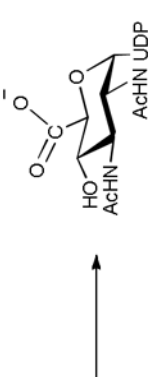
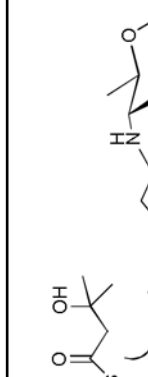
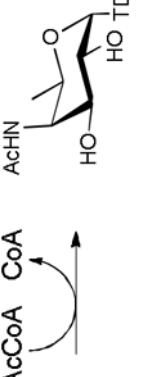
| Organism   | PDB (Variant)   | Ligands   | Ref | Enzyme & Reaction   |
|--|---|---|-----|---|
| <i>Bacillus anthracis</i>  | 3SC6  | NADP  |     |  <p>dTDP-4-keto-6-deoxy-Rha <math>\xrightarrow{\text{NAD(P)H}}</math> dTDP-6-deoxy-Man</p>   |
| <b>Glucose-Fructose OxidoReductase (GFOR) superfamily (Section 4.3.2. in the review)</b> |   |   |     |   |
| <i>Actinomadura kijaniata</i>  | 3RBV<br>3RC1<br>3RC2<br>3RC7(Y186F)<br>3RC9(K102A)<br>3RCB(K102E) | NADP<br>NADP/dTDP-benzene<br>NADP/dTDP-benzene<br>NADP/dTDP-benzene<br>NADP/dTDP-benzene<br>NADP/dTDP-benzene | 136 |  <p>dTDP-3,4-diketo-2,6-dideoxy-Glc <math>\xrightarrow{\text{NADPH}}</math> dTDP-4-keto-2,6-dideoxy-Glc</p>  |
| <b>Sugar aminotransferases (SAT) (EC 2.6.1.1.-) (Section 4.4 in this review)</b>         |   |   |     |   |
| <b>AAT fold (Section 4.4. in the review)</b>   |   |   |     |   |
| <i>Streptomyces venezuelae</i>   | 2FO3  | dTDP-4-amino-4,6-dideoxyglucose   | 143 |  <p>dTDP-4-amino-6-deoxy-D-glucose transaminase (DesI) (EC2.6.1.33)</p> <p>Glu <math>\alpha</math>-ketoglutarate <math>\xrightarrow{\text{PLP}}</math> dTDP-4-amino-6-deoxy-Glc</p> |
| <i>Streptomyces venezuelae</i>   | 2OGA<br>2OGE  | Ketamine with Glu<br>internal aldimine  | 149 | dTDP-3-amino-3,6-dideoxy- $\alpha$ -D-glucose transaminase (DesV) (EC 2.6.1.89)   |

| Organism   | PDB (Variant)               | Ligands                                    | Ref      | Enzyme & Reaction  |
|--|-----------------------------|--|----------|--|
| <i>Thermoanaerobacterium thermosaccharolyticum</i> | 3FRK                        | PLP;dTDP-3-aminoquinovose aldimine         | 148      |  <p>dTDP-3-keto-4,6-dideoxy-Glc</p> <p>dTDP-3-amino-4,6-dideoxy-Glc</p> |
| <i>Caulobacter crescentus</i><br>CB15              | 3BN1<br>3DR7<br>3DR4(K186A) | internal aldimine<br>GDP-3-deoxyperosamine | 142, 150 |  <p>GDP-4-keto-6-deoxy-Man</p> <p>GDP-4-amino-6-deoxy-Man</p>           |

| Organism   | PDB (Variant)                      | Ligands   | Ref | Enzyme & Reaction   |
|--|------------------------------------|---|-----|---|
| <i>Helicobacter pylori</i>   | 2FN6<br>2FNI<br>2FNU               | Apo<br>PLP<br>PMP/UDP-GlcNAc  | 145 | <p>UDP-2-acetamido-4-amino-6-deoxy-β-L-AINAc transaminase (PseC)</p> <p>UDP-4-keto-6-deoxy-AINAc      UDP-4-amino-6-deoxy-AINAc</p> |
| <i>Salmonella enterica serovar Typhi</i>   | IMDX<br>IMDO<br>IMDZ               | PLP/Oxaloglutaric acid<br>PLP<br>Cycloserine-PLP                                  | 144 | <p>UDP-4-amino-L-arabinose transaminase (ArnB) (EC 2.6.1.87)</p> <p>UDP-4-keto-arabinose      UDP-4-amino-arabinose</p>             |
| <i>Campylobacter jejuni</i>  | 1O69                               | 2-amino-PLP   | 146 | <p>UDP-4-amino-6-deoxy-D-GlcNAc transaminase (PglE)</p> <p>UDP-4-keto-6-deoxy-GlcNAc      UDP-4-amino-6-deoxy-GlcNAc</p>            |
| <i>Pseudomonas aeruginosa</i>  | 3NYU<br>3NYS(K188A)<br>3NYT(K185A) | Internal aldimine<br>PLP<br>PLP external aldimine adduct with UDP-3-amino-GlcNAcA | 147 | <p>UDP-3-amino-2-acetamido-glucuronic acid transaminase (WbpE)</p> <p>UDP-GlcNAc-3-keto-A      UDP-GlcNAc-3-amino-A</p>             |
| <b>Sugar O-/N-/C-methyltransferases (SMTs) (EC 2.1.1.-) (Section 5.1 in this review) (Rossmann fold)</b> |                                    |   |     |   |

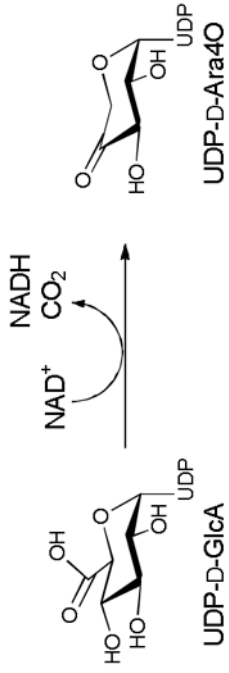
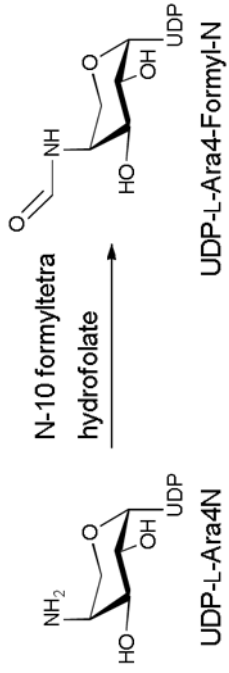
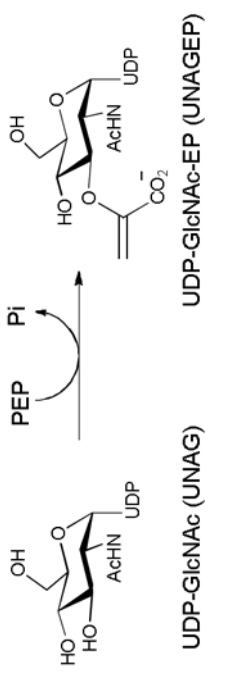
| Organism                             | PDB (Variant)                              | Ligands  | Ref | Enzyme & Reaction   |
|--------------------------------------|--|--|-----|---|
| <i>Lechevalieria aerocolonigenes</i> | 3BUS                                       | SAH  | 158 |  <p>D-Glc</p> <p>4-methoxy-D-Glc</p>   |
| <i>Streptomyces spheroides</i>       | 2WK1                                       | SAH  | 159 |  <p>Novobiocin 4'-O-MeT (NovP)</p> <p>4-demethoxy-L-noviose</p> <p>L-noviose</p>                     |
| <i>Micromonospora griseorubida</i>   | 3SSO<br>3SSN<br>3SSM                       | SAH<br>SAH/Mycinaamicin VI<br>SAH  | 160 |  <p>Mycinamicin 2-O-methyltransferase (MycE)</p> <p>6-deoxyallose</p> <p>2-methoxy-6-deoxyallose</p> |
| <i>Streptomyces fradiae</i>          | 3PFH<br>3PFG<br>3PX3(H123A)<br>3PX2(H123N) | SAH/dTDP-Quip3N<br>SAM/dTDP-Phenol<br>SAH/dTDP-Quip3N<br>SAH/dTDP-Quip3N | 161 |  <p>C3-N,N,-dimethyltransferase (TylM1)</p> <p>dTDP-3-amino-6-deoxy-Glc</p> <p>dTDP-mycaminose</p> |

| Organism   | PDB (Variant)                        | Ligands   | Ref    | Enzyme & Reaction  |
|--|--------------------------------------|---|--------|--|
| <i>Streptomyces venezuelae</i>                       | 3BXO                                 | SAM/UDP-Phenol  | 162    |  <p>Pikromycin C3-N,N,N-dimethyltransferase (DesVI)</p> <p>dTDP-3-amino-4,6-dideoxy-Glc + SAM → dTDP-desosamine + SAH</p>   |
| <i>Micromonospora chalybeata</i>                     | 3NDJ<br>3NDI                         | SAH/dTDP-3-amino-2,3,6-trideoxy-4-keto-3-methyl-D-glucose,<br>SAH/dTMP            | 163    |  <p>Kijanimicin affiliated C3-methyltransferase (tetronitrose) (TcaB9)</p> <p>dTDP-3-amino-2,3,6-trideoxy-4-keto-3-methyl-D-glucose + SAM → dTDP-3-amino-2,3,6-trideoxy-4-keto-3-methyl-D-glucose + SAH</p> |
| <b>5.2. Sugar N-Acetyltransferases</b>               |                                      |   |        |  |
| <b>left-handed-β-helix motif (LβH) superfamilies</b> |                                      |   |        |  |
| <i>Escherichia coli</i>                              | 2O15<br>2O16<br>2O17                 | UDP-GlcNAc/AcCoA<br>UDP-GlcNAc/CoA/GlcN-1-P<br>UDP-GlcNAc/desulpho-CoA/GlcNAc-1-P | 80     |  <p>Glucosamine-1-phosphate N-acetyltransferase (GlmU) (EC 2.3.1.157; EC 2.7.7.23)</p> <p>Glc3N-1-P + UDP-GlcNAc/AcCoA → Glc3NAc-1-P + UDP-GlcNAc/CoA/GlcN-1-P</p>  |
| <i>Mycobacterium tuberculosis</i>                    | 2QKX<br>3D8V<br>3D98<br>3D14<br>3DK5 | GlcNAc-1-P<br>UDP-GlcNAc<br>Apo<br>UDP-GlcNAc<br>Apo                              | 73, 74 |  <p>Glc3N-1-P + UDP-GlcNAc/AcCoA → Glc3NAc-1-P + UDP-GlcNAc/CoA/GlcN-1-P</p>  |
| <i>Yersinia pestis</i>                               | 3FWW                                 | Apo   |        |  |
| <i>Thermoanaerobacterium thermosaccharolyticum</i>   | 3FSC<br>3FS8<br>3FSC                 | CoA/dTDP-3,6-dideoxy-3-amino galactose<br>AcCoA<br>CoA/dTDP-3-amino-fucose        | 169    |  <p>dTDP-Quip3N-3-N-acetyltransferase (QdtC)</p> <p>dTDP-Quip3N + AcCoA → dTDP-Quip3NAc + CoA/dTDP-3,6-dideoxy-3-amino galactose</p>  |

| Organism                                       | PDB (Variant)  | Ligands   | Ref      | Enzyme & Reaction   |
|--|--|---|----------|---|
| <i>Campylobacter jejuni</i>                    | 3BSW<br>3BSY<br>3BSS<br>3VHE<br>3BFP<br>2NPO                         | Apo<br>AcCoA<br>UDP-2-acetamido-4-amino-6-deoxy-glucose<br>CoA<br>Apo<br>Apo  | 170, 171 | <p>UDP-QuitNAc4N-4-N acetyltransferase (PgID)</p>  <p>UDP-QuitNAc4N → UDP-QuitNAc4Nac</p>                              |
| <i>Bordetella petrii</i>                       | 3MQH<br>3MQG   | CoA/UDP-3-amino-2-acetamido-2,3-dideoxy glucuronic acid<br>AcCoA/UDP  | 172      | <p>UDP-GlcNAc3NA 3-N-acetyltransferase (WlbB) (EC 2.3.1.B6)</p>  <p>UDP-GlcNAc3NA → UDP-GlcNAc3Nac</p>                 |
| <i>Bacillus cereus</i>                         | 3VBJ<br>3VBI<br>3VBK(S84A)<br>3VBL(S84C)<br>3VBN(D94A)<br>3VBP(D94N) | dTDP/3-hydroxybutyryl-CoA<br>dTDP-4-amino-4,6-dideoxyglucose/CoA<br>dTDP-4-amino-4,6-dideoxyglucose/CoA<br>dTDP-4-amino-4,6-dideoxyglucose, CoA<br>dTDP, CoA<br>dTDP, CoA | 173      | <p>Acyltransferase (AnrD)</p>  <p>dTDP-4-amino-4,6-dideoxyglucose + CoAS → dTDP-4-amino-4,6-dideoxyglucyl-CoA + HO</p> |
| <b>GCN5-related N-acetyltransferase (GNAT)</b> |  |   |          |   |
| <i>Escherichia coli</i>                        | 2FT0<br>2FS5   | AcCoA<br>Apo  | 174      | <p>dTDP-fucosamine acetyltransferase (WecD)</p>  <p>dTDP-fucosamine + AcCoA → dTDP-fucosamine-4-Nac + CoA</p>        |
| <b>5.3. Oxidoreductases (N-oxidases)</b>       |  |   |          |   |

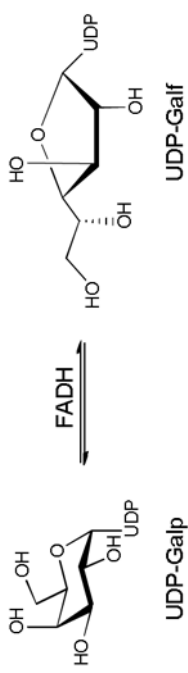
| Organism  | PDB (Variant) | Ligands | Ref | Enzyme & Reaction                  |
|---|---------------|---------|-----|------------------------------------|
| <i>Micromonospora carbonacea</i> var. <i>africana</i> | 3MXL          | Apo     | 182 | <p>Nitrosynthase-EvdC or ORF36</p> |
| <i>Actinonadurea kijaniata</i>                        | 3M9V          | dTDP    | 185 | <p>Nitrosynthase-KijD3</p>         |

5.4. Dehydrogenase/Decarboxylase and Formyltransferase (ArmA) (EC 1.1.1.305)

| Organism   | PDB (Variant)  | Ligands   | Ref  | Enzyme & Reaction  |
|--|--|---|--|--|
| <i>Escherichia coli</i>  | IU9J<br>2BLL<br>IYRW<br>I273(S433A)<br>I274(R619Y)<br>I275(R619M)<br>I27B(R619E)<br>I27E<br>2BLN | Apo-DH domain<br>Apo-DH domain<br>Apo-FDH domain<br>Apo-DH domain<br>Apo-DH domain<br>Apo-DH domain<br>ATP, UDP-GlcA-Full-length<br>UMP, N-5FDH- FDH domain | 188, 189, 195, 196   |  <p>UDP-D-GlcA → UDP-D-Ara4O + NADH + CO<sub>2</sub></p>  <p>UDP-L-Ara4N + N-10 formyltetrahydrofolate → UDP-L-Ara4-Formyl-N + hydrofolate</p> |
|  |  |   |  |  |
| <b>5.5. UDP-N-acetylglucosamine Enolpyruvyltransferase (MurA) (EC 2.5.1.7)</b> |  |   |  |  |
| <i>Escherichia coli</i>  | IUA6<br>1A2N(C115A)<br>3ISS(N67D)<br>2Z2C  | UDP-GlcNAc/fofomycin<br>Tetrahedral intermediate<br>Enolpyruvyl-UNAM<br>Cnicin  | 201, 204, 206  |  <p>UDP-GlcNAc + PEP → UDP-GlcNAc-EP (UNAGEP) + Pi</p>   |
|  | <i>Enterobacter cloacae</i>  | INAW<br>IDLG (C115S)<br>IEYN<br>IEJD<br>IEJC<br>IQ3G (D305A)<br>IRYW(C115S)<br>3LTH<br>3KQJ<br>3KR6<br>IYBG<br>3KQA   | Apo<br>Apo<br>ANS<br>Apo<br>Apo<br>Tetrahedral intermediate<br>UDP-GlcNAcEP<br>UDP-GlcNAc, fofomycin<br>UDP-GlcNAc<br>Fofomycin<br>Inhibitor T6361<br>Terreic acid | 202, 205, 209  |
| <i>Haemophilus influenzae</i>  | 2RL1<br>2RL2   | UDP-GlcNAc<br>UDP-GlcNAc, fofomycin   | 203  |  |

5.6. UDP-galactopyranose mutase (UGM) (EC 5.4.99.9)



| Organism                          | PDB (Variant)  | Ligands  | Ref      | Enzyme & Reaction   |
|-----------------------------------|--|--|----------|---|
| <i>Escherichia coli</i>           | 1I8T   | FAD  | 216      |  |
| <i>Mycobacterium tuberculosis</i> | 1V0J   | FAD  | 217      |   |
| <i>Klebsiella pneumoniae</i>      | 1WAM<br>2B17<br>2B18<br>3GF4<br>3INR<br>3INT<br>3KYB | FADH <sup>-</sup><br>FAD<br>FADH <sub>2</sub><br>UDP-Glc/FAD<br>UDP-Gal/FAD<br>UDP-Gal/FADH <sub>2</sub><br>FAD/FMIN | 217, 218 |   |
| <i>Deinococcus radiodurans</i>    | 3M14<br>3HE3<br>3HDO<br>3HDY                         | FAD/UDP-Gal(phosphonate analog)<br>FAD/UDP<br>FAD/UDP-Gal<br>FADH <sub>2</sub> /UDP-Gal                              | 219, 220 |   |