Division of Molecular Structural Biology, Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

STRUCTURAL BIOLOGY OF CARBOHYDRATE TRANSFER AND MODIFICATION IN NATURAL PRODUCT BIOSYNTHESIS

Magnus Claesson

Karolinska Institutet

Stockholm 2013

Published by Karolinska Institutet. Printed by Larseriks Digital Print AB

ISBN 978-91-7549-005-2

ABSTRACT

Certain organisms, can during periods of limited resources, adapt their metabolism to enable biosynthesis of secondary metabolites, compounds that increase competitiveness and chances of survival. The subjects of this thesis are enzymes acting on carbohydrate substrates during secondary metabolism.

The enzymatic attachment of carbohydrate moieties onto precursors of polyketide antibiotics such as anthracyclines, required for their biological activity, is performed by glycosyltransferases (GT). The anthracycline nogalamycin contains two carbohydrates: a nogalose moiety attached via an *O*-glycosidic bond to C7, and a nogalamine attached via an *O*-glycosidic bond to C1 and an unusual carbon-carbon bond between C2 and C5´´ of the sugar. Genetic and functional data presented in this thesis established the roles of SnogE as the GT performing the C7 *O*-glycosyl transfer of the nogalose moiety and SnogD as the *O*-GT attaching the nogalamine moiety onto the C1 carbon. The activity of SnogD was verified *in vitro* using recombinant protein, following establishment of a transglycosylation-like assay. The three-dimensional structure of the homo-dimeric SnogD was determined to 2.6 Å and consists of a GT-B fold. Mutagenesis of two active site residues, His25 and His301, evaluated *in vitro* and *in vivo*, suggested His25 to be the catalytic base, activating the acceptor substrate by proton abstraction from the C1-hydroxyl group. His301 provides a positive charge to stabilise the negative charge formed close to the diphosphate of the leaving group during glycosyl transfer. Genetic, functional and structural data together suggest the involvement of an additional or altogether different enzyme for the C-C bond formation.

The bifunctional enzyme aldos-2-ulose dehydratase (AUDH) from *Phanerochaete chrysosporium* catalyses the dehydration and isomerisation of the secondary metabolites glucosone and 1,5-anhydro-D-fructose (AF) into the antimicrobial compounds cortalcerone and microthecin (Mic), respectively. The three-dimensional structure of the dimeric AUDH was determined to 2.0 Å. The enzyme consists of a seven bladed β-propeller, two cupin folds and a lectin-like domain, in a novel combination. Two structural metal ions, Mg^{2+} and Zn^{2+}, are bound in loop regions. Two additional zinc ions are present at the base of two putative active sites, located in the β-propeller and the second cupin fold. The specific removal of these zinc ions eliminated catalytic activity, proving the metal dependency of the overall reaction. The structure of AUDH in complex with the reaction intermediate ascopyrone M bound at both putative active sites, and a complex of zinc-depleted enzyme with AF bound in the cupin fold have been determined by X-ray crystallography to 2.6 and 2.8 Å resolution, respectively. These observations support the presence of two distinct active sites located 60 Å apart, partly connected by an intra-dimeric channel. The dehydration reaction most likely follows an elimination reaction with the zinc ion acting as a Lewis acid to polarise the C2 keto group of AF. Abstraction of the C3 proton by the suitably located residue His155 would generate an enol intermediate, which is stabilised by the zinc ion. Return of the proton to the C4 hydroxyl group would generate a favourable leaving group.

LIST OF PUBLICATIONS

I. Vilja Siitonen, Magnus Claesson, Pekka Patrikainen, Maria Aromaa, Pekka Mäntsälä, Gunter Schneider and Mikko Metsä-Ketelä. Identification of late-stage glycosylation steps in the biosynthetic pathway of the anthracycline nogalamycin. ChemBioChem, 2012, 13, 120-128.

II. Magnus Claesson, Ylva Lindqvist, Susan Madrid, Tatyana Sandalova, Roland Fiskesund, Shukun Yu and Gunter Schneider. Crystal Structure of Bifunctional Aldos-2-Ulose Dehydratase/Isomerase from Phanerochaete chrysosporium with the Reaction Intermediate Ascopyrone M. J. Mol. Biol. 2012; 417, 279-293.

III. Magnus Claesson, Vilja Siitonen, Doreen Dobritzsch, Mikko Metsä-Ketelä and Gunter Schneider. Crystal structure of the glycosyltransferase SnogD from the biosynthetic pathway of nogalamycin in Streptomyces nogalater. FEBS J. 2012; 279, 3251-3263.

CONTENTS

LIST OF ABBREVIATIONS

AclK	*Streptomyces galilaeus* glycosyltransferase K
ACP	Acyl carrier protein
AknS	*Streptomyces galilaeus* glycosyltransferase S
AF	1,5-anhydro-D-fructose
AFOX	1,5-anhydro-D-fructose oxime
APM	Ascopyrone M
APP	Ascopyrone P
APT	Ascopyrone T
AUDH	Aldos-2-ulose dehydratase
BOG	β-octyl glycoside
CAZy	Carbohydrate Active Enzymes (http://www.cazy.org/)
CDP	Cytosine-5´-diphosphate
dUDP	2-deoxyuridine-5´-diphosphate
GDP	Guanine-5´-diphosphate
GT	Glycosyltransferase
EDTA	Ethylene-diamine-tetraacetic acid
FAS	Fatty acid synthase
LDP	Lignin degrading peroxidase
LGC	Lignocellulose
LIC	Ligation independent cloning
Mic	Microthecin
NMR	Nuclear magnetic resonance
NADPH	Nicotinamide adenine dinucleotide phosphate
PCR	Polymerase chain reaction
PDB	Protein Data Bank (http://www.ebi.ac.uk/pdbe)
PKS	Polyketide synthase
NDP	nucleotide-5´-diphosphate
SAM	S-adenosylmethionine
SAH	S-adensylhomocysteine
sno	*Streptomyces nogalater* gene cluster containing the genes required for biosynthesis of nogalamycin
SGC	Structural Genomics Consortium
SnogD	*Streptomyces nogalater* glycosyltransferase D
SnogDm	Reductively methylated form of SnogD
SnogE	*Streptomyces nogalater* glycosyltransferase E
SnogZ	Putative *Streptomyces nogalater* glycosyltransferase Z
rmsd	Root mean square deviation
TDP	Thymidine-5´-diphosphate
TDPG	Thymidine-5´-diphosphophoglucose
TTP	Thymidine-5´-triphosphate
UDP	Uridine-5´-diphosphate
wt	Wild type
Å	Ångström (10^{-10} m)

1 INTRODUCTION

1.1 SECONDARY METABOLISM AND ANTIBIOTICS

Certain organisms, including microbes, fungi, plants and animals, carry genes that are not obligate for survival but increase the survivability and fecundity of the organism. These genes enable the secondary or special metabolism, limited to periods of low growth rates, during which biosynthesis of e.g. antibiotics and pigments take place. The energy invested into biosynthesis of antibiotics is rewarded by a reduction in competition with other organisms for nutrients, providing an increased chance of survival and a competitive advantage in the microclimate of the organism [1]. Antibiotics are molecules with bactericidal or antibacterial effect, killing or limiting growth of bacteria, and include large groups of chemically diverse compounds.

The dawn of antibiotic research is attributed to the discovery of penicillin, from *Penicillium notatum,* in 1928 by Sir Alexander Flemming. The medical implications became obvious after introduction of stabilising modifications in the 1940's by Howard Florey and Sir Ernst Boris Chain, resulting in the first medical treatment using penicillin. The apparent potential of natural products as sources of bioactive compounds sparked large scale world-wide screening in the 1950's to 1970's, bringing attention to the *Streptomyces* genus as one of the most important sources of secondary products. Since then lichens and fungi have attracted interest as additional sources of natural products. The soil dwelling gram positive *Streptomyces,* belonging to the Actinobacteria phylum, produce a great diversity of bioactive compounds, with only a subset proving to have a useful pharmacology, i.e. to be biologically active but not excessively toxic.

Originally derived from natural sources, antibiotics are today mainly generated either chemically or by modification of naturally produced compounds in a semisynthetic fashion. Production via modification of natural compounds is particularly important due to the innate complexity of the chemistry involved, which prevents synthesis either altogether or in sufficient quantities at an acceptable cost.

The biosynthesis of natural products has been extensively studied at the genetic level, and this is particularly true for *Streptomyces*. Moreover, more and more details at the protein level have emerged during the last 20 years. The resulting genetic, structural and enzymatic insights have revealed many of the molecular requirements for biosynthesis, and have highlighted the potential for the production of new compounds with better pharmacological properties by combinatorial biosynthesis or enzyme redesign [2].

1.2 POLYKETIDE ANTIBIOTICS

Polyketide natural products have profound commercial and medical importance, stemming from their extensive chemical diversity [3]. The biosynthesis of polyketides and fatty acids have several common features, e.g. utilisation of basic metabolic building blocks as starting material [4], [5]. Polyketide biosynthesis is initiated by a polyketide synthase (PKS) [6–8]. Three major superfamilies of PSKs have been identified; type I and II, which act in a manner similar to that of the fatty acid synthase (FAS) and both utilise acyl carrier protein (ACP), and type III, which in contrast do not require ACP [7]. The type I PKS include both modular and iterative synthases. The modular type I PKS are megasynthases consisting of large multifunctional proteins,

where the biosynthesis reactions proceed in different active sites in a manner resembling an assembly-line, and produce reduced polyketides. The iterative type PKS produce either reduced or aromatic polyketides. The type III iterative PKS, which are present in plants, fungi and bacteria, consist of a single polypeptide chain, containing multifunctional active sites performing all biosynthesis steps. Biosynthesis by these enzymes typically yields aromatic polyketides.
The type II PKS also use an iterative mode of chain elongation and consist of an assembly of several distinct polypeptide chains harbouring the active sites, which catalyse individual steps in the biosynthesis of the typically aromatic polyketide. The anthracyclines are produced by a PKS type II, and the following discussion will be focused on the anthracyclines.

1.3 ANTHRACYCLINES

The anthracyclines include compounds with anti-bacterial (oxytetracycline/rifamycin), anti-fungal (pramidicin), cytostatic (doxorubicin), anti-viral (A-74528), cholesterol reducing (lovastatin), antiparasitic (frenolicin) and immunosuppresant (FK506) activities. Following the isolation of anthracyclines from rhodomycin producing strains of *Streptomyces purpurascens* [9], soil sample screening in the 1950's resulted in compounds with anticancer activity, sparking the "golden age" of antibiotic discovery. Amongst the thousands of compounds isolated, only a fraction have proven to be of sufficiently low toxicity to be therapeutically useful. In 1974 doxorubicin was approved by the Food and Drug Administration for treatment of cancer, and today several anthracycline drugs are amongst the most frequently used compounds for treatment of cancer. Therapeutic use of anthracyclines is associated with a cumulative toxicity, affecting primarily the cardiomyocytes and causing lifelong diastolic or systolic dysfunction, which restricts their long-term use [10]. The underlying mechanisms causing toxicity are not completely understood. The current toxicity-models are linked to oxidative stress, and/or partial intracellular metabolism of the drug, which reduces drug efflux by introduction of alcohol groups resulting in the accumulation of a persisting toxic reservoir [11].

1.3.1 Anthracycline biosynthesis

Anthracycline polyketides are synthesised from common metabolic intermediates such as acetyl- and malonyl-CoA, and synthesis is initiated by the PKS. The PKS synthesis is primed by co-enzyme A activated esters of short chain fatty acids (e.g. acetyl-CoA), with subsequent condensation of extender units (e.g. malonyl-CoA) through Claisen condensation followed by decarboxylation, resulting in a linear chain (Fig. 1.1). Cyclases, aromatases, hydroxylases and methylases modify the polyketide, resulting in the planar aromatic and tetracyclic 7,8,9,10-tetrahydro-5,12–naphtaceno-quinone structure. Chemical diversity is introduced by variations of the substitution pattern of the tetracyclic core and addition of carbohydrates [12].

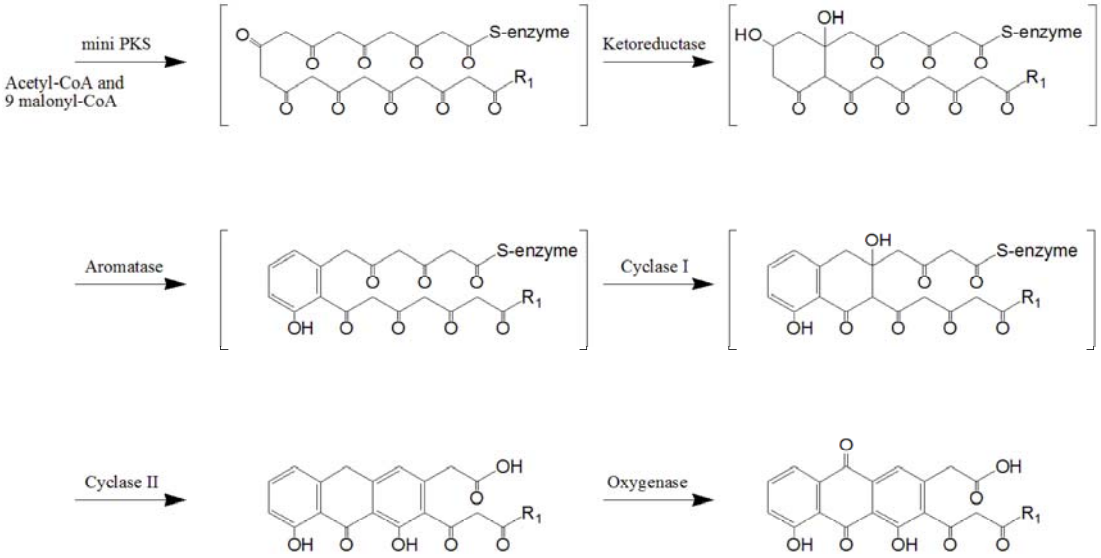

Figure 1.1 – Schematic representation of polyketide assembly. In nogalonic acid R1 is an ethyl group, for aklanonic acid R1 is a methyl group and the consumed metabolites are one propionate and 9 acetates.

The anthracycline nogalamycin (**1**) is produced by *Streptomyces nogalater* and contains two unusual deoxy-carbohydrates: the amino-sugar nogalamine attached in an unusual bicyclic configuration and the neutral nogalose (Fig. 1.2). The structural features of these carbohydrates make this compound interesting from a biosynthetic point of view. The structure of nogalamycin was determined by X-ray crystallography in 1983 [13], and subsequent complex structures with DNA provided detailed information on binding interactions [14–16]. Extensive efforts to generate new compounds based on nogalamycin were made during the 1970´s, but these experiments failed as a result of poor toxicity profiles [17]. Menogaril, which emerged as the most promising candidate, failed to proceed beyond phase II clinical trials during the early 1990´s.

The polyketide core of nogalamycin, nogalamycinone, is synthesised from one acetyl-CoA and nine malonyl-CoA units [18], by the action of an iterative PKS type II pathway [19]. The mini PKS type II consists of four distinct subunits; ACP, malonyl-CoA malonyltransferase, ketosynthase and the chain length factor subunits, which regulate the chain length. The highly reactive poly-β -ketone is cyclised, starting with the D ring, by cyclases and aromatases, which enforce the formation of the correct tetracyclic core of the anthracyclines [20]. Oxidation at C12 by the small cofactor-independent monoxoygenase SnoaB produces the nogalonic acid [21] (Fig. 1.2). Following *O*-methylation of the C14-hydroxyl group by SnoaC [22], the fourth and last ring is closed by an intramolecular aldol condensation reaction catalysed by SnoaL [23]. Ketoreduction at C7 by the nicotinamide adenine dinucleotide phosphate (NADPH) dependent SnoaF results in a hydroxyl group, which in turn is the point of attachment of the noagalose moiety – a reaction catalysed by SnogE [24]. The final tailoring step of the aglycone is introduction of a hydroxyl group at C1 by the recently discovered two-component monoxygenase SnoaW/SnoaL2, thus enabling subsequent glycosyl transfer of the second carbohydrate, the nogalamine moiety by SnogD [24–26]. Following glycosyl transfer, additional modifications of the carbohydrates are introduced. The importance of the attached carbohydrate for biological activity is well established; the sugar moieties are important for solubilisation, uptake and interaction with the biological targets [27], [28].

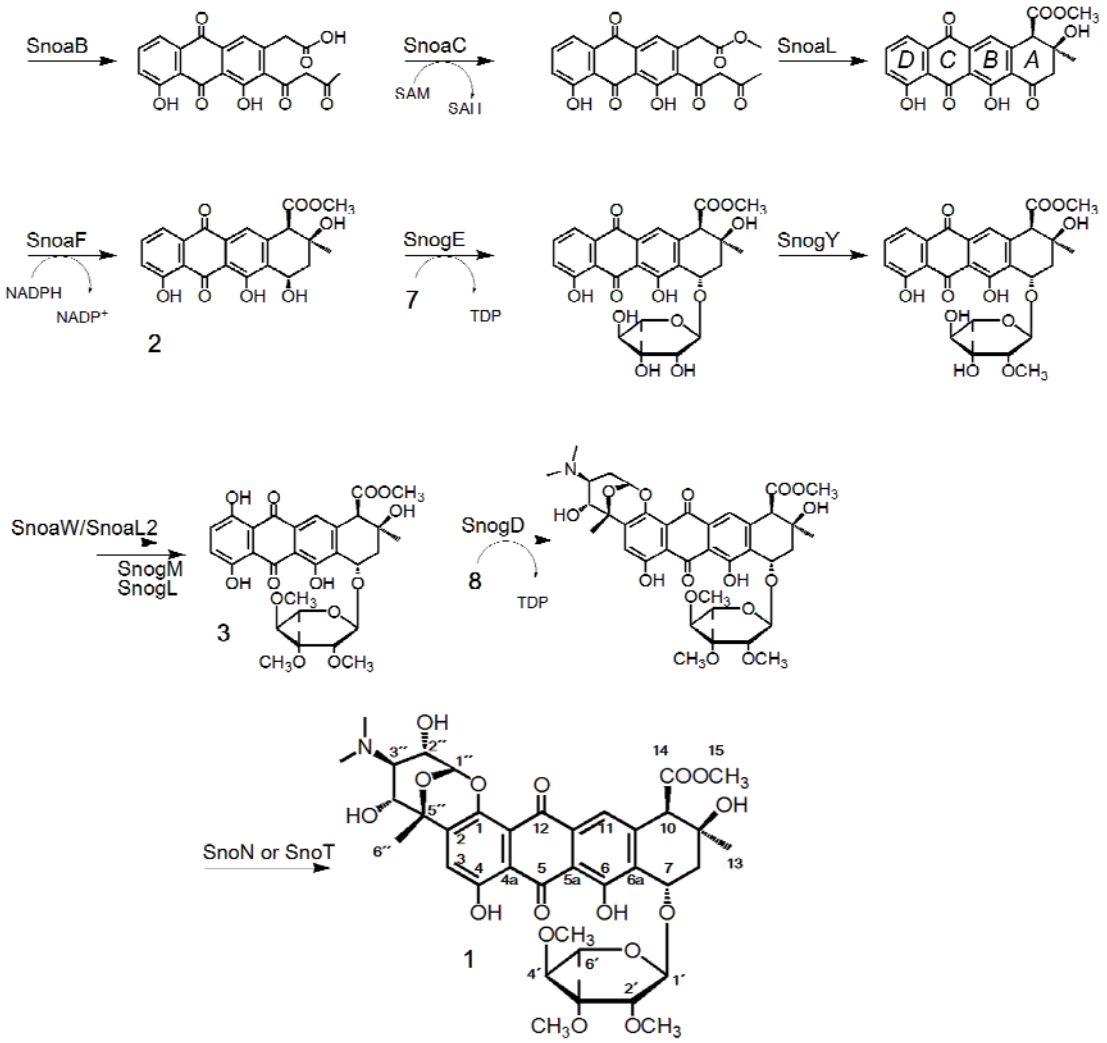

Figure 1.2 – Model pathway for biosynthesis of nogalamycin (**1**), from nogalonic acid (continuation from Fig. 1.1), via the recently discovered intermediates nogalamycinone (**2**) and 3´,4´-demethoxy-nogalose-1-hydroxynogalamycinone (**3**) [24]. The likely donor substrates for glycosyl transfer are TDP-2,3,4-tridemethoxy nogalose (**7**) and TDP- L -acosamine (**8**).

1.3.2 Enzymes from nogalamycin biosynthesis with previously determined structures

The structures of SnoaB, SnoaL and SnoaL2 from the nogalamycin biosynthetic pathway have previously been determined in our group (Fig. 1.3). The fold of SnoaB resembles the ferrodoxin-type α + β sandwich fold (Fig 1.3A) [21], and the cofactor independent monoxygenation reaction introduces oxygen to the C12 carbon, via a carbanion mechanism. The enzyme deprotonates the substrate, which reacts with molecular oxygen via a single electron transfer. The formed hydroperoxy- anion intermediate is subsequently protonated, resulting in nogalonic acid and water [21]. The structures of SnoaL and SnoaL2 are similar and superimpose with a root mean square deviation (rmsd) of 2.4 Å, in spite of only 20% sequence identity and quite different chemistry catalysed. The overall fold of the two proteins resembles a distorted α + β barrel (Fig. 1.3B&C). The novel cyclisation reaction of SnoaL does not proceed via a Schiff-base, nor does it require any cofactors. Instead proton abstraction from the C10 carbon atom is facilitated by acid-base chemistry using an invariant

aspartic acid (Asp121). The resulting enolate intermediate is stabilised by delocalisation over the π-system of the neighbouring rings. The cyclisation reaction is completed by a nucleophilic attack of the enolate onto the C9 carbon, followed by a proton transfer yielding nogalaviketone [23]. The mechanism of C1 carbon hydroxylation was recently proposed to proceed via a SnoaW catalysed reduction of the antraquinone ring in an NADPH dependent manner. The formed dihydroquinone would subsequently activate molecular oxygen yielding a C1 peroxy-intermediate, which following protonation by SnoaL2 generates the C1 hydroxylated product [26].

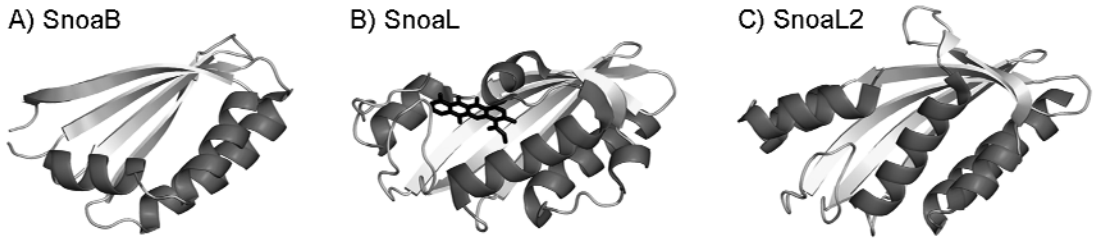

Figure 1.3 – Cartoon representations of previously determined structures of nogalamycin biosynthetic enzymes. A) The monooxygenase SnoaB (PDB ID: 3KNG, resolution: 1.7 Å). B) The cyclase SnoaL in complex with the product nogalaviketone shown as sticks (PDB ID: 1SJW, resolution: 1.35 Å) C) The C1-hydroxylase SnoaL2 (PDB ID: resolution: 2GEX, 2.5 Å).

1.3.3 Nogalamycin carbohydrate biosynthesis in *S. nogalater*

Biosynthesis of the two carbohydrate moieties of nogalamycin is predicted based on gene cluster homology to require a multitude of enzymes, metabolising the common precursor TDP-glucose into the neutral deoxysugar nogalose and the dideoxy aminosugar nogalamine [22]. Both carbohydrates originate from the common metabolite α-D-glucose-1-phosphate, which is transferred onto the nucleotide by the thymidylyltransferase SnoaJ, producing the activated form of the carbohydrate (Fig. 1.4A). The nucleotide-activated carbohydrate undergoes 4′,6′-dehydratisation to the 4′-keto-6′dehydroxy-form, catalysed by SnogK. From this metabolite the carbohydrate biosynthesis diverges (Fig 1.4B&C).

In nogalose biosynthesis, a 3′,5′-epimerisation by SnogF follows, generating the TDP-4′-keto-6′-deoxy-L- mannose. This is likely achieved by a similar mechanism as in the well-studied reaction of RmlC from *Salmonella enterica*, proceeding via deprotonation of C3 and C5 by the conserved His65, with the second member of the catalytic dyad Asp171 facilitating proton abstraction [29], [30]. The resulting enolate intermediates are stabilised by Lys74, while the subsequent protonation that completes the epimerisation step is mediated by Tyr140. Methylation of C3′ is predicted to be performed by SnogG2, following the mechanism of the homologous *C*-methyltransferase TylC3 from the biosynthesis of tylosin in *Streptomyces fradiae*, proceeding via proton abstraction from C3. The resulting enolate intermediate reacts with the electrophilic methyl group of the co-substrate S-adenosylmethionine (SAM) [31]. Reduction of the 4′-ketone is putatively catalysed by SnogC. The subsequent reactions which produce the nogalalose moiety were suggested to occur after carbohydrate transfer onto the aglycone [22], [32], a prediction supported by recent *in vivo* data (paper I). *O*-methylation of the C2′ carbon atom is performed by SnogY and *O*-methylations of the C3′ and C5′ carbon atoms are probably associated with the putative *O*-methyltransferases SnogM and SnogL (Fig. 1.2) [22], [24].

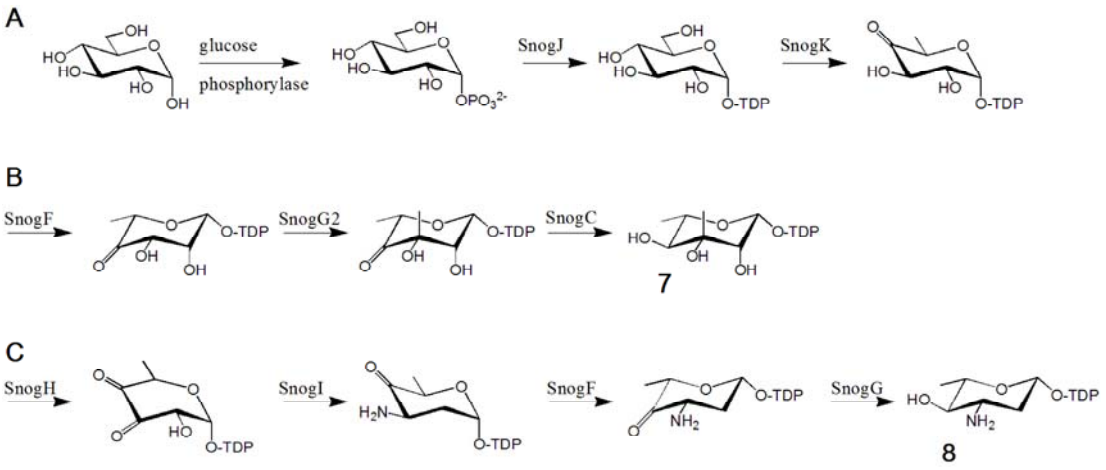

Figure 1.4 – Biosynthesis of nogalamycin carbohydrate moieties. A) Generation of TDP-4′-keto-6′-deoxy-α-D-glucose. The subsequent steps are shown for the nogalose moiety in B) and the nogalamine moiety in C), resulting in the activated forms of the carbohydrates **7** and **8**, likely transferred by the respective glycosyltransferases (B- SnogE, C-SnogD [22], [32], [33]).

Formation of TDP-nogalamine is predicted to follow the typical pathway of aminosugar biosynthesis (Fig 1.4C). The 4′-keto-6′dehydroxy-form of the TDP-carbohydrate, formed by SnogK, is converted into a reactive 3′,4′-diketo-2′-dehydroxy intermediate by SnogH. This reaction may proceed as a dehydration reaction similar to that catalysed by TylX3 [34], using a Zn^{2+} activated water molecule as base for the C3 deprotonation or to stabilise the enolate intermediate. The intermediate subsequently undergoes β-elimination resulting in the ketone form of the C2′′-oxygen, followed by stereo-specific introduction of a solvent derived proton at C2′′ [34], [35]. The resulting bi-ketide form of the carbohydrate would enable the subsequent transamination at C3′′. This reaction, putatively catalysed by SnogI, is thought to follow a mechanism that is homologous to the pyridoxal 5′-phosphate (PLP)-dependent transamination reaction catalysed by DesV from the D-desosamine biosynthesis of *Streptomyces venezuelae,* using glutamic acid as amine donor [36]. The subsequent 5′′-epimerisation and 4′′-ketoreduction steps are proposed to be carried out by SnogF and SnogG, respectively [22]. As in the case of nogalose biosynthesis, additional tailoring reactions are performed after glycosyl transfer of the TDP- L -acosamine moiety by SnogD. The two *N*-methylation steps at the 3′amino group are probably performed by SnogA and SnogX, and followed by hydroxylation at C2′′ by the gene product of either *snoN* or *snoT*.

1.3.4 Glycosyltransferases

The attachment of sugar moieties onto biological macromolecules such as proteins, other carbohydrates, organic and inorganic substances is performed by a particular class of enzymes, the glycosyltransferases. The opposite process of removing carbohydrates is catalysed by hydrolases such as glycosidases, performing in essence a transfer onto water. The biosynthesis and hydrolysis of carbohydrates accounts for the bulk of anabolic biotransformation reactions in nature [37]. GT enzymes exist as globular soluble and membrane associated proteins. There are considerably more biochemical and structural data accumulated from the globular soluble enzymes [32]. Intracellular GT enzymes are present in all kingdoms of life, with additional GTs in the

pereplasmic space of bacteria and the sub-cellular compartments of eukayrotes, e.g. the endoplasmic reticulum [38] and Golgi [39].
The Carbohydrate Active Enzymes database (CAzY) classifies GT enzymes using mono- or di-phosphate nucleotide, lipid phosphate and phosphate activated donors into distinct sequence based families [40–42]. A total of 94 families are defined, based on the reaction performed and the substrates used, and more than 100000 carbohydrate interacting modules are described.
The reaction catalysed by GT enzymes (EC.2.4.x.x) is the transfer of an activated carbohydrate moiety from a donor-substrate onto an acceptor substrate, resulting in a glycosidic bond. The acceptor substrates are commonly other carbohydrates, but also include proteins, nucleic acids, lipids and small molecules such as antibiotics. The carbohydrate donors are typically classified into two groups, the nucleotide-activated (Leloir type) and those activated by other groups such as phospho-groups (non-Leloir type).
In terms of three-dimensional structure the individual GT enzymes belong to one of two occurring fold families, the GT-A and the GT-B, with the members of each family predicted to share the same fold (Fig. 1.5) [42], [43]. The GT-A fold, which was first observed in the structure of SpsA from *Bacillus subtilis* [44], consists of two dissimilar domains of different size, whereas the GT-B fold is characterised by two domains of similar size and fold, and was first observed in the T4 β-glucosyltransferase [45]. An open skewed β-sheet constitutes the centre of the GT-A fold, which is surrounded by α-helices. The fold bears resemblance to the Rossmann-like nucleotide binding fold, with the two β/α/β domains interacting with distinct acceptor- and nucleotide-substrate binding sites. GT-A enzymes frequently contain an Asp-X-Asp signature motif, which coordinates a divalent cation and/or ribose by the side chain carboxyl groups [46], [47]. Amino acid variations are however not uncommon amongst these residues, arguing against an overall sequence conservation [48]. In the GT-B fold the two β/α/β Rossmann-fold like domains are separated, and are interacting to a lesser degree compared with the GT-A fold. The central cleft formed between the two domains encompasses the active site, and the substrate binding occurs at the domain interface.

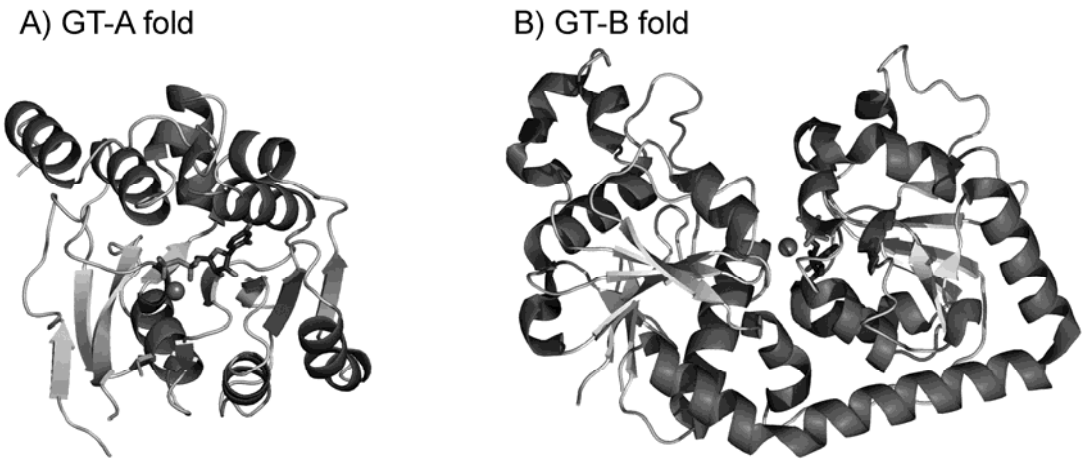

Figure 1.5 – The two folds of glycosytransferases. A) GT-A fold, as exemplified by SpsA from *Bacillus subtilis* (PDB ID: 1qgq) [49] in complex with UDP and Mn^{2+}. The domains are separated horizontally at the centre B) GT-B fold, as exemplified by T4 β-glucosyltransferase (PDB ID: 1jg7), in complex with UDP and Mn^{2+} [45], with domains separated vertically at the centre. Bound dinucleotide ligands are shown as sticks and metal ions as spheres.

GT catalysed transfer typically results in oxygen linkage, but other acceptor nucleophiles such as sulphur (thioglycosides in plants), nitrogen (*N*-linkages in glycoproteins) and carbon (*C*-linked glycoside antibiotics) have also been described [43], [50]. The GT catalysed reactions proceed through transition states similar to non-enzymatic sugar transfers, where a nucleophile and a leaving group interact weakly with a reaction centre that frequently carries a high degree of positive charge [51].

The GT enzymes are additionally classified into one of two classes, retaining or inverting, based on the stereochemical outcome of the catalysed reaction [52]. The carbohydrate transfer reaction results in either an inversion or retention of the donor anomeric carbon configuration. Each outcome is the result of an individual type of reaction chemistry (Fig. 1.6), analogous to the reactions catalysed by glycosidases [43].

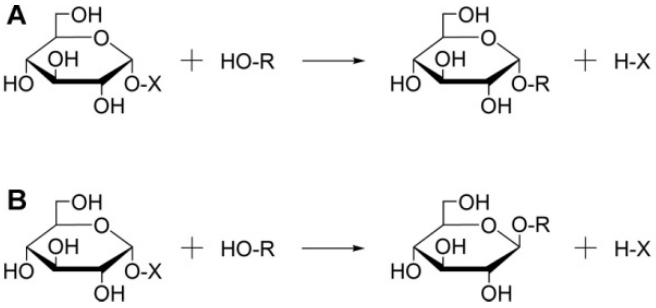

Figure 1.6 – The two stereochemical outcomes of glycosyl transfer by GT enzymes. A) Retaining reaction, maintaining the configuration of the anomeric carbon. B) Inverting reaction, causing an inversion of the anomeric carbon configuration.

Traditionally the reaction of retaining GT was postulated to proceed by removal of the donor-carbohydrate from its activating partner as a consequence of a nucleophilic attack performed by the enzyme, a process aided by a divalent cation (Mn^{2+}, Mg^{2+}) which is ubiquitously observed at the active site. The metal ion is coordinated by side chain carboxyl groups of acidic residues (Asp, Glu). Presence of a divalent cation stabilizes the developing negative charge of the donor substrate leaving group, thus facilitating the reorganisation of the covalent bond. Analogous to glycosyl hydrolases a mechanism for retaining glycosyl transfer was suggested proceeding via a double-displacement mechanism (Fig. 1.7A), during which a covalent intermediate between the enzyme nucleophile and the anomeric carbon of the donor-carbohydrate would be formed, as was observed for hen egg-white lysozyme [53]. This intermediate would subsequently be cleaved by a second nucleophilic attack, performed by the acceptor substrate aglycone, thus completing the carbohydrate transfer and regenerating the enzyme nucleophile for a subsequent reaction. The low degree of structural conservation at the postulated location of the catalytic nucleophile does however reduce the plausibility of this reaction chemistry [43], as does the absence of structures of GT enzymes with trapped covalent species [54].
In recent years there is increasing evidence for an alternative reaction mechanism, proceeding via an “internal-return” type mechanism, also referred to as S_Ni (substitution nucleophilic internal). Here the nucleophile would attack from the same face of the donor-carbohydrate as the leaving group with the glycosyl transfer proceeding via a transition state oxocarbenium ion, which is stabilised by the enzyme [43], [54] (Fig. 1.7B). The GT related results presented in this thesis concern two

inverting glycosyltransferases, both belonging to class 1, and hence the following description of GT enzymes will be limited to this class.

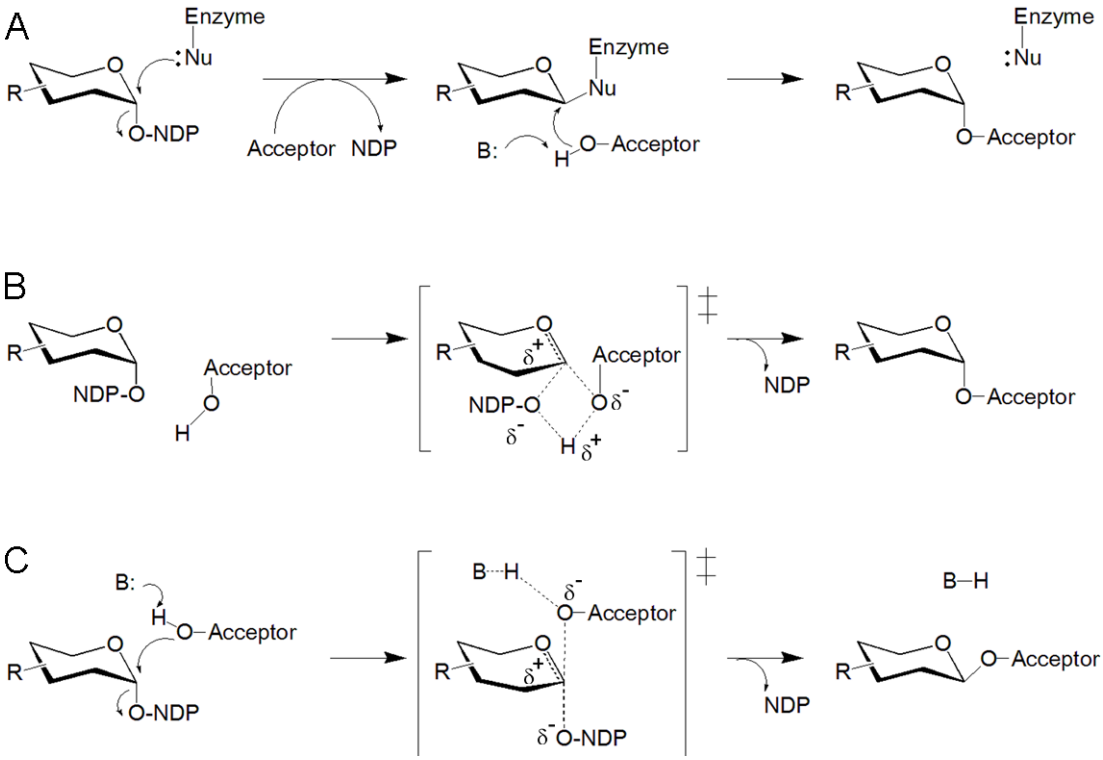

Figure 1.7 –The two major types of reaction mechanisms of GT. A) The double displacement mechanism for retaining glycosyl transfer. B) The alternate S_NI oxycarbenium intermediate mechanism of retaining glycosyl transfer, this has recently been suggested to proceed in two steps [54] C) The single displacement S_N2 type mechanism of inverting glycosyl transfer.

The inverting reaction catalysed by the GT-1 class proceeds via a single displacement (S_N2) reaction, where the acceptor substrate performs a nucleophillic attack onto the anomeric carbon (Fig. 1.7C). This process is often facilitated by the abstraction of a proton from the accepting hydroxyl group by an enzymatic base, commonly Asp or His side chains [55–59]. As the new bond between the acceptor substrate and the donor substrate is forming, the developing negative charge of the leaving group is stabilised by a positive charge in the vicinity, commonly supplied by enzyme side chains or helix dipoles rather than a divalent cation [43]. Today 22 structures of class 1 GT enzymes have been added to CaZY. In spite of their structural similarity, the overall sequence homology is moderate (Table 2.1).

Table 2.1 – Structural homology detected by DALI [60], between SnogD (PDB ID 4amb) and glycosyltransferases annotated in CAZy [42] as belonging to class 1.

Glycosyltransferase	Organism	Domain[a]	DALI score					PDB ID[e]
			Z-score	rmsd[b]	lali[c]	Nres[d]	% seq. id.	
Calicheamicin GT **CalG3**	*Micromonospora echinospora*	B	42.6	2.6	359	379	37	3oti[B]
NDP-olivose: tetracycline β-olivosyltransferase **SsfS6**	*Streptomyces sp. SF2575*	B	37.8	2.6	337	356	30	4g2t[A]
D-olivosyltransferase **UrdGT2**	*Streptomyces fradiae T#2717*	B	37.4	2.6	345	382	28	2p6p[A]
TDP-β-L-Rha: spynosin 9-O-α-L-rhamnosyltransferase **SpnG**	*Saccharopolyspora spinosa NRLL18537*	B	37.3	2.6	343	373	30	3uyk[A]
Calicheamicin GT **CalG1**	*Micromonospora*	B	37.1	3.6	355	391	30	3otg[A]

	echinospora							
TDP-desosamine: erythronolide desosaminyltransferase, **EryCIII**	*Saccharopolyspora erythraea NRRL 2338*	B	35.8	3.0	346	408	34	2yjn[A]
Calicheamicin GT **CalG2**	*Micromonospora echinospora*	B	33.2	3.1	344	397	23	3rsc[A]
Oleandomycin GT **OleI**	*Streptomyces antibioticus ATCC 11891*	B	33.2	2.7	338	392	25	2iya[A]
Calicheamicin GT **CalG4**	*Micromonospora echinospora*	B	31.3	3.1	341	397	26	3ia7[A]
Oleandomycin glycosyltransferase **OleD**	*Streptomyces antibioticus ATCC 11891*	B	28.3	4.4	339	394	23	2iyf[B]
UDP-β-L-4-epi-vancosamine: vancomycin-pseudoaglycone vancosaminyltransferase **GtfD**	*Amycolatopsis orientalis ATCC19795*	B	27.7	3.8	335	400	22	1rrv[B]
dTDP-β-L-4-epi-epivancosamine: epivancosaminyltransferase **GtfA**	*Amycolatopsis orientalis A82846*	B	27.5	3.5	332	391	24	1pn3[A]
UDP-Glc : flavonoid β-GT **UGT71G1**	*Medicago truncatula*	E	26.3	3.5	331	454	11	2acw[B]
multifunctional UDP-Glc : (iso)flavonoid β-GT **UGT85H2**	*Medicago truncatula*	E	26.0	3.0	320	443	16	2pq6[A]
UDP-Glc: sinapoyl-alcohol-, 2,5-DHBA-, 3,4-DHBA-GT **UGT72B1**	*Arabidopsis thaliana*	E	25.9	3.3	329	461	17	2vce[A]
TDP/UDP-Glc: aglycosyl-vancomycin: GT **GtfB**	*Amycolatopsis orientalis ATCC19795*	B	25.8	4.1	326	382	20	1iir[A]
UDP-Glc : (iso)flavonoid β-glucosyltransferase **UGT78G1**	*Medicago truncatula*	E	25.6	3.1	320	443	12	3hbf[A]
UDP-Glc: anthocyanidin 3-O-glucosyltransferase **VvGT1**	*Vitis vinifera*	E	25.5	3.2	315	434	15	2c1x[A]
UDP-GlcA: β-glucuronosyltransferase 2B7 **Ugt2b7**	*Homo sapiens*	E	18.3	2.3	152	166	18	2o6l[B]
UDP-N-acetylglucosamine transferase subunit **ALG13**	*Saccharomyces cerevisiae S288c*	E	10.9	3.4	143	201	14	2ks6[A]

[a] A-archaea, B –bacteria, E-eukayrota
[b] root mean square distance
[c] number of structurally equivalent residues
[d] number of residues in target protein
[e] percentage of identical amino acids over structurally equivalent residues of respective homologue to SnogD
[f] DALI matched chain in brackets

1.4 SECONDARY METABOLITES PRODUCED DURING DEGRADATION OF WOOD MATERIAL

Lignocellulose (LGC) biomass is the second most prominent organic polymer on earth, superseded only by cellulose. LGC is estimated to contain 30% of non-fossil organic carbon in the biosphere - a reservoir upheld by *de novo* biosynthesis in plants and some types of algae and degradation by certain fungi and bacteria [61]. LGC is composed by cellulose and hemicellulose polymers tightly cross-linked by lignin, and is present in the cell wall, for which the cross-linked polysaccharides provide mechanical stress resistance. The composition of lignin is heterogenous, with low restriction of primary structure, and the macromolecular assemblies may exceed 10000 Daltons in mass. The lignin building blocks are the monolignol units; p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, which vary in the degree of methoxylation. Cross-linking within the lignin polymers is typically extensive, and arises from radical-radical coupling reactions initiated by oxidative enzymes, by formation of monoglino radicals [61]. The complex and heterogeneous cross-linking of LGC requires a specific degradation machinery [62][63]. Ligninases performing part of the cleavage are present in a limited number of organisms belonging to the kingdoms of fungi and bacteria. Degradation of the lignin component and thereby mobilisation carbon, is performed by haem containing lignin peroxidases (LDP) (E.C.1.11.1.14), manganese peroxidases (E.C.1.11.1.13), versatile peroxidase (E.C.1.11.1.16) and copper containing laccases (E.C.1.10.3.2) [64]. The peroxidase typically generate the free radicals required for the depolymerisation reaction from hydrogen peroxide.

White rot fungi, belonging to the Basidomycota phyla, are predominant degraders of wood material, with the capacity to degrade lignin, cellulose and hemicellulose, commonly resulting in the typical white fibrous deposits, which are rich in cellulose. The brown rot fungi are less numerous (representing only 7% of wood rotting Basidomycota), which degrade cellulose following oxidation of and partial modification of lignin cellulose, and to a much lesser extent lignin [61]. *Phanerochaete chrysosporium* is the most extensively studied white rot fungus, and is regarded as an important organism for industrial pulp and biofuel production. It generates the required hydrogen peroxide substrate of the lignin peroxidase, using the flavine dependent enzyme pyranose-2-oxidase, which oxidizes pyranoses at the C2 position to the corresponding C2 ketoses [65–67]. The C2 ketose produced from glucose, presumably derived from cellulose, glucosone (D-arabino-hexosulose) may re-enter the carbohydrate metabolism after NADPH dependent reduction by pyranose-2-reductase into glucose. Alternatively it may be further enzymatically converted into the secondary metabolite cortalcerone (2-hydroxy-6H-3-pyrone-2-carboxaldehyde hydrate) [66], [68], [69] (Fig. 1.8). The discovery of cortalcerone from *Corticium coeruleum* extracts was reported in 1976 [70], and the enzyme catalysing the reaction, aldos-2-ulose dehydratase was later isolated and characterised from the red algae *Gracilariopsis lemaneiformis* [71], the morels *Morchella costata* and *M. vulgaris* [68] and the white rot fungus *Phanerochaete chrysosporium* [66], [69], [72].

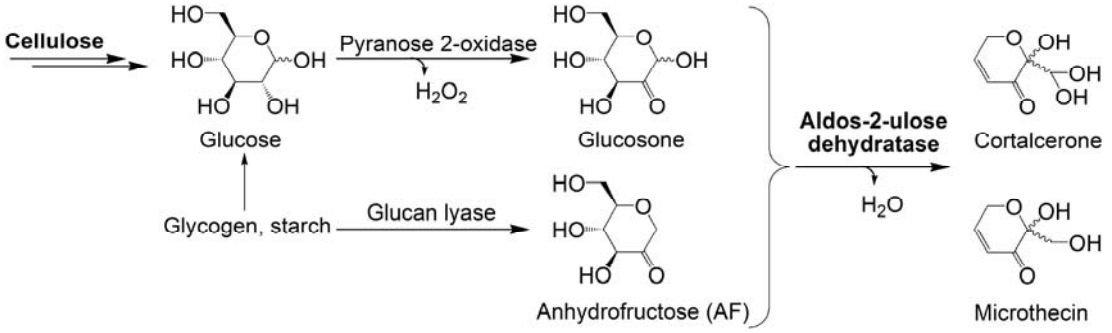

Figure 1.8 – Sources of glucosone and 1,5 – anhydro-D-fructose (AF), and enzymatic conversion into the secondary metabolites cortalcerone and microthecin (Mic).

In certain fungi and red marine algae, the bifunctional enzyme aldos-2-ulose dehydratase (AUDH) can also catalyse the conversion of 1,5-anhydro-D-fructose (AF), to the related metabolite microthecin (Mic). This secondary metabolite exhibits antibacterial activity against Gram-positive and Gram-negative bacteria, such as *Pseudomonas aeruginosa*, and cytotoxic actitiy against certain malignant blood cell lines [73]. In other fungi such as *Anthracobia melaloma*, AF is converted by 1,5-anhydro-D-fructose dehydratase (EC 4.2.1.111) into ascopyrone M (APM), which is subsequently modified by ascopyrone tautomerase (EC 5.3.3.15) resulting in ascopyrone P (APP) [74]. The metabolite APM is spontaneously hydrated in aqueous solutions to form the saturated acopyrone T, albeit at a low rate at neutral pH [75]. In bacteria and humans a NADPH-dependent reductase can convert AF into 1,5-anhydro-D-glucitol or 1,5-anhydro-D-mannitol [76][77].

1.4.1 The bifunctional enzyme aldos-2-ulose dehydratase

The AUDH catalysed production of microthecin proceeds in two steps, an initial dehydration of AF to APM and a subsequent complex isomerisation into the final product Mic [68], [71], [72], [78] (Fig. 1.9). The bifunctionallity of AUDH sets it aside amongst dehydratases from carbohydrate metabolism, where one enzyme commonly catalyse a single reaction [79].

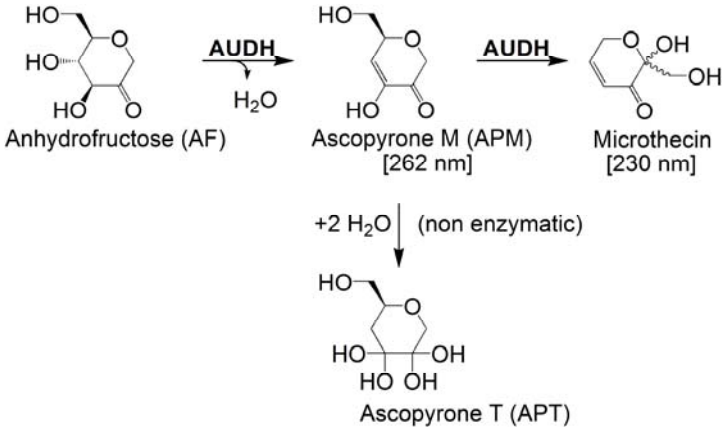

Figure 1.9 – The two reactions catalysed by AUDH.

The activity of AUDH from *P. chrysosporium* has been studied biochemically, where the two independent reaction steps can be followed spectroscopically at absorption maxima of the reaction intermediate APM and product Mic (262 and 230 nm, respectively) without interference by the substrate AF [72].

The second isomerisation reaction step catalysed by AUDH is altogether less straight forward than the dehydration reaction, with no examples of similar chemistry found by the author. Based on the structures of APM and Mic the isomerisation is easier to imagine proceeding via ring opening by addition of water, since extensive chemical modifications would otherwise be required to form Mic. These processes appear unlikely to be catalysed by a single enzyme. The dehydrated ring form of APM is however not hydrolysed spontaneously in aqueous solution, although addition of water to form ascopyrone T (APT) may occur [78]. This would indicate that the isomerisation reaction is performed enzymatically in a biological setting.

2 AIM OF THIS THESIS

The biosynthesis of medically relevant anthracyclines by *Streptomyces* has been studied since the emergence of doxorubicin/daunorubicin in the 1960's. These studies have resulted in novel antibiotics, as well as improved methods for and understanding of combinatorial biosynthesis. The role of the carbohydrate moieties has to a great extent been elucidated, however the carbohydrate biosynthesis and conjugation is less understood from a structure/function perspective. This is particularly the case for modified carbohydrates and unusual carbohydrate moieties, which likely require unidentified chemistry and where the overall carbohydrate biosynthesis can at the best predicted based on gene cluster analysis. Knowledge of these steps can prove valuable for combinatorial biosynthesis, with detailed information about catalysis and substrate specificity, thus greatly facilitating development of new antibiotics, potentially exhibiting improved toxicity profiles. Therefore we aimed to structurally characterise the three putative glycosyltransferases involved in nogalamycin biosynthesis, to elucidate their activities and to provide insights into both the substrate specificity and the catalytic reaction, which is particularly interesting due to the unusual C-C bond produced.

Structural elucidation of the bifunctional AUDH was motivated by the enigmatic catalysis performed by this large protein, which has no full length sequence homologues and shows only partial homology to non-characterised putative proteins. The intermediate and the final product, which both have anti-microbial activity, could be starting points for drug design. In addition the isomerisation step could be exploited for generation of new compounds of similar structure, with potentially enhanced biological activity.

3 RESULTS AND DISCUSSION

3.1 GLYCOSYL TRANSFER IN THE BIOSYNTHESIS OF NOGALAMYCIN (PAPERS I AND III)

The polyketide antibiotic nogalamycin, produced by *Streptomyces nogalater*, contains two carbohydrate moieties attached at opposite sides of the aglycone (Fig. 3.1). The nogalose moiety attached at C7 is similar to the L-rhamnose moieties incorporated into the macrolide spinosyn [80], the aromatic polyketide elloramycin [81] and the enedieyne calicheamines type antibiotics [82], but the bicyclic attachment of the amino-sugar nogalamine is considerably more exotic. In addition to the conventional *O*-glycosyl bond between the C1 hydroxyl group and the C1´´ of the carbohydrate, a covalent carbon-carbon (C-C) bond exists between C2 of the aglycone and the C5´´ of the nogalamine. The atoms forming the bonds between the deoxysugar and the aglycone are connected by an oxygen atom forming an ether bond. C-C bond attachment of carbohydrates is present in a limited number of other natural products, such as urdamycin [83], gilvocarcin [84], hedamycin [85] and granaticin [86], but the combination with the *O*-glycosyl bond is specific for nogalamycin. Hence the sequence of bond formation and chemistry resulting in the C-C bond between aglycone and the nogalamine moiety are intriguing. At the outset of this study characterisation of the three predicted glycosyltransferases from the nogalamycin biosynthetic pathway was expected to provide insights into the mechanisms of carbohydrate transfer, and in particular potentially into the formation of the unusual C-C bond linkage.

Until this study the late stage glycosylations and modifications of nogalamycin biosynthesis were not proven experimentally, but proposed based on gene cluster homology to different pathways [22]. Modifications such as *O*-methylations of carbohydrates were thought to occur after glycosyl transfer, based on lack of suitable genes predicted to encode TDP-binding and *O*-methyl transfer activity within the *sno* gene cluster.

3.1.1 *In vivo* studies of glycosyl transfer and late stage modifications during biosynthesis of nogalamycin

The establishment of the pSnogaori/pIJTZOMLT complementation system provided the possibility to study the late stage glycosylation and modification steps of nogalamycin biosynthesis *in vivo*, since all genes annotated as required for biosynthesis of aglycone and deoxysugar were included. This was indeed the case as the production of nogalamycin (**1**) in the heterologous host *Streptomyces albus* was observed (compounds presented in Fig. 3.1). The pSnogaori alone gave rise to the novel compounds 3´,4´-demethoxynogalose-1-hydroxynogalamycinone **(3)**, Nogalamycin F **(4)** and Nogalamycin R **(5)**, with SnogD responsible for rhodosamine and 2-deoxyfucose transfer (Fig. 3.1). The individual knock-outs of the GT genes *snogE* and *snogD* from the pSnogaori/pIJTZOMLT system produced the compounds **2** and **3** respectively. Hence SnogD is responsible for transfer of the nogalamine moiety and SnogE for the nogalose moiety (most likely in the forms of TDP- L -acosamine (**8**) and TDP-2,3,4-tridemethoxy nogalose (**7**), respectively). In addition to *snogD* and *snogE*, the gene cluster contains a third predicted GT gene, *snogZ*. The *snogZ* gene is however not required for either of the two *O*-glycosyl transfers as the compounds **3**, **4** and **5**

were produced in the absence of *snogZ*, using the pSnogaori vector. Furthermore formation of the C-C bond of **5** rules out the need of snogZ for the *C*-glycosyl linkage. Based on the *in vivo* data the *snogZ* would appear redundant.

Following transfer of the nogalose moiety by SnogE, the *O*-methylations of the C3´and C5´positions are likely catalysed by SnogM and SnogL. Hydroxylation of the C1 position of **3** is catalysed by SnoaW/SnoaL2, as the step preceding nogalamine transfer by SnogD (Fig. 2.2). Dimethylation of the C3´´ amino group of the nogalamine moiety by SnogA and SnogX occurs after carbohydrate transfer to the aglycone.

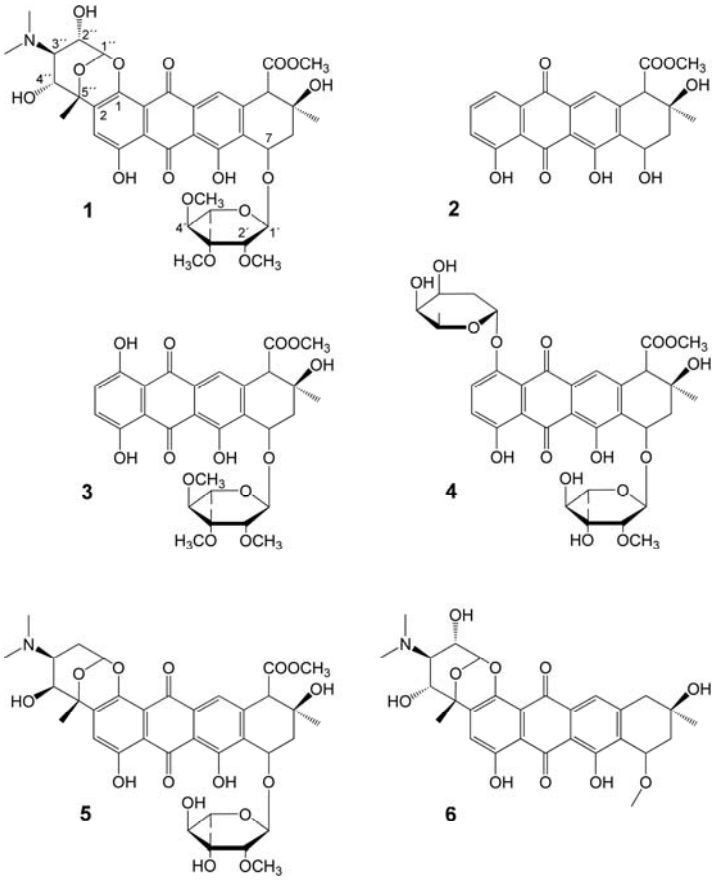

Figure 3.1 – Structures of the anthracycline compounds included in papers I and III. **1**, nogalamycin; **2**, nogalamycinone; **3**, 3´,4´-demethoxynogalose-1-hydroxynogalamycinone; **4**, nogalamycin F; **5**, nogalamycin R; **6**, menogaril. The compound enumeration used here is in accordance with paper III, with addition of compound **6**.

3.1.2 Recombinant protein production

To enable *in vitro* experiments, s*nogD* was cloned from genomic *Streptomyces nogalater* DNA into pET-based vectors (Fig. 3.2A), followed by solubility screening to optimise the production of soluble recombinant protein. Proteins resulting from these constructs were purified in quantities exceeding 2 mg/l *E. coli* culture, but each sample suffered from precipitation, indicating poor stability.

Therefore a multi-construct approach was established, similar to that developed by the Structural Genomics Consortium (SGC) [87]. This included ligation independent cloning (LIC) [88] and extensive solubility screening, which together were required to produce sufficient amounts of SnogD for crystallisation trials and activity experiments (Fig. 3.2B).

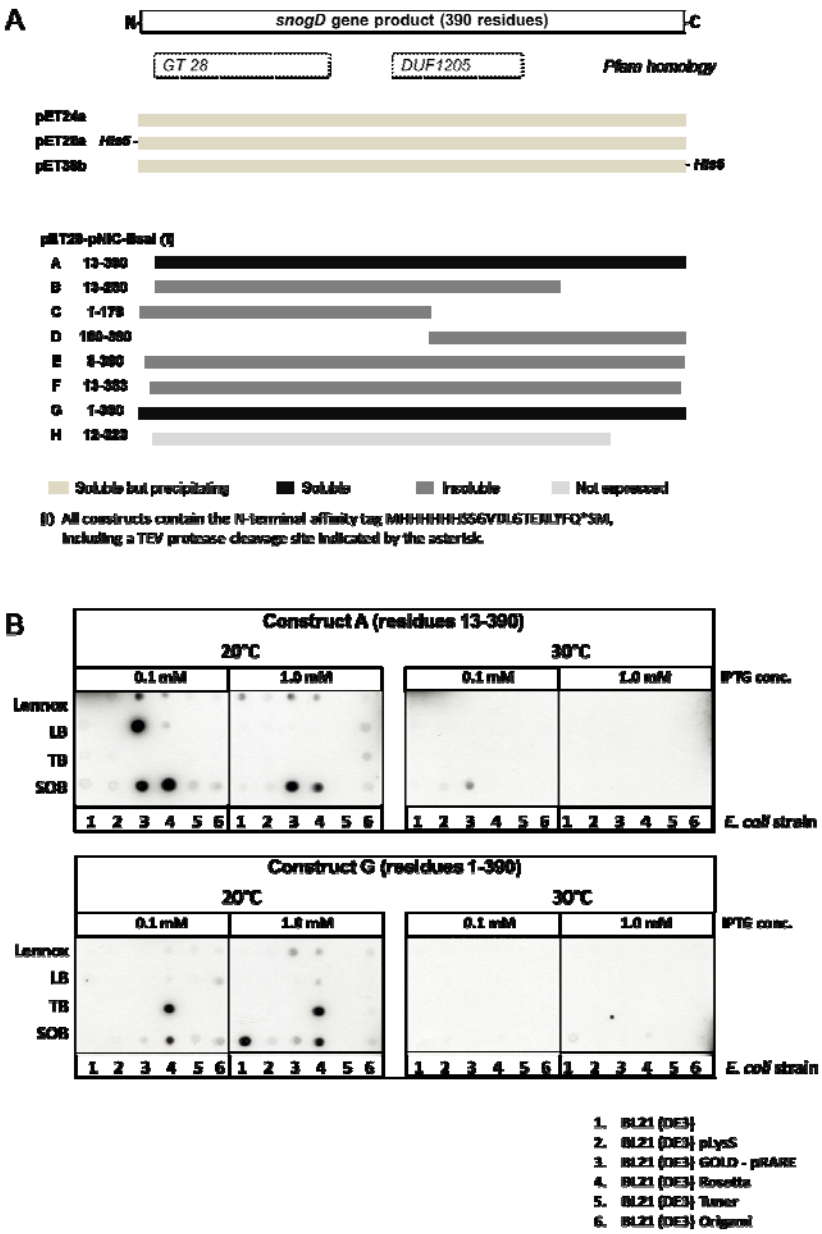

Figure 3.2 – A) Cloned constructs of *snogD*. B) Dot-blot detection of soluble recombinant SnogD from expression screening with the constructs A and G.

The SGC pipeline, optimised for cloning of human genes, had to be adapted to facilitate cloning of the high GC-content DNA of *Streptomyces* (the GC content of the genes investigated here is 73/73/75 % for the genes *snogD/snogE/snogZ* respectively). This was achieved by extended denaturation of template at high temperature (typically 5 minutes at 371 K) and extensive use of dimethyl sulfoxide (DMSO) and glycerol (concentrations up to 14% and 10% respectively) during polymerase chain reaction (PCR), to decrease the strand and primer separation temperatures. Of the DNA polymerases tested only a subset (Phusion, Finnzymes and *pfu* polymerase, Stratagene) successfully amplify the genes when combined with DMSO. At the recombination step during LIC, the insert to vector ratio was typically increased to 6:1 to produce transformants.

Two of the cloned constructs resulted in microgram amounts of soluble protein detected by dot-blot [89] (Fig. 3.2B). Screening for an optimal expression condition and optimization during scale up, by use of cold-shock prior to induction was

performed. This resulted in one condition producing soluble protein of the construct "A", encoding residues 13-390. The recombinant SnogD protein could be purified to homogeneity by three steps of liquid chromatography in amounts of 1 mg / litre culture, and was used for crystallisation and enzymatic experiments. Addition of trace metal ions was later found to enhance the soluble yield during production of SnogD [90]. The precipitation problem associated with the initial constructs was overcome using the multi-construct approach, however long term protein stability was still a limiting factor. Studies of SnogD were possible by rapid and frequent protein purification, directly followed by experiments.

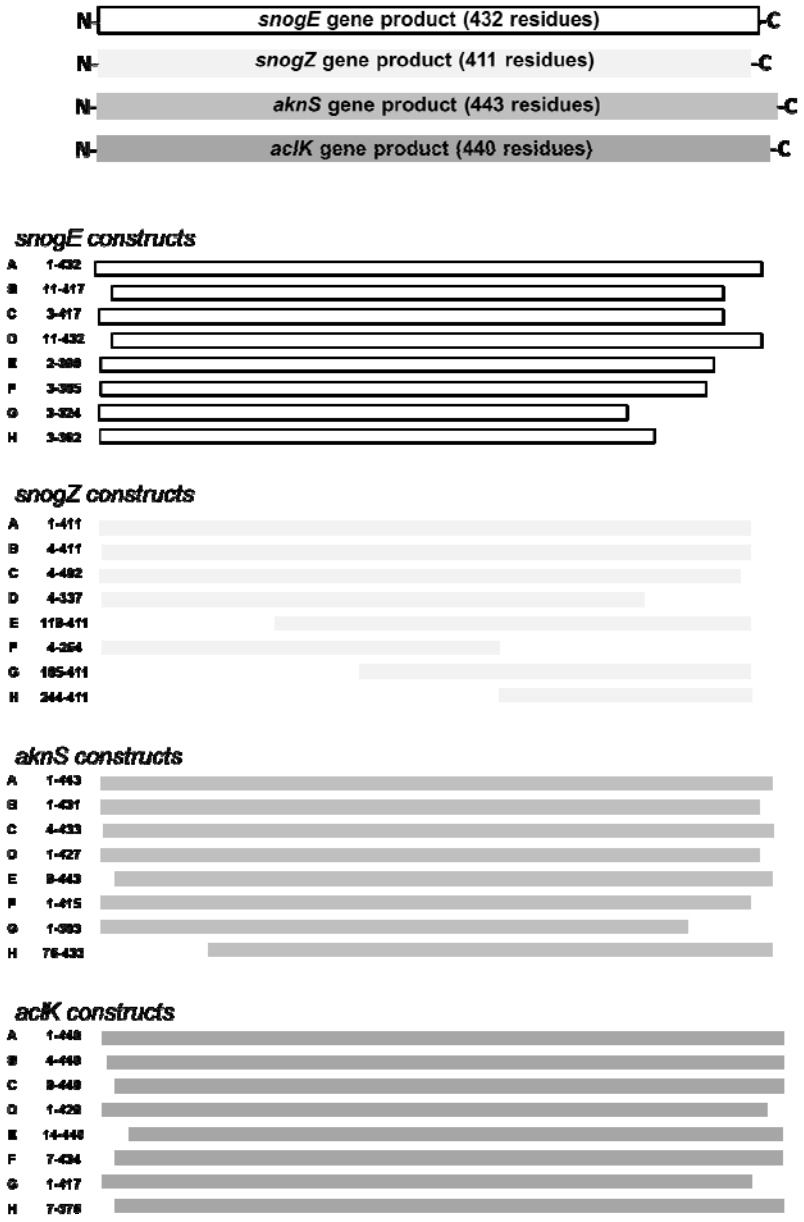

Figure 3.3 – Graphical representation of the cloned constructs of *snogE, snogZ, aclK and aknS*. All constructs were cloned into the pET28-pNIC-BsaI vector.

The genes encoding *snogE* and *snogZ* were also LIC cloned into the pET28-pNIC-BsaI vector, following the procedure described for *SnogD*, resulting in insufficient protein yields (<0.1 mg/l culture), also suffering from precipitation. The multi-construct approach was applied to *snogE*/*snogZ* and their homologs *aknS*/*aclK* from the

aclacinomycin pathway of *Streptomyces galilaeus* (58.4 and 30.3 % sequence identity to SnogE/SnogZ respectively)(Fig. 3.3). However of the constructs designed, no clone producing soluble recombinant protein above microgram levels was obtained.
With the majority of recombinant SnogD found in inclusion bodies, over-expression of *E. coli* chaperones (dnaK-dnaJ-grpE, groES-groEL, tig) was performed in an attempt to increase the soluble yield of the SnogD constructs A and G, with no improvement observed even at 293 K. The high GC-contents of these genes and poor recombinant protein solubility/stability limited studies of the GT enzymes from the nogalamycin biosynthesis. Gene synthesis with codon adaptation for the expression host, alternatively use of an expression host with inherently high GC DNA, and design of additional truncation-constructs could provide a solution for future studies.

3.1.3 Studies of SnogD catalysed glycosyl transfer

In the absence of known and available natural substrates for SnogD at the time, and the complexity of obtaining such, an enzymatic assay was set up which was inspired by the transglycosylation experiments of Thorson and colleagues [91]. In this system the activity of SnogD could be studied in the "reverse" direction of natural biosynthesis, i.e. transfer of the carbohydrate from the aglycone to a dinucleotide, thus providing an alternative path for activity studies. This was particularly appealing at a time when the predicted donor-substrate TDP- L -acosamine was not available, and the proteins predicted to convert TDP-5´-glucose into the required carbohydrate (SnogK/ SnogH/ SnogI/ SnogF/ SnogG, Fig. 1.4A&C) were not characterised [22]. The described two-step GT catalysed transfer of a carbohydrate from one aglycone to another, via a nucleotide-5´-diphosphate (NDP), exploits the relaxed substrate specificity reported for several GT enzymes and would allow generation of NDP-activated carbohydrates [92–95] (Fig. 3.4).

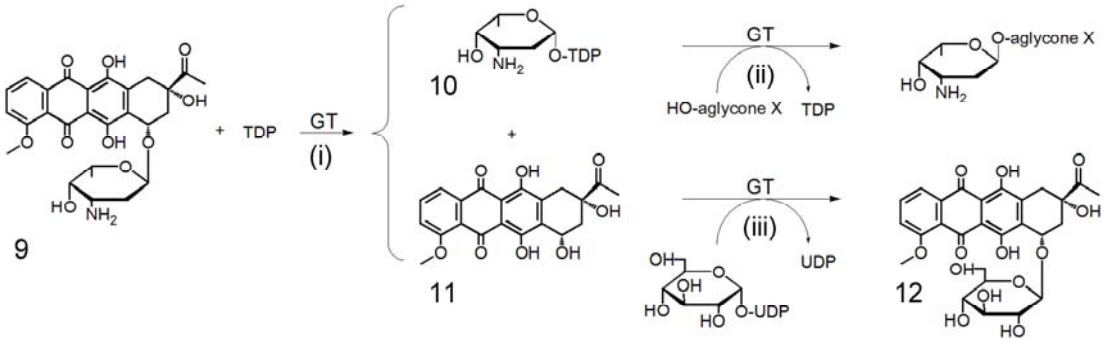

Figure 3.4 – Schematic representation of transglycosylation reactions to study glycosyltransferase (GT) catalysed reactions. (i) Glycosyl transfer from 13-deoxdaunorubicin (9) to TDP, producing the activated TDP-L-daunosamine (10) and the aglycone (11). Both products can be used as substrates for subsequent glycosyl transfer reactions. (ii) Glycosylation of a different aglycone (X), with the carbohydrate **10** derived from **9**. (iii) Glycosylation of **11** using a different donor sugar, exemplified by UDP-5´-glucose, resulting in the not naturally occurring compound **12**.

Cultivation of *Streptomyces lividans* supplied with the majority of the *sno* gene cluster yielded an extract of nogalamycin-type compounds. Activity of SnogD could be observed through changes in the relative amounts of these compounds upon addition of the enzyme, but only in the presence of UDP in molar excess over the anthracycline substrates (Fig. 3.5A).

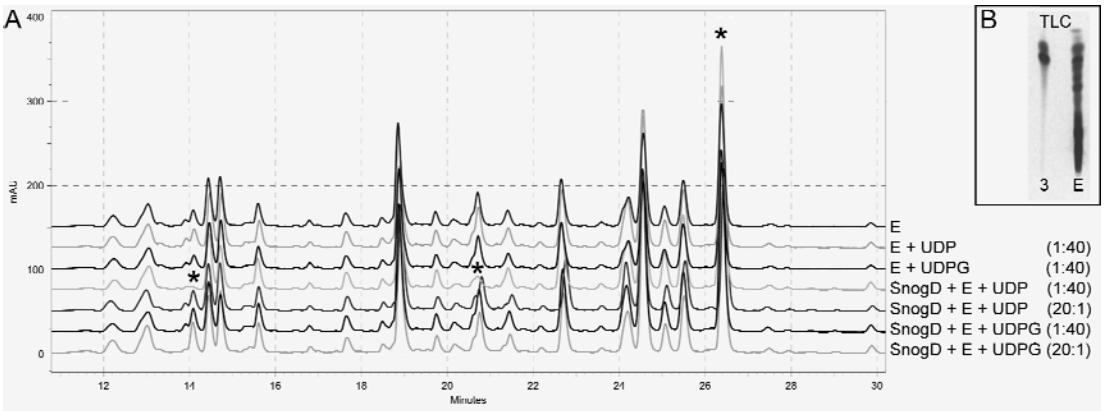

Figure 3.5 A) HPLC chromatograms of SnogD reactions with the extract (E) and UDP or UDP-glucose (UDPG), the molar ratios of extract to UDP/G are presented in parenthesis by each trace. The peaks with clearly altered intensity for the "SnogD+E+UDP (1:40)" reaction are indicated by asterisks (reduced at; 14.1, 20.7 and increased at 26.4 min respectively). B) TCL of partially purified compound **3** and the extract.

The discovery of the compounds **2**, **3**, **4** and **5** enabled more detailed enzymatic activity studies of SnogD. The *O*-glycosyl transfer activity at the C1-hydroxyl of the aglycone, observed *in vivo*, was verified using recombinant SnogD by the deglycosylation of **4** resulting in **3** (Fig. 3.6). The reaction is dependent on a pyrimidine type dinucotide but not selective for TDP, the nucleotide used during biosynthesis, since presence of UDP also resulted in deglycosylation to a comparable extent. Glycosyl transfer from TDP-5´-glucose by SnogD onto **3** did not occur *in vitro*, suggesting limitations to 2-deoxy carbohydrates such as rhodosamine and 2-deoxyfucose. The *in vivo* production of both **4** and **5** would imply a specificity for 2,6-dideoxy forms of NDP-activated carbohydrates (rhodosamine and 2-deoxyfucose), but the stereochemistry of the C3-hydroxyl of the hexose appears less stringent as this differs from **1** in both compounds. Furthermore only rhodosamine was incorporated in the bicyclic configuration typical for nogalamycin, perhaps indicating the C3´´-NH_2 moiety is required for formation of the C-C bond or substrate-binding.

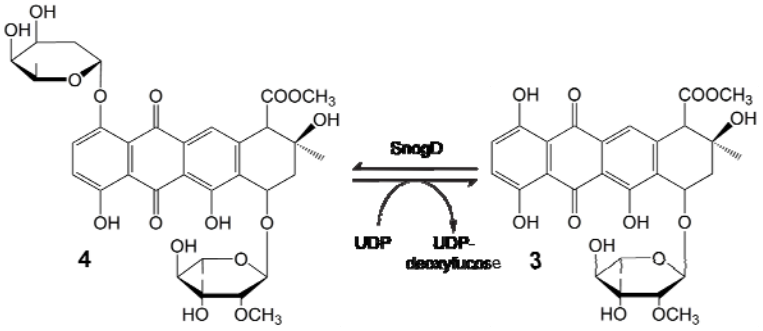

Figure 3.6 –The TDP/UDP dependent reaction catalysed by SnogD.

Incubation of SnogD with the C7-glycosylated compound **3** did not result in SnogD catalysed deglycosylation in the presence of TDP/UDP, indicating that the C7-carbohydrate is required for acceptor-substrate recognition and binding. Nor did the incubation of SnogD with the daunosamine containing 13-deoxydaunorubicin (**9**) result in glycosyl transfer onto TDP/UDP. Carbohydrate transfer from **9** to TDP/UDP would have generated the nucleotide activated TDP-L-daunosamine (**10**), which only differs from the postulated donor substrate of SnogD (**8**) at the stereochemistry of the C4´ hydroxyl group, thus could have provided a potential donor-substrate (Fig.

1.4&3.4). SnogD could not remove the carbohydrate of **5**, suggesting that these reactions either require an additional partner/activation, or perhaps that only the *O*-glycosidic bond was cleaved. Taking these results together, the relaxed substrate specificity of SnogD would enable generation of new anthracycline compounds, limited to 2-deoxy carbohydrates with a requirement for an attached C7-carbohydrate.

Inhibition experiments with topoisomerase I and II and the novel nogalamycin-type compounds **3**, **4** and **5** visualised the respective roles of the attached carbohydrates in comparison to **1** and **6**. The compounds **3, 4**, and **5** did inhibit human topoisomerase I [96], the target of nogalamycin inhibition. Topoisomerase II was inhibited only by **6** and **1**, implying the importance of the C2-C5´´ C-C bond and the stereo-chemistry of the C6´´ methyl group for an optimal interaction with DNA. The C2´´-hydroxyl group appears important for DNA-anthracycline complex stabilization by hydrogen bonding to major groove purine bases [16], as the inhibition effect was significantly reduced for **5**.

3.1.4 Crystallisation of SnogD and SnogDm

Recombinant SnogD of the construct A was crystallized in the space group $P2_12_12$, in complex with the donor substrate homologue 2-deoxyuridine-5´-diphosphate (dUDP). Due to difficulties in reproduction of diffraction quality crystals, reductive methylation (RM) was required to overcome the reproduction hurdle and allow additional data collection. Fisher *et al.* reported the presence of a methylated pyridoxal phosphate in glycogen phosphorylase in 1958 [97] and a procedure using formaldehyde as methyl group donor was described ten years later [98]. RM has been utilized to enhance the crystallisation propensities of proteins and for salvaging soluble proteins recalcitrant of crystallisation [99]. The methyl group donor formaldehyde forms a Schiff-base adduct with solvent exposed free amines of the protein, i.e. the N-terminal amine and the ε-amine of lysines. Reduction of the Schiff base by a strong reductant generates the final methyl-adduct, which can subsequently undergo a second step resulting in the tertiary amine (Fig 3.7).

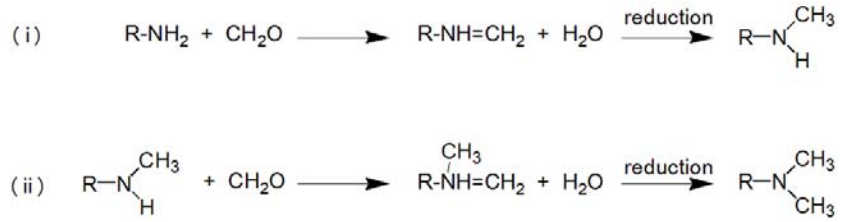

Figure 3.7 –Reaction scheme for reductive methylation of solvent exposed amines. The generation of the secondary and tertiary amines is shown in i) and ii) respectively.

RM has been indicated to alter isoelectric point, solubility and hydropathy, which may promote crystallisation by facilitating crystal packing [100]. The biochemical activities of methylated proteins have in several cases been reported as unchanged post methylation when compared with wild type enzyme, with small or no changes in three dimensional structure [100]. Methylation increases the lysine interaction radius by 1-1.2 Å, replacing long range (4.2Å) ε-amine interaction, with shorter (>3.3 Å) and stronger interactions to their respective oxygen/nitrogen partner [100]. Interactions of methylated lysines are reported to include carboxyl- and main chain carbonyl groups as well as side chains of arginine and histidine residues. Stronger interactions of the ε-

amine are associated with a reduction in local entropy, which would be beneficial for crystallisation.

RM was performed on partially purified SnogD based on a generic protocol [99], with the formaldehyde solution prepared by depolymerisation of inexpensive solid paraformaldehyde immediately prior to use. RM of SnogD resulted in a mass increase corresponding to complete di-methylation of the N-terminal nitrogen and all but one of the four lysine residues. A substantial loss of material was observed during methylation of SnogD (typically exceeding 50%), however the propensity of the protein to aggregate appeared reduced (Fig. 4.8). The reason behind the observed reduced aggregation is not clear, but is possibly a result of precipitation of unstable protein during the harsh chemical treatment.
Precipitating with SnogD was a commonly co-purifying contaminant from the expression host, DnaK, which could be removed with a fraction of recombinant SnogD during purification. With knowledge of the chaperone contaminant, ATP, high concentrations of NaCl/urea (<0.5 and <1 M), glycerol and mild detergents as β-octyl glycoside (BOG) were added during purification of SnogD. However no observed turn over and release of SnogD was observed.

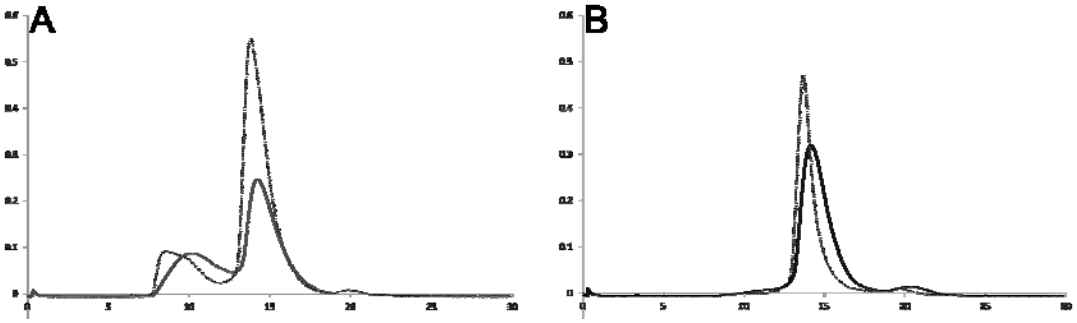

Figure 3.8 – Size exclusion chromatograms of purified recombinant SnogD compared with the methylated protein displaying the reduction in aggregation post methylation. A280 traces of A) Native SnogD (dotted line), sample after 48 h (full line). B) Methylated SnogD (dotted line), sample after 48 h (full line). The slight deviations in peak elution volume and peak width are associated with different sample volumes.

Methylated SnogD was crystallised reproducibly by streak-seeding from second or later generations of SnogDm microcrystals, in the space groups $P2_12_12$ and P2. The optimised crystallisation condition was associated with multiple nucleation events, often resulting in showers of microcrystals within one hour even at 277K. Reduction in precipitant concentration to reduce nucleation rate, resulted in protein precipitation and no increase in crystal size. Screening for additives promoting crystal growth and packing, using commercially available screens, did not improve crystal growth. The additive screening did not provide crystallisation "poison", able to reduce the nucleation rate [101], nor did addition of detergents such as BOG or organic solvents. Even pre-mixing SnogD with precipitant and filtration, to remove initial nucleation sources, did not reduce the nucleation. Hence streak seeding at drop setup, using methylated SnogD, was the only way well diffracting crystals could be reproduced.

Co-crystallisation experiments with SnogD were performed to obtain a ternary complex. Dinucleotides (TDP, TDPG, UDP, CDP and GDP) alone and in combination with polyketides (**1**, **3**, **4**, nogalonic acid methyl ester, 1,5 and 1,8-dihydroxyanthraquinone) were screened using both wild type and methylated SnogD. The polyketides

do however require solubilisation in organic solvents, which often resulted in a final solvent concentration too great for the protein, in order to achieve the desired 5-10 fold molar excess of polyketide. The solvent tolerance of SnogD was determined using a simplistic screening method, where solvent was added to the concentrated protein until signs of precipitation were observed, by native gel and under microscope. Stepping back from the critical solvent concentration, and optimizing the mixing order of the solutions, enabled co-crystallisation experiments to be performed with ligand concentrations in molar excess and without detectable aggregation of the enzyme. Reduction in protein concentration and addition of ligand close to the solubility limit with subsequent co-concentration was also utilised. The optimal mixing order for SnogD was determined to be addition of ligand to buffer and solvent, followed by protein and rapid mixing, typically resulting in a solution suitable for use, with no precipitation or minor amounts of brightly coloured polyketide ligand present as micro-crystals.

Protein crystals were obtained following co-crystallisation of SnogDm with **1**, **3**, nogalonic acid methyl ester and 1,5-dihydroxyanthraquinone, in combination with UDP. In the presence of **1** and **3** these crystals accumulated a purple/red colour, indicating an accumulation of the respective polyketide within the crystal, however the X-ray diffraction of these crystals did not extend further than 4 Å with smeary spots and signs of anisotropy. Soaking experiments with ligands were performed and although these did not cause visible changes in the crystal morphology, protein diffraction beyond 20 Å was never observed.

3.1.5 Structure determination of SnogD

The structure of wild type SnogD (PDB ID: 4amb) was determined by molecular replacement and refined to a resolution of 2.6 Å with 2-deoxyuridine-5´-diphosphate (dUDP) bound (data collection and refinement statistics in paper III). Structures of methylated SnogD with and without dUDP were determined by molecular replacement and refined to 2.7 and 2.6 Å respectively (PDB ID: 4an4, 4amg), using the structure of SnogD-wtdUDP as search model.

The overall structure of SnogD belongs to the GT-B fold and shares the canonical twin domain Rossmann-fold [102] of this fold class [43], with the active site located at the subunit interface (Fig. 3.9). The quaternary structure of SnogD was determined to be dimeric in solution, correlating well with the content of the asymmetric unit of the $P2_12_12$ crystals. The molecules of the biological dimer are oriented head to tail, and are related by a twofold non-crystallographic symmetry axis. The tetramer assembly observed in P2 consists of a dimer of dimers, induced by a variation in crystal packing interactions.

The N-terminal domain (residues 1–209) consists of a seven-stranded parallel β-sheet, flanked by eight α-helices and two 3_{10} helices, distributed three respectively four per side. The N-terminal 7-stranded parallel β-sheet is extended by an additional parallel β-strand formed by residues (215-217) from the interdomain linker (residues 210-227). The two domains are connected by a 17 residue well defined interdomain linker, contributing to the 1400 $Å^2$ large dimer interface. The C-terminal domain (residues 228–390) is of similar topology, with a six-stranded parallel β-sheet flanked by six α-helices and four 3_{10}-helices. The last C-terminal helix crosses over to complete the N-terminal domain, through residues Pro378-Gly390, and includes a kink between residues Glu374 and Pro377, a common feature of the GT-B fold.

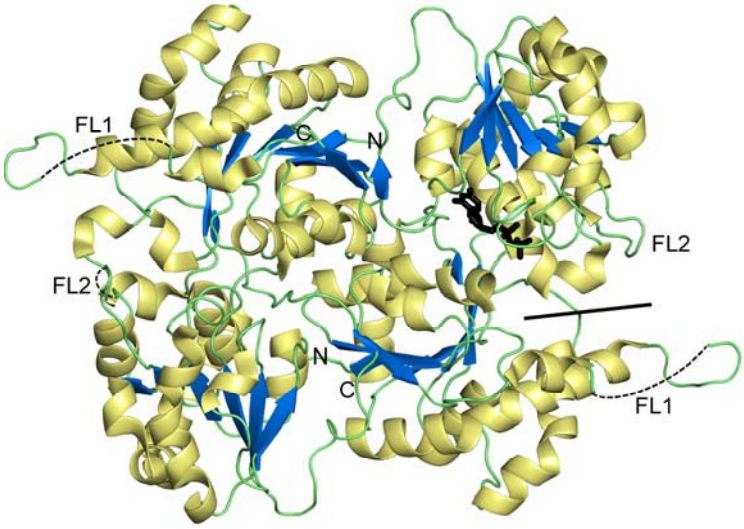

Figure 3.9 – The 3D structure of the SnogD dimer, coloured by secondary structure. The A chain is presented on the left and on the right side the B chain with bound dUDP shown as a black stick model. The flexible loops involved in substrate binding, FL1 and FL2, are illustrated as dashed lines. The putative location of acceptor substrate binding is indicated by the black bar to the right.

3.1.6 Nucleotide binding and the active site

The donor substrate mimic dUDP co-purified from the expression host, and was present in one chain per dimer (Fig. 3.9). Nucleotide binding is associated with rearrangement of two loops; n1 comprising part of the α-phosphate binding site and n2 which is shifted out to accommodate the pyrimidine ring of the bound nucleotide (Fig. 5 paper III). The more outward conformation of n1 and the condensed conformation of n2 enabled optimal crystal packing interactions with a symmetry related molecule in the absence of nucleotide. The nucleotide containing subunit is not part of such crystal packing interaction, and the additional packing interaction in the absence of nucleotide could explain the selection for and incorporation of nucleotide free and half-occupied dimers in the crystals. dUDP binds in the domain cleft similarly to nucleotide binding observed in other class 1 GT structures (Fig. 3.10).

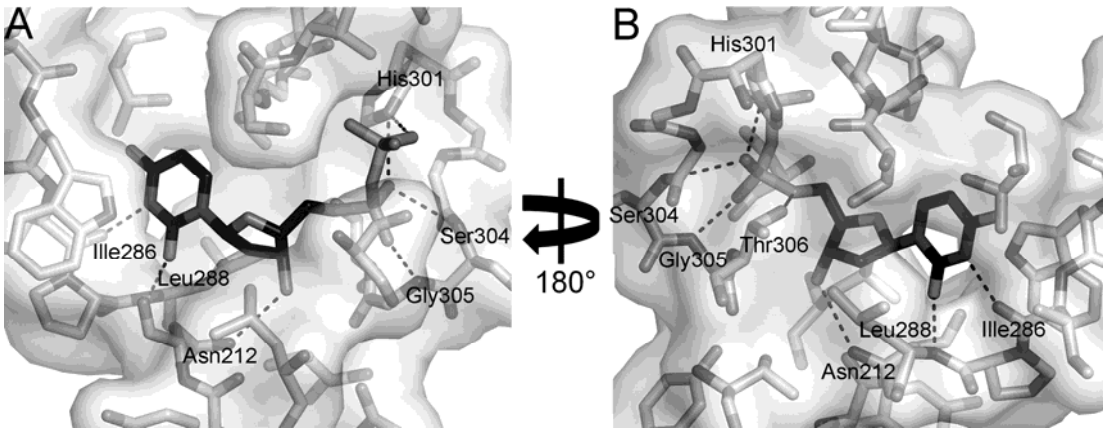

Figure 3.10 – The nucleotide binding in SnogD shown in two orientations. A) dUDP is shown as black sticks. Residues within 4 Å distance of dUDP are shown as light blue sticks, and their surface is shown semi-transparent. Hydrogen bonds (<3.2 Å) are indicated by dashed lines. B) 180° rotated view of A

Hydrogen bonds to the ligand dUDP are provided by the C-terminal domain plus one residue of the interdomain linker. The conserved PPi-binding motif is in SnogD represented by GGSG (Fig. S1 paper III), and forms hydrogen bonds to the α-phosphate of dUDP. The highly conserved His301 is suitably located to provide

hydrogen bonds to the O1 oxygens of the α- and β-phosphates. The presence of a positive charge close to the leaving group would help stabilising the developing negative charge of the di-phosphate during glycosyl transfer, a function commonly fulfilled by a helix dipole or imidazole group [43].
Leu288 is located in the vicinity of the expected position of the donor substrate 2´-hydroxyl group (Fig 3.10), but is unlikely to enforce 3´deoxy-ribose dinucleotide preference as was suggested for the glycosyltransferase GtfA [55], correlating with the observations from the *in vitro* activity experiments. This, combined with an unoccupied volume by the C2 of the uracil ring large enough to accommodate the additional methyl group of TDP, would explain the lack of dinucleotide type selectivity observed during enzymatic experiments with SnogD.

Hydrogen bonding to the deoxyribose 3´-hydroxyl of dUDP is provided by Asn212 of the interdomain linker, and not by protein co-ordinated water or the side chain of a glutamic acid residue following the PPi-motif as seen in UDP discriminating GT [55], [103], [104]. In SnogD the corresponding residue is Thr309, which is not interacting with the deoxyribose group. The hydrogen bond by Asn212 to the ribose moiety contributes to the alternative position of the interdomain linker observed in SnogD as well as SpnG and SsfS6 (Fig. 3.11).

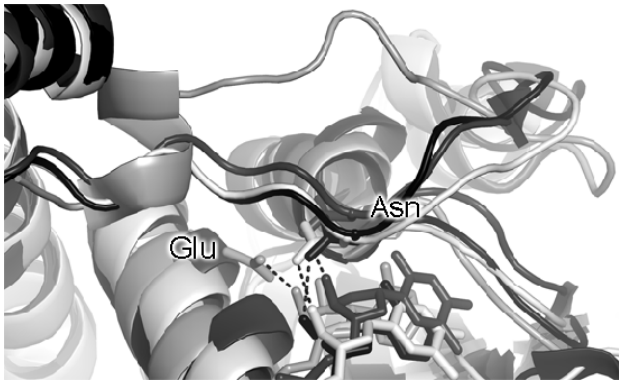

Figure 3.11– Interdomain linker organization upon hydrogen bonding of Asn212 to the C3´ of the deoxyribose hydroxyl of dUDP, SnogD (black, PDB ID: 4AMB), SpnG (dark grey, PDB ID: 3UYL [105]), SsfS6 (white, PDB ID: 4G2T, [106]) and UGT78G1 (light grey, PDB ID: 3HBF, [107]). For clarity the respective N-terminal domains are not shown. The Asn-residues and the Glu360 of UGT78G1 are shown as sticks, coloured as the protein.

The two flexible loops associated with substrate binding (FL1 and FL2), located at the domain-interface, could not be completely built in all structures due to weak or missing electron density. Binding of substrates would likely involve both loops, with FL1 folding over the acceptor substrate and FL2 interacting with the carbohydrate moiety of the donor-substrate similar to the previous observation in homologous GTs such as CalG3 [108]. The commonly observed D/E motif associated with hydrogen bonding to the C2´´-C4´´ hydroxyl groups of the donor carbohydrate is not observed in SnogD. However the polar residues of the FL2 loop may form hydrogen bonds to the 3´´ amino and 4´´-hydroxyl groups.
The residues flanking FL1 contribute hydrophobic residues to a shared dimer-dimer hydrophobic interaction, formed with residues of the crystallographically related molecule (Fig. 3.12). The resulting hydrophobic cluster forces the FL1 loop into an adjacent solvent pocket. Ligand binding by SnogD would fold FL1 over the substrate, and likely result in a protrusion at the dimer-dimer interface, as seen in CalG3.

Formation of such protrusion and disruption of the hydrophobic cluster would disturb the observed crystal packing or even prevent it, perhaps explaining the difficulties in obtaining well diffracting SnogD crystals in complex with an acceptor-ligand.

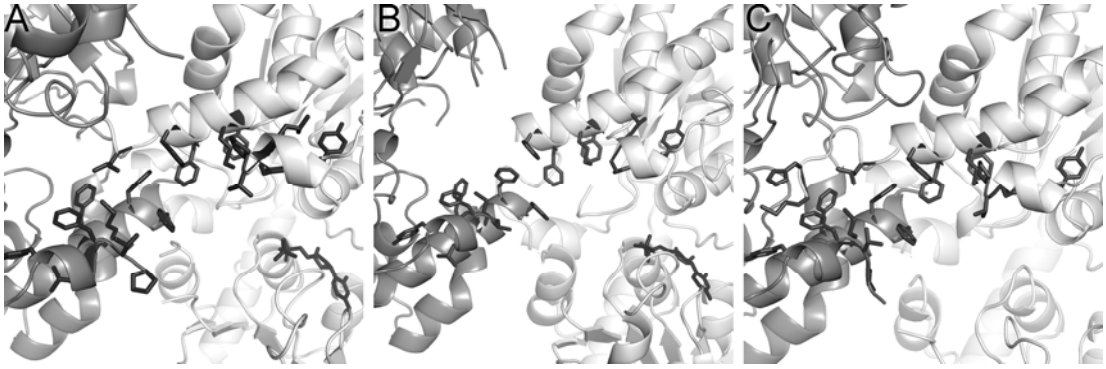

Figure 3.12 – The hydrophobic cluster formed at the dimer-dimer interface. A) SnogDwt. B) SnogDm(dUDP). C) SnogDm. The two interacting subunits are shown in a cartoon representation, coloured dark and light grey respectively. The dimer-dimer interface is located vertically at the centre in each representation. The hydrophobic residues are shown as dark grey sticks. The bound dUDP is shown as grey sticks.

The formation of a shared intra-molecular 4-stranded β-sheet in the P2 crystal form resulted in an alternative orientation of the crystal contact present in $P2_12_12$. The residues forming the contact are present as either α-helical/loop or antiparallel β-sheet, giving rise to two distinct modes of dimer-dimer interactions related by a 180° rotation along the a-axis. Reductive methylation of SnogD resulted in an additional intramolecular salt bridge, arising within the crystal lattice between the methylated Lys384 and Gly374* of an adjacent molecule. The dimer-dimer interface along the a-axis is thus slightly different in crystals of the methylated protein.

3.1.7 Active site mutagenesis

Based on the dUDP complex structure of SnogD and active site residue conservation, four active site mutants were selected (His25, Asp128, Asp238 and His301). The high and varying GC-content of *snogD* made mutagenesis challenging, primarily the acquisition of PCR products, which required a step-down protocol, long primers and addition of DMSO to generate any product. The mutations were introduced by PCR, with mutations in the 5´-end of primers, hence the resulting gene fragments were only overlapping by three base pairs. The dsDNA sequence of SnogD was subsequently produced from the fragments and cloned into the pET28-pNIC-BsaI vector using restriction enzyme digestion and ligation. The mutants His25Ala, His25Asn and His301Ala were successfully cloned, and purified following the procedure developed for the construct A. Circular-dichroism of the mutants was performed to verify foldness.

Table3.1 – Relative *in vitro* activity of SnogD mutants.

	Relative activity of triplicates (%)	Standard deviation (%)
No enzyme	1.1	1.7
No UDP	0.2	0.1
His25Ala	1.5	0.6
His25Ans	1.9	1.3
His301Ala	2.4	1.6
Wild type	100	5.8

Activity of the mutants was investigated *in vitro* using the established deglycosylation assay, resulting in a significant loss of activity observed for all mutants compared with wild type enzyme (Fig. 7 paper III, and Table 3.1). The mutant activities *in vivo* were also investigated, using the SnogD knock-out and the generated mutants. The activity *in vivo* showed the same trend, by a reduction in the production of **5** (Fig. 8 paper III).

3.1.8 Reaction chemistry of SnogD

The binary complex of SnogD with **3** and TDP-nogalamine was modelled, based on the SnogDwt structure and the complex of *Vitis vinifera* 3-*O*-glucosyltransferase (VvGT1) with the donor substrate analog uridine-5´-diphosphate-2-deoxy-2-fluoro-α-D-glucose and the acceptor kaempferol [59] (Fig. 3.13). The location of the donor-substrate carbohydrate was modelled in the active site, imposing (i) the restrictions from the covalent bond to the dinucleotide, (ii) the small carbohydrate binding pocket closed by FL2 and (iii) an axial orientation of the C1´-hydroxyl group required for carbohydrate transfer.

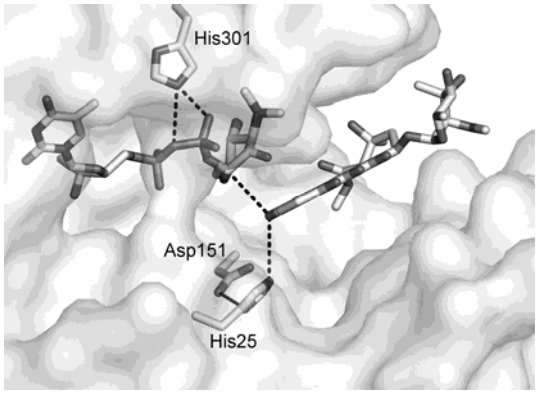

Figure 3.13 – Model of the Michaelis complex of SnogD. The acceptor substrate **3** and TDP-nogalamine are shown as sticks in the active site. Asp151 and the two catalytic histidines are also shown as sticks, with putative hydrogen bonds indicated with dashed lines.

For catalysis by inversion of the anomeric carbon, which SnogD most likely proceeds by, the C1- hydrogen bond and the *O*-glycosyl bond to the α-phosphate will be in a strained conformation. In the large hydrophobic groove of SnogD the planar aglycone was positioned, and restricted by the proximity requirement for a correct distance to the anomeric carbon of the carbohydrate and the position of the catalytic base His25. Hence the C7-carbohydrate of **3** was modelled towards solvent, without clashes. In the model the position of the conserved His25 is in proximity of the C1-hydroxyl position and adjacent to a conserved Asp151, which is suitably located to coordinate the histidine side chain and aid proton abstraction by hydrogen bonding to Nε2 of the histidine.

The glycosyl transfer reaction catalysed by SnogD (Fig. 1.2) is likely to resemble that described for the macrolide GT enzyme OleI from *Streptomyces antibioticus* [103] (Fig. 3.14). The conserved His25 would be the catalytic base, activating the nucleophile by abstracting the proton of the C1-hydroxyl group, which subsequently performs an attack on the anomeric C1 of the carbohydrate. During formation of the carbohydrate-aglycone bond, His301 supplies a positive charge to compensate for the developing negative charge of the diphosphate leaving group as postulated previously [104]. These roles of His25 and His301 are supported by the reduction in enzymatic activity

of the respective mutants, with a nominally larger effect for His25. This observation is supported by the conservation of these two histidine residues and biochemical studies that suggested similar functions of the corresponding residues in the catalytic mechanism of other inverting GT enzymes [55], [59], [103], [104], [108], [109]. In the modelled complex a partially conserved Asp151 is suitably located to enforce a for catalysis optimal protonation state of His25.

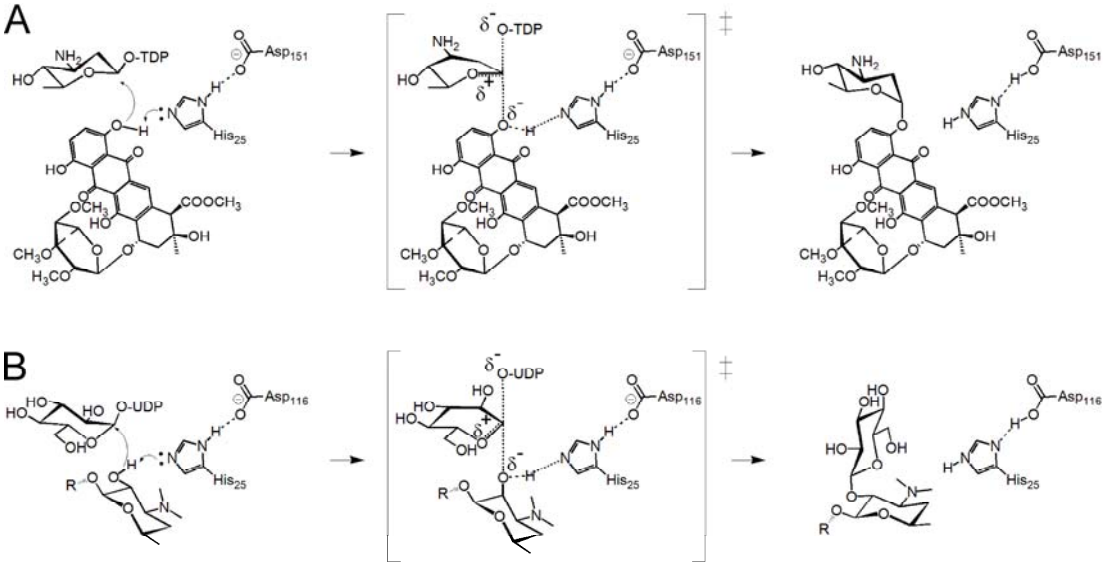

Figure 3.14 – Proposed reaction mechanisms for A) SnogD and B) OleI from *Streptomyces antibiotic*us [103], sharing with SnogD the conserved active site base His25 and an Asp residue in an equivalent position suitably located to aid the acceptor substrate deprotonation.

The substrate specificity of SnogD was observed to be relaxed with respect to the type of dinucleotide, where both TDP and UDP could be utilized for the deglycosylation reaction. This is in agreement with the structure of SnogD in complex with dUDP, in which either of the diphospho-nucleotides could be accommodated. Selection for a TDP-activated carbohydrate in SnogD cannot occur since only one hydrogen bond is provided to the ribose of the nucleotide, instead of the canonical two hydrogen bonds, typically from a glutamic acid following the PPi motif to the C2 and C3 hydroxyl groups of the ribose. In fact, the only observed hydrogen bond to the free hydroxyl groups of the ribose is provided from the interdomain linker in SnogD, a feature shared with SpnG [105] and SsfS6 [106] (Fig. 3.11). This hydrogen bond shifts the position of the interdomain linker closer to the dinucleotide, in comparison to homologous GT enzymes such as UGT78G1 [107] and CalG3 [108]. The significance of this is not obvious, but the residue conservation of the hydrogen bonding asparagine, interacting with the ribose moiety of the donor substrate, among these three GT and the presence of a short side chain residue at the typical TDP-discrimination position suggests an alternate pattern of donor-substrate selection amongst these enzymes.

3.1.9 *C*-glycosyl bond formation during secondary metabolism

Aryl-*C*-glycosidic bonds are found in bacteria and plants. Two mechanisms for carbohydrate transfer have been postulated (Fig. 3.15A) [110] proceeding i) similarly to a Friedel-Crafts aromatic substitution where appropriate positioning of the substrates affords a direct attack by the aglycone, or ii) *O*-glycosylation of the phenolic hydroxyl group analogous with *O*-glycosyltransferases followed by an *O*-*C*-bond rearrangement suggested for *ortho* -*C*-glycosidic attachments.

During the biosynthesis of the angucycline urdamycin the GT UrdGT2 transfers TDP-activated D-olivose onto the aglycone resulting in a C-C bond between the C1 carbon atom of the carbohydrate and the C9 carbon atom of the aglycone [83]. Structural analysis of UrdGT2 suggested activation of the sp^2 hybridized C9 by the deeply buried aspartic acid 137 [50] (Fig. 4.16A). Asp137 would act as a general base and via a water molecule stabilise the development of the C8-phenolate ion, thereby reduing the pK_a of the C9 and C11-hydrogens. The aryl-*C* bond would subsequently form through an S_N2 reaction coupled with the abstraction of the C9 proton by a base supplied by the enzyme, following a direct transfer mechanism (Fig. 3.15B).
The rational design of the *O*-glycosyltransferase LanGT2 into an aryl-*C*-glycosyl-transferase was recently performed based on the structure of UrdGT2 [111]. The mutagenesis work proved the importance of the Asp137 residue for catalysis, as the D137A mutant was inactive, but suggest a direct interaction of Asp137 with the donor substrate and the involvement of a serine residue side chain (Ser16 of LanGT2) for stabilisation of the increasing electron density by the C8 and C9 carbons.
The granaticin type of polyketides contain a unique bicyclic carbohydrate moiety attached by two C-C bonds [86], [112]. Carbohydrate transfer of the 4-keto-2,6-dideoxyhexose is assigned to the only annotated GT gene within this cluster *gra14*. Two routes of C-glycosylation have been proposed for granaticin i) *C*-glycosylation of the 8-hydroxylated benzoisochromanequinone aglycone followed by an intramolecular aldol condensation, or ii) intermolecular aldol condensation at C10 followed by hydroxylation at C8 and formation of an intramolecular *C*-glycoside bond (Fig. 3.15C) [112].

Neither of these types of C-C bond formation chemistry seems plausible in the case of nogalamycin, where the SnogD catalysed *O*-glycosylation at C1´´ occurs prior to the C-C bond formation [33]. The attached aminosugar moiety does not contain any keto-group, and can therefore not undergo aldol condensation. Furthermore the activated-TDP is consumed during the *O*-glycosylation and therefore no suitable leaving group remains, hence the C-C bond formation in nogalamycin would formally require a hydride removal from either i) the C5´´ carbon of the nogalamine moiety or ii) from the C2 carbon of the aglycone (Fig. 3.15D). Alternatively a radical mechanism could be proposed. In either case one would expect participation of cofactors such as NAD(P)H, flavine or an iron-sulphur cluster within the C-C bond forming enzyme of the nogalamycin biosynthesis to support such chemistry.
The structure of SnogD and the modelled Michaelis complex do not provide structural motifs capable of binding the required cofactor, nor does the protein sequence suggest presence of such motifs.

In addition no good candidates within the gene cluster have been identified as likely to assist or indeed catalyse the C-C bond formation. Involvement of proteins encoded outside the cluster to perform this reaction would seem unlikely, since this type of glycosyl moieties is quite unusual and limited to the nogalamycin biosynthesis.

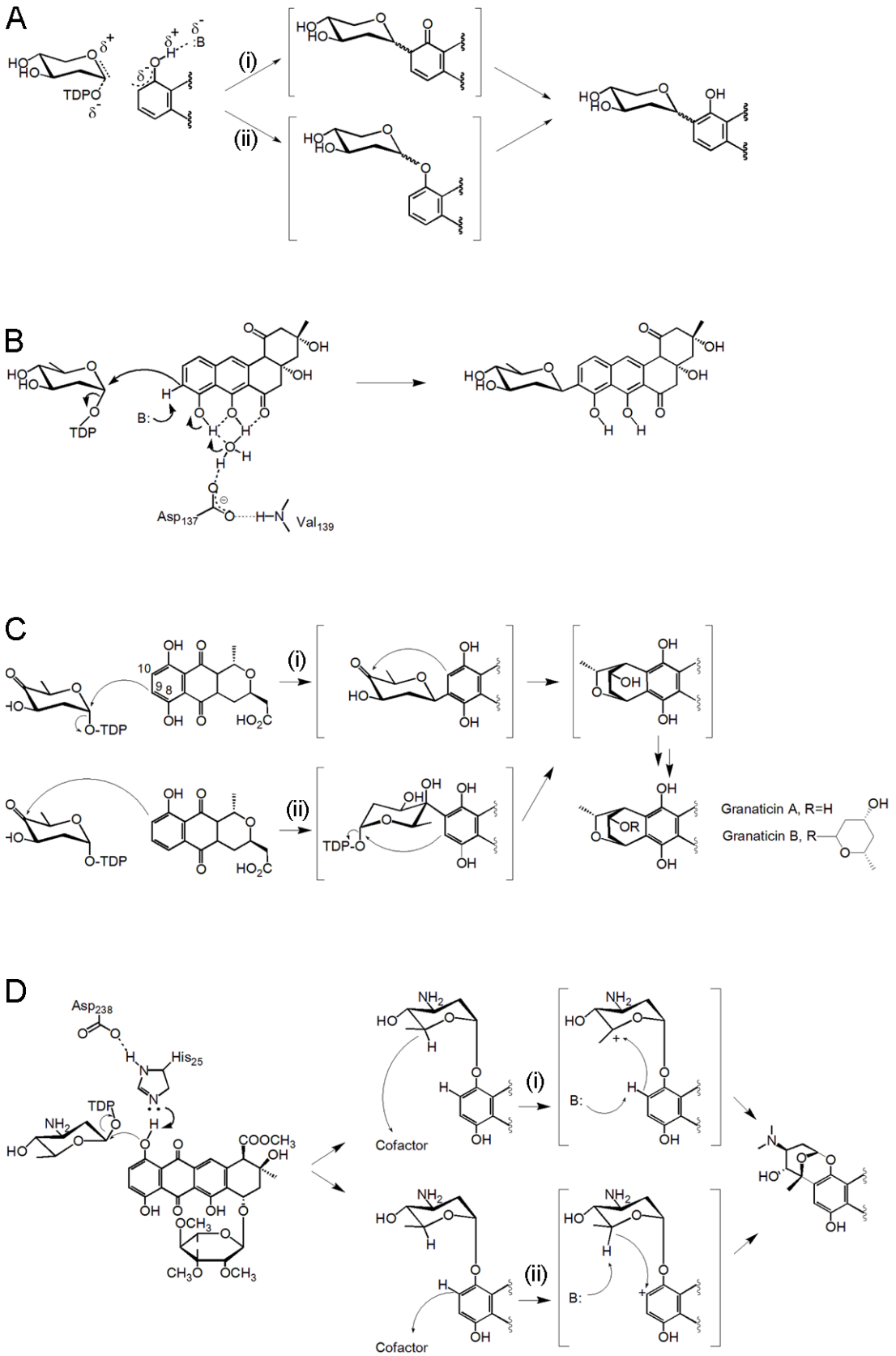

Figure 3.15 – Aryl-*C*-glycosyl and C-C bond formation. A) The two proposed mechanisms of aryl-*C*-glycosylation; i) direct addition resembling Friedel-Crafts aromatic substitution and ii) *O*-glycosylation and O-C rearrangement. B) The proposed aryl-*C*-glycosyl bond formation catalysed by UrdGT2. C) The two suggested routes for formation of the two C-C bonds present in granaticin, including an intermolecular (i) or intramolecular (ii) adol condensation. D) Initial *O*-glycosyl transfer by SnogD and illustration of the need for a hydride transfer during the C-C bond formation of nogalamycin. Enzyme supplied base is indicated by B:. Alternatively the enzyme mechanism could proceed via a radical mechanism.

For certain GT enzymes the catalytic rate is enhanced by helper proteins, with sequence homology to cytochrome p450 enzymes, but lacking signature motifs such as the invariant cysteine residue coordinating the heme iron center [32]. The activity increase was first observed during coexpression of the DesVII/DesVIII pair in *E. coli*, and resulted in an enhancement of DesVII catalysed glycosyl transfer by a factor of 10^3 [113], [114]. Similarly the glycosyl transfer of TDP-D-deosamine onto 3-α-mycarosylerythronolide B by EryCIII was enhanced in the presence of the helper protein EryCII. The structural basis of rate enhancement excerpted by EryCII has not been unambiguously deduced even in light recent of structural data of the EryCIII/EryCII complex, but was suggested to arise from a change of protein conformation of the GT [115], or by the helper protein stabilising its partner and acting as an allosteric activator [116]. Conformational stabilization of the active form was also suggested in the case of the GT AknS, where presence of AknT resulted in a 40-fold increase of K_{cat} [117]. AknT shares a 37.5% sequence identity to the putative protein SnogN encoded by the *sno* gene cluster. In addition to the AknT homology, SnogN has a sequence identity above 35% to several cytochrome p450 enzymes, particularly from Actinobaceria. It is at this stage tempting to suggest the involvement of SnogN during C-C bond formation, perhaps acting as a helper protein to SnogD similarly to AknS, but no genetic and biochemical data support this today.

In view of the lack of biochemical data the mechanism of C-C bond formation during nogalamycin biosynthesis remains elusive.

3.2 STRUCTURAL ENZYMOLOGY OF THE BIFUNCTIONAL DEHYDRATASE/ISOMERASE ALDOS-2-ULOSE DEHYDRATASE FROM PHANEROCHAETE CHRYSOSPORIUM (PAPER II)

3.2.1 Recombinant protein production and sequencing

AUDH from *Phanerochaete chrysosporium* (UNIPROT ID: P84193) was cloned by our collaborators at Danisco, Denmark, based on genome sequence information [118] and the partially sequenced protein [72]. The protein was produced in the methanotropic yeast *Hansenula polymorpha*, followed by amino acid sequencing of the full length protein. No signal peptide sequence or associated cleavage site was identified when analysing the sequence of the translated gene and up to 180 nucleotide bases of the upstream neighbouring region with Signal P 4.0 [119], potentially arguing against an extracellular localisation for AUDH.

Initial database searches for sequence homologues provided little or no information with regards to putative domain architecture, 3D folds, structural or ligand binding motifs. Partial sequence homologs of moderate homology were found, but none of these corresponded to characterised enzymes. Within the last year additional sequences of putative proteins have been deposited from Ascomycota, Basidomycota and Streptomycetaceae, with up to 51% sequence identity to AUDH (Fig. 3.16).

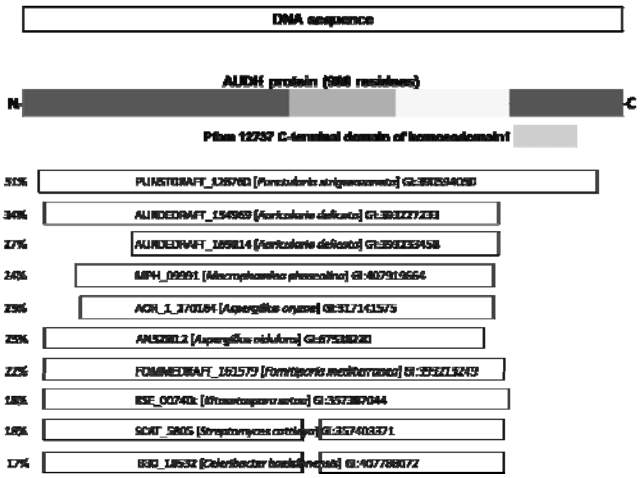

Figure 3.16 – Graphical representation of AUDH sequence homologues. The ten sequences that were received the highest scores by pBLAST (NCBI) (all with E-values below $7e^{-63}$) are presented with homologous regions represented by the boxes. The respective sequence identities to AUDH are listed on the left. For these ten sequences the residues binding the site_1 and site_3 zinc ions are strictly conserved and the site_2 Zn^{2+} and the Mg^{2+} binding motifs are conserved.

3.2.2 Crystallisation and structure determination

AUDH was crystallised in the space group $P2_12_12$, resulting in well diffracting crystals (Fig. 3.17). The structure of AUDH was solved by the single wavelength anomalous diffraction method, using a mercury derivative, at 2.1 Å resolution. The three deposited structures of AUDH were refined to: 2.0 Å for the native protein (PDB ID: 4a7k), 2.6 Å in complex with APM (PDB ID: 4a7z) and 2.8 Å for the zinc depleted AF complex (PDB ID: 4a7y) (data collection and refinement statistics in paper II). The presence of zinc was detected by X-ray fluorescence scanning, and collection of additional anomalous data at the zinc K adsorption edge enabled identification of the three zinc binding sites.

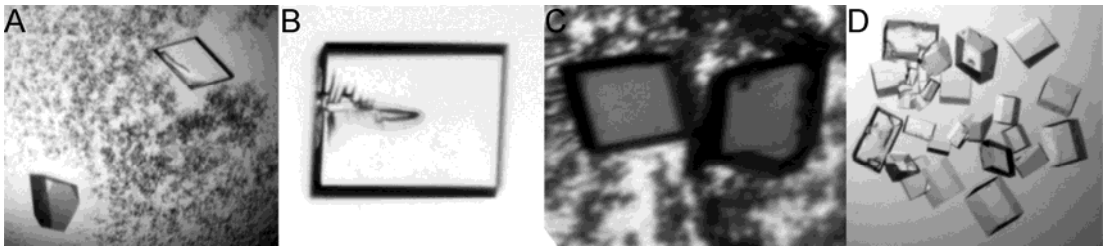

Figure 3.17 – Crystals of AUDH. A) and B) Native crystals. C) and D) Co-crystallisation with AF and 1,10-phenathroline respectively.

3.2.3 AUDH is an all β-protein

AUDH is an all β-protein and consists of three domains; an N-terminal seven bladed β-propeller, a central domain consisting of two cupin folds and a C-terminal domain showing structural homology to the ricin type lectins (Fig. 3.18) [120]. In spite of a lack of significant overall sequence identity to other proteins, the three domains have structural homologues in the Protein Data Bank (PDB). However there are no structures of proteins with the same domain composition as AUDH.

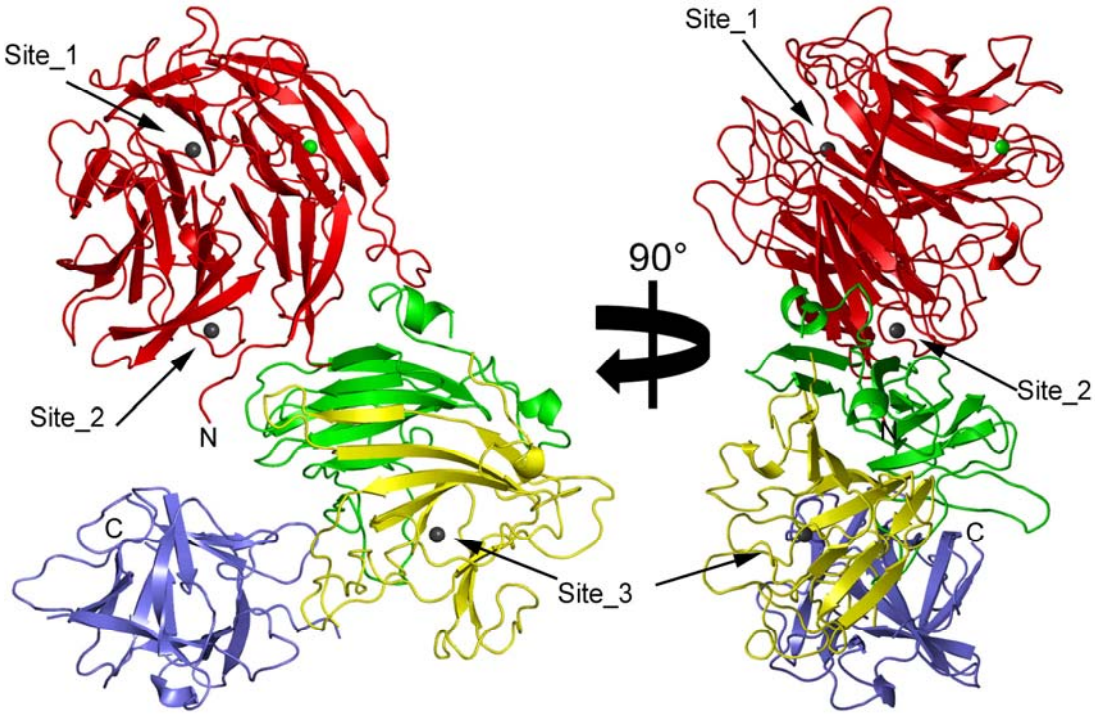

Figure 3.18 – The structure of the AUDH monomer, shown in two orientations. The β-propeller domain is coloured red, the first cupin fold green, the second cupin fold yellow and the lectin domain is coloured blue. Bound ions are shown as spheres, with the two catalytic zinc ions (Site_1 and Site_3) and the structural Zn^{2+} (Site_2) coloured gray, and the structural Mg2+ coloured green.

The N-terminal seven bladed β-propeller domain encompasses almost half of the 900 residue long polypeptide chain (residues 1-443), and includes two Zn^{2+} binding sites and an additional Mg^{2+} binding site. The magnesium ion and the site_2 zinc ions are coordinated with an octahedral geometry in loop regions, suggesting structural roles (Fig. 3.19A&B). The β-propeller does not contain the commonly observed central pore [121]. Instead, side chains of the central β-strands occlude approximately half of the distance of the propeller axis (Fig 3.20).

The three histidine residues coordinating the site_1 zinc ion are located above the plane of the propeller blade at the "top" end of the pore occlusion. Their side chain imidazole rings bind the zinc ion in a tetrahedral coordination with a water molecule as the fourth ligand (Fig 3.19C). These histidines and the zinc ion form the base of the active site pocket, which is lined and capped by the relative long loops connecting the β-strands of the propeller blades. The observed conformation of these loops suggests substrate entry and product exit following loop rearrangement.

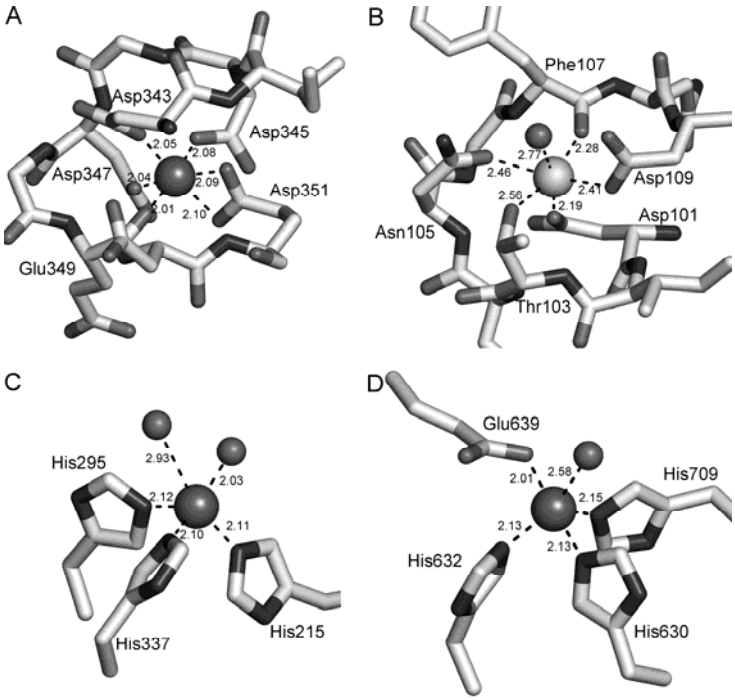

Figure 3.19 – The metal ion coordination in AUDH. A) Structural zinc ion at site_2 in. B) Structural magnesium ion. C) Site_1 zinc ion. D) Site_3 zinc ion. The metal-ligand distances are marked in Å.

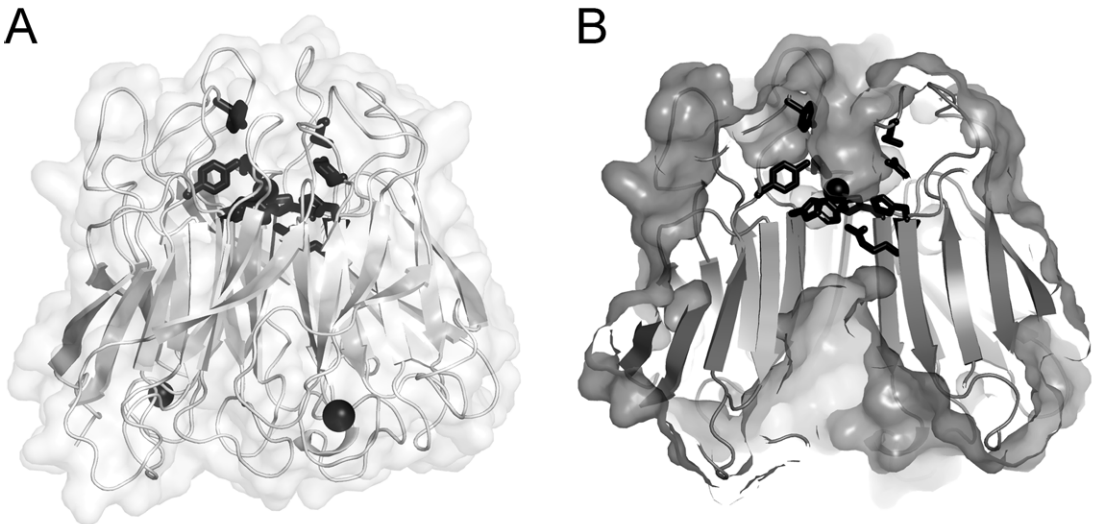

Figure 3.20 – The β-propeller domain of AUDH. A) The loops β- connecting the strands extends up to 15 Å away from the propeller blades, lining and capping the active site by site_1. This site is shown at the top centre with the zinc ion as a black sphere and the ligand coordinating residues as black sticks. The structural metal ions are shown as black spheres, with the site_2 Zn^{2+} located at the lower left and the structural Mg^{2+} located at the lower right. B) Cross-section of the β-propeller showing the propeller pore occlusion and the site_1 volume. The site_1 zinc ion is located at the base of the active site accessible from the top.

The central domain consists of two cupin folds, which are oriented similarly to bi-cupin fold proteins [122], [123] (Fig. 3.21). Automated database searches for structural homologues provided structural homologues belonging to the bicupins. However the individual cupin folds of AUDH only superimposed well on one of the two folds of the respective bicupins, due to differences in orientation of the respective cupin folds. The second cupin fold contains the site_3 zinc ion, which is coordinated at the base of the *cupa* in a trigonal-bipyramidal geometry by histidine and glutamic acid side chains and one water molecule (Fig. 3.19D).

Figure 3.21 – The central bicupin domain of AUDH. The two cupin folds are represented as semi-transparent surface and cartoon, with the first cupin fold coloured black and the second site_3 fold coloured pale grey. The site_3 zinc ion is shown as a black sphere.

The C-terminal domain displays a β-trefoil, first reported for ricin [120] and subsequently in hydrolases, cytokines, toxins and lectins (Fig 3.22). The trefoil gives rise to an internal pseudo-three fold symmetry, with each motif consisting of four β-strands. Of the three motifs one is involved in the dimer interface, the second is facing towards the barrel-opening of the second cupin fold and the third motif faces the solvent (Fig. 3.22).

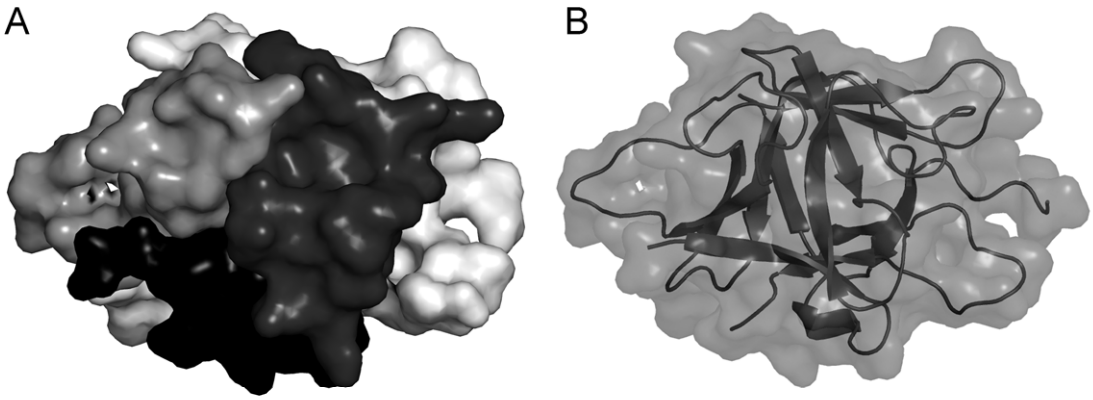

Figure 3.22 – The pseudo 3-fold symmetry of the lectin fold in AUDH. A) Surface representation showing the polypeptide segment connecting the bicupin and lectin domains in light grey, and the three motifs are coloured: light gray (facing outwards), dark grey (involved in the subunit interface) and black (facing the cupin opening). B) Same orientation with semi-transparent surface showing a cartoon representation of the polypeptide chain.

The role of this domain is unclear, but the presence of this fold in many poly-saccharide binding proteins could indicate a similar carbohydrate binding role, perhaps enabling enrichment of AUDH at a location of carbohydrate degradation and substrate generation. The presence of residues corresponding to this domain is not conserved among the sequence homologues of AUDH (Fig. S3 paper II), perhaps indicating a species specific function of this domain in *P. chrysosporium*.

3.2.4 AUDH requires zinc ions for activity

Bioinformatics analysis using the protein sequence of AUDH did not predict any metal binding motifs in AUDH, however a reduction in activity in the presence of the chelator ethylenediaminetetraacetic acid (EDTA) was previously reported for AUDH from *P. chrysosporium* [69]. Such reduction in activity could by us only be achieved using the chelator 1,10-phenanthroline. Presence of micromolar concentrations of the polyaromatic chelator complicated spectroscopic-based activity studies, due to the strong inherent absorption of 1,10-phenanthroline in the wavelength range of interest for activity studies (230-262 nm). Extensive dialysis led to an almost complete removal of the chelator, and subsequent activity experiments showed an overall dependency on zinc ions for catalysis (Fig 3.23).

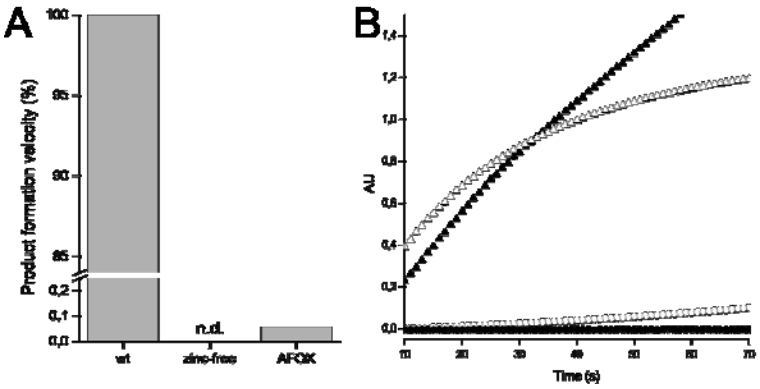

Figure 3.23 – The zinc dependency of AUDH and 1,5-anhydro-D-fructose oxime (AFOX) inhibition. A) Relative reaction rates of wild-type (left), zinc-free (middle) and AFOX inhibited AUDH (right). Formation of the product was monitored at 230 nm, the maximum absorbance of microthecin. B) Time-course monitoring formation of the reaction intermediate (APM), at 262nm as open symbols, and the product microthecin, at 230 nm as filled symbols, for wild-type (triangles) and zinc-free (circles) AUDH.

3.2.5 Co-crystallisation with substrate and intermediate

Extensive screening for a ligand complex with AUDH was performed in pursuit of structural data providing information of ligand binding and enzyme catalysis. Co-crystallisation and soaking of AUDH crystals was performed in a rather wide pH range (4.5-9.0) in presence and absence of ligands (AF, AF/Mic, Mic, 2-keto-D-glucose, various alcohol sugars and hexoses such as sorbitol and glucose) in concentrations up to 0.5 M. Optimisation of the conditions resulted in crystals that diffracted well enough for data collection and more than 30 data sets with a resolution higher than 3 Å were collected. However, none of them revealed additional electron density of sufficient size or shape to accommodate the respective ligands.

Eventually, a complex with APM was obtained by co-crystallisation of a modified form of 1,5-anhydro-D-fructose, the 1,5-anhydro-fructose-oxime (AFOX) [124]. After structure determination, positive difference electron densities of a mainly planar shape were found at both site 1 and 2, of a size and shape that best fitted APM (Fig. 5

and S1, paper II). The serendipitous trapping of APM likely resulted from a slow turnover of AFOX in the crystallisation drop, in combination with a greater affinity of the enzyme for APM, compared to the product. This is supported by the observation that AFOX inhibits the overall reaction of AUDH (Fig. 3.23A). No electron density was observed that could be attributed to potential hydroxylamine-protein adducts or peptide bond cleavage events, indicating that the reduced catalytic activity could not be due to amination or peptide-bond cleavage effects, or interference with ligand binding.

After discovery of the zinc dependency of AUDH catalysis, the enzyme was crystallised in presence of the metal chelator 1,10-phenathroline and soaked with AF. The calculated electron density maps showed the specific removal of the catalytic zinc ions at site_1 and site_3, with the structural metal sites essentially intact. Additional electron density at site_3 was observed that in size and shape suggested bound AF, with the C2 and C3 hydroxyl groups in equatorial positions (Fig. 6 and S2, paper II). The product Mic could not be modelled into this density, due to the considerable bulkiness of this molecule.

3.2.6 Reaction chemistry of AUDH

The two step reaction catalysed by AUDH is likely to occur in two distinct catalytic centres, that are located in the β-propeller and bicupin domain. This is supported by the observation of APM and AF binding to these sites, presence of the zinc ions at the bottom of these ligand binding cavities, and the specific interactions with the respective ligands in the two complexes. Accumulation of APM prior to steady-state concentrations was observed (Fig 3.24), indicating release of APM and diffusion to a second active site, which is in accordance with NMR data showing APM release [75] and formation of APT following non-enzymatic dihydration of APM [72].

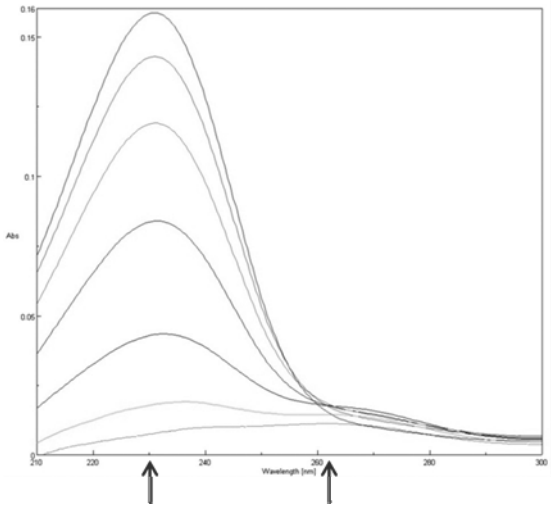

Figure 3.24 – Lag phase of Mic production by AUDH. The pre-steady state accumulation of APM is observed by the absorption maxima of APM 262 nm (right arrow) and the lag in Mic formation at the absorption maxima of Mic 230 nm (left arrow). The first time point is the bottom line, with the curves recorded at subsequent time points accumulating on top.

The AUDH homodimer is formed through tight subunit interactions between the bicupin and lectin domains over a 2800 $Å^2$ large surface. The subunit interface is not continuous, but contains two channels formed between the subunits (Fig. 3.25), potentially enabling tunnelling of molecules between the two active sites. In the AUDH-AF complex additional electron density was observed at the centre of the dimer

interface, adjacent to one of the channel segments and near to the entrance of the second. This was modelled as AF with partial occupancy.
Carbohydrate accumulation at this location close to the centre of the dimer could potentially mimic one of the stop points during diffusion within or between the channels. The presence of two tunnel segments covering a large part of the distance between the two active sites would restrict random diffusion of the intermediate and facilitate compound transfer between the sites, likely increasing the catalytic rate, as was recently observed for the bifunctional enzyme dethiobiotin synthetase–diamino-pelargonic acid aminotransferase from *Arabidopsis thaliana* [125].

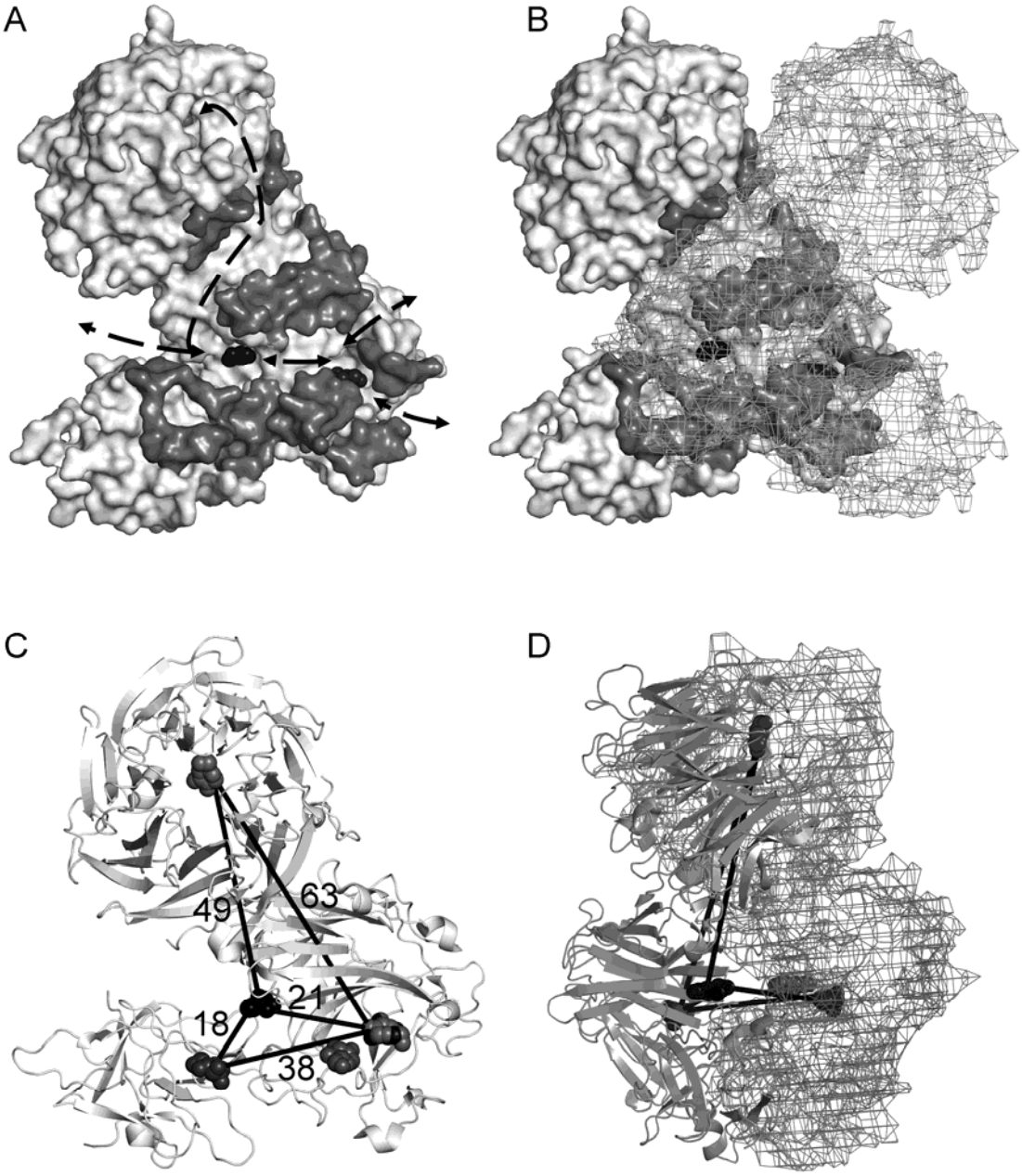

Figure 4.25 – A) Illustration of the subunit-subunit interface of the AUDH dimer with access paths to the two sites through channels and pores indicated by arrows. Interface residues are coloured dark gray, AF molecules are shown as black spheres, and APM as gray spheres. B) The loaction of the second subunit illustrated by mesh tracing. C) Cartoon representation of the monomer with ligand distances given in Å. D) As B, but rotated 90 degrees with the distances from C indicated by black lines.

Due to the lack of suitable acid-base catalysts at site_3, and the presence of such at site_1 we hypothesised that site_1 would perform the initial dehydration reaction. This reaction would resemble that of 2-keto-3-deoxy-D-arabinonate dehydratase [126], with a magnesium ion polarising a carbonyl carbon of the substrate, which in turn facilitates proton abstraction at the adjacent carbon. Based on the position of AF at site_1 in the modelled Michaelis complex, several putative active-site residues can be identified in the vicinity of AF (Fig. 7, paper II), with His155, located between the C3 and the C4 hydroxyl group of AF as a good candidate for proton transfer.
The dehydration reaction would by this model proceed via an elimination mechanism, with His155 abstracting the hydrogen from the C3 carbon (Fig. 3.26). Proton abstraction would be facilitated by the zinc ion acting as a Lewis acid on the C2 carbonyl group, thereby reducing the pK_a of the C3 hydrogen. The formed enolate may be stabilised by the zinc ion. Proton transfer from His155 to the C4-hydroxyl group would result in a better leaving group, the departure of which would complete the dehydration reaction. The closely located side chain of Arg156 forms stacking interactions with the imidazole ring of His155, enforcing a suitable deprotonated state of His155 at the start of catalysis.

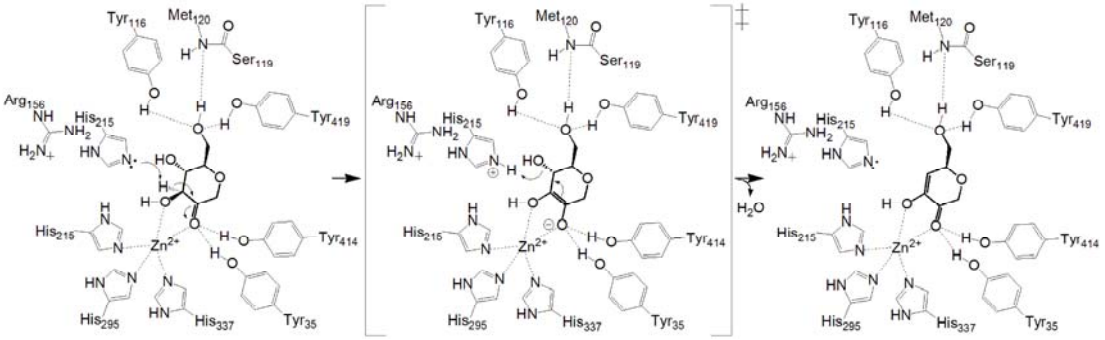

Figure 3.26 – The prostulated elimination reaction mechanism for the AUDH catalysed dehydration of AF.

Two of the four residues located within hydrogen bonding distance of AF/APM at site_3 are the zinc-liganding residues Glu639 and His641. The other two hydrogen bonding partners are the carbonyl group of Ala627 and the side chain of Trp726. In the structure of the zinc depleted AUDH an additional hydrogen bond is formed between His630 and the C2 hydroxyl group of AF, consequently pulling the hydroxyl group close to the position of the zinc ion observed in the native and APM-complex structures. However, the absence of suitable residues for acid/base catalysis at this site and the presence of such at site_1 prompted us to suggest AF binding to site_3 is an artefact resulting from the absence of the zinc ion. The absence of AF at site_1 in the ADUH-AF complex could perhaps be explained by a requirement of the zinc ion for ligand binding.

Both reaction steps of the bifunctional AUDH require further biochemical investigation to elucidate the reaction chemistry. Active site mutagenesis targeting the individual reaction centres would prove or disprove the hypothesis that site_1 is responsible for the dehydration. Subsequent biochemical characterisation of individual point mutants with exchanges at the respective active sites would provide insights into the reaction chemistry.

4 CONCLUSIONS

This thesis focuses on two distinct classes of enzymes, however they share several key features. Both are involved in natural product biosynthesis during secondary metabolism, and both participate in complex chemical reactions that are likely to require novel reaction mechanisms. For both of these enzymes a combination of structural and biochemical studies have provided new insights into their catalytic activities and the mechanism by which these are performed.

The data obtained after *in vivo* reconstitution of the nogalamycin biosynthesis identified the specific activities of the glycosyltransferases SnogD and SnogE. SnogE is the *O*-glycosyltransferase performing the first glycosyl transfer onto the C7 carbon of the aglycone, using TDP-2,3,4-tridemethoxy nogalose (**7**) as donor substrate.
The genetic, functional data and biochemical data obtained for SnogD defines its role as the C1 carbon *O*-glycosyltransferase, transfering the nogalamine moiety from TDP-L-acosamine (**8**). Additionally, activity by SnogD requires a C7 glycosylated acceptor substrate, suggesting an additional selective binding interaction of this carbohydrate moiety. Furthermore the SnogD catalysed *O*-glycosyl linkage precedes formation of the *C*-glycosyl bond.
It remains however unclear whether SnogD is directly involved in the subsequent *C*-glycosylation step. This would require activation of the substrate by hydride transfer or a radical mechanism, since no leaving group or functional group enabling formation of the C-C bond is present. This reaction requires either an additional helper protein recruited by SnogD or the action of a different enzyme. However the number of suitable gene candidates for this is very short within the nogalamycin gene cluster. Studies to identify this protein and characterise the mechanisms involved promise exciting discoveries in the future, with potential applications in combinatorial biosynthesis.

The structure of the bifunctional enzyme AUDH showed a novel combination of folds, with two reaction centres separated by 60 Å. A channel is formed at the homo-dimer interface, which could enable substrate transport between the two active sites. The discovery of the metal binding sites was, based on protein sequence homology, unexpected. The binding position and coordination geometry of the site_1 and site_3 zinc ions suggested their involvement in catalysis, which was verified by the loss of activity upon specific removal using a chelator. The structure of the zinc depleted form of AUDH was very similar structurally, ruling out loss of activity from structural disturbances due to metal removal. The ligand complexes with APM and AF provided insights into how the respective substrate and intermediate bind at the respective sites, as well as identifing potential active site residues. This information enabled a reaction mechanism for the dehydration step to be postulated, based on acid/base catalysis.
Further biochemical studies are required to dissect the reaction chemistry of the respective active sites. In particular the reaction chemistry of the final complex isomerisation reaction poses an interesting biochemical challenge, potentially revealing an alternate mode of carbohydrate isomerisation.

These studies have themselves provided new insights into both enzymes and now provide the foundation for future studies to fully elucidate the unusual mechanisms of the two enzymes. This would ultimately enable modification of the chemistry to enable production of new products such as novel antibiotics.

5 ACKNOWLEDGEMENTS

First I would like to thank you Gunter for accepting me as an undergraduate and subsequently as a PhD student. The enthusiasm and drive you display has been and still is inspiring. The support and guidance throughout my years in the group is most appreciated. Many thanks to you and Ylva for nurturing a nice working environment, and activities such as group-dinners.

I would also like to express my gratitude to you Doreen, for taking me on as an undergraduate student and introducing me into the field. Nice parties and food just came with the science it seemed.

Thanks to my collaborators at the University of Turku in Finland. Great times working with you, and great times off hours. Especially Mikko for sharing my black humour in the office, and Vilja for analysing all the extracts I sent. Also thanks to Pauli, Thadee and Igor, for interesting discussions. Thanks to Shukun and my collaborators at Danisco, that was quite some protein.

Many thanks to the past and present members of the Molecular Structural Biology group. Åggi the best Norrlands-j I have ever met, you are great company come rain or shine! Thanks Eddie you too have been a good companion, ever since the start and not only whilst cultivating *E. coli* again and again and again. Robert you are always helpful with scientific issues, and a deep source of discussion (off)topics, sometimes wiser than we realised. Hanna for helping me understand the temperament of the polyketides, and always being quick to a laugh and friendly banter. Rajesh a man full of optimism and devotion, and keen on discussing science at any time. Maria we had nice synchrotron trips, lab times and some confusing teaching moments, at least when it came to the discussions. I still think more cable-car related experiments with local experts are required. Talking of such, thanks Jean-Marie for conversations and insights into the minds of us Swedes. Jason, the first person I met that truly understands and follows my more obscure popular cultural references and hints, those facilitated the deep thoughts we shared. Dominik always hungry for new approaches to science and great company outside the lab. My current office mates, Ahmad for running the lab, always helpful and up for a discussion. Dominic as the runner up, I enjoyed the political banter and the discussions from chemistry to medicine. From now I will not compete for the seminar room anymore. Ming-Wei soon the border disputes are over and you can finally expand. “Let’s go” Ömer – leaving us a step behind, always brewing a burning topic for discussion. Thanks Jola for the scientific tricks, and for joining me in clearing out the darker corners of the lab. Katja and Cyprian for understanding the sleepless nights and sharing discussion on diapers - now I know what it is all about. A special thanks to Bernie our local Coot-man, for the tips and tricks, in addition to the helpful and observant eye when it comes to science. And of course for the introduction into molecular biology and protein chemistry mixed up with pub-crawls. Thanks to you Tanja, for the nice times we shared at the synchrotron and in the lab, it really should be that straightforward all the time. Katharina, sorry for dragging you down the TLC-route, but it seemed to work in the end. Nice getting to know you Eva-Maria, sorry there was not much time for Swedish conversations.

Thanks Laura for sharing the bright and the dark times in the 96 wells. The people of the early days… Thanks Catrine and Stina for help and support, and for being nice company as bench-mates. Special thanks to you Eva, not only for the “recruitment”, but also for help over the years. Thanks also to Alessandra and Victoria for guiding me through the paperwork, as well as nice moments out of office.

I would like to thank the teaching unit at MBB; Åke, Susanne, Agneta and Susie. I appreciated the teaching and the discussions about how it was and could be performed. It was quite an experience. Special thanks to Susanne for scientific and not so scientific conversation and off hours access to equipment and chemicals.

Thanks to Olle, Ylva, Essam, Andreas, Sergio, Emma and Aussi, for the parties we had and to a much lesser extent, their aftermaths. Thomas your never fading interest stands alone, also thanks for lending me the cardigan, it did not break the fall but at least it kept me warm. Thanks to the football crew of old, when even I played; Jakko, Juha, Massa, Mikko, Seth and Phillip. I guess, also thanks Jakko for the matchmaking skills you did not know you had.

Stort tack till familjen och bekanta. Martin, alltid välönskande och stöttande, du är till viss del en orsak till detta. Det kanske ligger något i KB trots allt. Tack Pappa och mamma för stöd under vägen och uthållighet. Marie, det var “mycke detaljer” ett tag, men det blev inte bättre. Tack Mike för alla minnesvärda stunder och strapatserna vi delat, jag kommer dock fortfarande inte reka i Dalarna. Alex tack för din övermänskliga förmåga att förgöra vila och sömn, samt leendena som avbröt dem.

Jodie tack för allt.

6 REFERENCES

[1] D. H. Williams, M. J. Stone, P. R. Hauck, and S. K. Rahman, “Why Are Secondary Metabolites (Natural Products) Biosynthesized?,” *Journal of Natural Products*, vol. 52, no. 6, pp. 1189–1208, Nov. 1989.

[2] F. T. Wong and C. Khosla, “Combinatorial biosynthesis of polyketides--a perspective.,” *Current opinion in chemical biology*, vol. 16, no. 1–2, pp. 117–23, Apr. 2012.

[3] C. Hertweck, “The biosynthetic logic of polyketide diversity.,” *Angewandte Chemie (International ed. in English)*, vol. 48, no. 26, pp. 4688–716, Jan. 2009.

[4] K. J. Weissman, “Introduction to polyketide biosynthesis.,” *Methods in enzymology*, vol. 459, no. 09, pp. 3–16, Jan. 2009.

[5] A. Das and C. Khosla, “Biosynthesis of aromatic polyketides in bacteria.,” *Accounts of chemical research*, vol. 42, no. 5, pp. 631–9, May 2009.

[6] S. Smith and S.-C. Tsai, “The type I fatty acid and polyketide synthases: a tale of two megasynthases.,” *Natural product reports*, vol. 24, no. 5, pp. 1041–72, Oct. 2007.

[7] C. Hertweck, A. Luzhetskyy, Y. Rebets, and A. Bechthold, “Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.,” *Natural product reports*, vol. 24, no. 1, pp. 162–90, Feb. 2007.

[8] S.-C. S. Tsai and B. D. Ames, *Structural enzymology of polyketide synthases.*, 1st ed., vol. 459, no. 09. Elsevier Inc., 2009, pp. 17–47.

[9] H. Brockmann, K. Bauer, and I. Borchers, “Rhodomycin, ein rotes Antibioticum (Antibiotica aus Actinomyceten, VII. Mitteil.,” *Chemische Berichte*, vol. 84, no. 8, pp. 700–710, Sep. 1951.

[10] G. Minotti, P. Menna, E. Salvatorelli, G. Cairo, and L. Gianni, “Anthracyclines: Molecular Advances and Pharmacologic Developments in Antitumor Activity and Cardiotoxicity,” *Pharmacological Reviews*, vol. 56, no. 2, pp. 185–229, 2004.

[11] P. Menna, O. Gonzalez Paz, M. Chello, E. Covino, E. Salvatorelli, and G. Minotti, “Anthracycline cardiotoxicity.,” *Expert opinion on drug safety*, pp. 1–16, Jun. 2011.

[12] G. Schneider, “Enzymes in the biosynthesis of aromatic polyketide antibiotics.,” *Current opinion in structural biology*, vol. 15, no. 6, pp. 629–36, Dec. 2005.

[13] S. K. Arora, “Molecular structure, absolute stereochemistry, and interactions of nogalamycin, a DNA-binding anthracycline antitumor antibiotic,” *Journal of the American Chemical Society*, vol. 105, no. 5, pp. 1328–1332, Mar. 1983.

[14] Y. C. Liaw, Y. G. Gao, H. Robinson, G. a van der Marel, J. H. van Boom, and a H. Wang, “Antitumor drug nogalamycin binds DNA in both grooves simultaneously: molecular structure of nogalamycin-DNA complex.,” *Biochemistry*, vol. 28, no. 26, pp. 9913–8, Dec. 1989.

[15] Y. G. Gao, Y. C. Liaw, H. Robinson, and a H. Wang, “Binding of the antitumor drug nogalamycin and its derivatives to DNA: structural comparison.,” *Biochemistry*, vol. 29, no. 45, pp. 10307–16, Nov. 1990.

[16] C. K. Smith, G. J. Davies, E. J. Dodson, and M. H. Moore, “DNA-nogalamycin interactions: the crystal structure of d(TGATCA) complexed with nogalamycin.,” *Biochemistry*, vol. 34, no. 2, pp. 415–25, Jan. 1995.

[17] L. H. Li and W. C. Krueger, “The biochemical pharmacology of nogalamycin and its derivatives.,” *Pharmacology & therapeutics*, vol. 51, no. 2, pp. 239–55, Jan. 1991.

[18] P. F. Wiley, D. W. Elrod, and V. P. Marshall, “Biosynthesis of the anthracycline antibiotics nogalamycin and steffimycin B,” *The Journal of Organic Chemistry*, vol. 43, no. 18, pp. 3457–3461, Sep. 1978.

[19] B. Shen and C. R. Hutchinson, “Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 13, pp. 6600–4, Jun. 1996.

[20] J. Kantola, G. Blanco, A. Hautala, T. Kunnari, J. Hakala, C. Mendez, K. Ylihonko, P. Mäntsälä, and J. Salas, “Folding of the polyketide chain is not dictated by minimal polyketide synthase in the biosynthesis of mithramycin and anthracycline.,” *Chemistry & biology*, vol. 4, no. 10, pp. 751–5, Oct. 1997.

[21] T. Grocholski, H. Koskiniemi, Y. Lindqvist, P. Mäntsälä, J. Niemi, and G. Schneider, “Crystal structure of the cofactor-independent monooxygenase SnoaB from Streptomyces nogalater: implications for the reaction mechanism.,” *Biochemistry*, vol. 49, no. 5, pp. 934–44, Feb. 2010.

[22] S. Torkkell, T. Kunnari, K. Palmu, P. Mäntsälä, J. Hakala, and K. Ylihonko, “The entire nogalamycin biosynthetic gene cluster of Streptomyces nogalater: characterization of a 20-kb DNA region and generation of hybrid structures,” *Molecular Genetics and Genomics*, vol. 266, no. 2, pp. 276–288, Oct. 2001.

[23] A. Sultana, P. Kallio, A. Jansson, J.-S. Wang, J. Niemi, P. Mäntsälä, and G. Schneider, “Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation.,” *The EMBO journal*, vol. 23, no. 9, pp. 1911–21, May 2004.

[24] V. Siitonen, M. Claesson, P. Patrikainen, M. Aromaa, P. Mäntsälä, G. Schneider, and M. Metsä-Ketelä, “Identification of late-stage glycosylation steps in the biosynthetic pathway of the anthracycline nogalamycin.,” *Chembiochem: a European journal of chemical biology*, vol. 13, no. 1, pp. 120–8, Jan. 2012.

[25] P. Beinker, B. Lohkamp, T. Peltonen, J. Niemi, P. Mäntsälä, and G. Schneider, “Crystal structures of SnoaL2 and AclR: two putative hydroxylases in the biosynthesis of aromatic polyketide antibiotics.,” *Journal of molecular biology*, vol. 359, no. 3, pp. 728–40, Jun. 2006.

[26] V. Siitonen, B. Blauenburg, P. Kallio, P. Mäntsälä, and M. Metsä-Ketelä, “Discovery of a Two-Component Monooxygenase SnoaW/SnoaL2 Involved in Nogalamycin Biosynthesis.,” *Chemistry & biology*, vol. 19, no. 5, pp. 638–46, May 2012.

[27] A. C. Weymouth-Wilson, “The role of carbohydrates in biologically active natural products.,” *Natural product reports*, vol. 14, no. 2, pp. 99–110, May 1997.

[28] B. La Ferla, C. Airoldi, C. Zona, A. Orsato, F. Cardona, S. Merlo, E. Sironi, G. D’Orazio, and F. Nicotra, “Natural glycoconjugates with antitumor activity,” *Natural Product Reports*, pp. 630–648, 2011.

[29] M. F. Giraud, G. A. Leonard, R. A. Field, C. Berlind, and J. H. Naismith, “RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase.,” *Nature structural biology*, vol. 7, no. 5, pp. 398–402, May 2000.

[30] C. Dong, L. L. Major, V. Srikannathasan, J. C. Errey, M.-F. Giraud, J. S. Lam, M. Graninger, P. Messner, M. R. McNeil, R. a Field, C. Whitfield, and J. H. Naismith, “RmlC, a C3’ and C5' carbohydrate epimerase, appears to operate via an intermediate with an unusual twist boat conformation.,” *Journal of molecular biology*, vol. 365, no. 1, pp. 146–59, Jan. 2007.

[31] H. Chen, Z. Zhao, T. M. Hallis, Z. Guo, and H. Liu, “Insights into the Branched-Chain Formation of Mycarose : Methylation Catalyzed by an,” *Angewandte Chemie (International ed. in English)*, vol. 40, no. 3, pp. 607–610, 2001.

[32] C. J. Thibodeaux, C. E. Melançon, and H. Liu, “Natural-product sugar biosynthesis and enzymatic glycodiversification.,” *Angewandte Chemie (International ed. in English)*, vol. 47, no. 51, pp. 9814–59, Jan. 2008.

[33] V. Siitonen, M. Claesson, P. Patrikainen, M. Aromaa, P. Mäntsälä, G. Schneider, and M. Metsä-Ketelä, “Identification of late-stage glycosylation steps in the biosynthetic pathway of the anthracycline nogalamycin.,” *Chembiochem : a European journal of chemical biology*, vol. 13, no. 1, pp. 120–8, Jan. 2012.

[34] H. Chen, G. Agnihotri, and Z. Guo, “Biosynthesis of mycarose: isolation and characterization of enzymes involved in the C-2 deoxygenation,” *Journal of the American …*, no. c, pp. 8124–8125, 1999.

[35] M. L. Davis, J. B. Thoden, and H. M. Holden, “The x-ray structure of dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase.,” *The Journal of biological chemistry*, vol. 282, no. 26, pp. 19227–36, Jun. 2007.

[36] E. S. Burgie, J. B. Thoden, and H. M. Holden, “Molecular architecture of DesV from Streptomyces venezuelae: a PLP-dependent transaminase involved in

the biosynthesis of the unusual sugar desosamine.," *Protein science : a publication of the Protein Society*, vol. 16, no. 5, pp. 887–96, May 2007.

[37] R. Kleene and E. G. Berger, "The molecular and cell biology of glycosyltransferases.," *Biochimica et biophysica acta*, vol. 1154, no. 3–4, pp. 283–325, Dec. 1993.

[38] L. Tu and D. K. Banfield, "Localization of Golgi-resident glycosyltransferases.," *Cellular and molecular life sciences : CMLS*, vol. 67, no. 1, pp. 29–41, Jan. 2010.

[39] P. Stanley, "Golgi glycosylation.," *Cold Spring Harbor perspectives in biology*, vol. 3, no. 4, Apr. 2011.

[40] J. Campbell, G. Davies, V. Bulone, and B. Henrissat, "A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities," *The Biochemical journal*, vol. 329 (Pt 3), p. 719, Feb. 1998.

[41] P. M. Coutinho, E. Deleury, G. J. Davies, and B. Henrissat, "An Evolving Hierarchical Family Classification for Glycosyltransferases," *Journal of Molecular Biology*, vol. 328, no. 2, pp. 307–317, Apr. 2003.

[42] B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.," *Nucleic acids research*, vol. 37, no. Database issue, pp. D233–8, Jan. 2009.

[43] L. L. Lairson, B. Henrissat, G. J. Davies, and S. G. Withers, "Glycosyltransferases: structures, functions, and mechanisms.," *Annual review of biochemistry*, vol. 77, pp. 521–55, Jan. 2008.

[44] A. Vrielinkl, W. Ruger, H. P. C. Driessen, and P. S. Freemont, "Crystal structure of the DNA modifying enzyme : -glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose," vol. 13, no. 15, pp. 3413–3422, 1994.

[45] S. Moréra, L. Larivière, J. Kurzeck, U. Aschke-Sonnenborn, P. S. Freemont, J. Janin, and W. Rüger, "High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding.," *Journal of molecular biology*, vol. 311, no. 3, pp. 569–77, Aug. 2001.

[46] C. a Wiggins and S. Munro, "Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 14, pp. 7945–50, Jul. 1998.

[47] C. Breton, E. Bettler, D. H. Joziasse, R. A. Geremia, and A. Imberty, "Sequence-Function Relationships of Prokaryotic and Eukaryotic," vol. 1009, pp. 1000–1009, 1998.

[48] J. E. Pak, P. Arnoux, S. Zhou, P. Sivarajah, M. Satkunarajah, X. Xing, and J. M. Rini, "X-ray crystal structure of leukocyte type core 2 beta1,6-N-

acetylglucosaminyltransferase. Evidence for a convergence of metal ion-independent glycosyltransferase mechanism.," *The Journal of biological chemistry*, vol. 281, no. 36, pp. 26693–701, Oct. 2006.

[49] S. J. Charnock and G. J. Davies, "Structure of the Nucleotide-Diphospho-Sugar Transferase , SpsA from Bacillus subtilis , in Native and Nucleotide-Complexed Forms †,‡," pp. 6380–6385, 1999.

[50] M. Mittler, A. Bechthold, and G. E. Schulz, "Structure and action of the C-C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin.," *Journal of molecular biology*, vol. 372, no. 1, pp. 67–76, Sep. 2007.

[51] M. L. Sinnott, *Carbohydrate Chemistry and Biochemistry*. The Royal Society of Chemistry, 2007.

[52] M. L. Sinnott, "Catalytic mechanism of enzymic glycosyl transfer," *Chemical Reviews*, vol. 90, no. 7, pp. 1171–1202, Nov. 1990.

[53] D. J. Vocadlo, G. J. Davies, R. Laine, and S. G. Withers, "Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate.," *Nature*, vol. 412, no. 6849, pp. 835–8, Aug. 2001.

[54] S. S. Lee, S. Y. Hong, J. C. Errey, A. Izumi, G. J. Davies, and B. G. Davis, "Mechanistic evidence for a front-side, SNi-type reaction in a retaining glycosyltransferase.," *Nature chemical biology*, vol. 7, no. 9, pp. 631–8, Sep. 2011.

[55] A. M. Mulichak, H. C. Losey, W. Lu, Z. Wawrzak, C. T. Walsh, and R. M. Garavito, "Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 16, pp. 9238–43, Aug. 2003.

[56] A. M. Mulichak, H. C. Losey, C. T. Walsh, and R. M. Garavito, "Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics.," *Structure*, vol. 9, no. 7, pp. 547–57, Jul. 2001.

[57] A. M. Mulichak, W. Lu, H. C. Losey, C. T. Walsh, and R. M. Garavito, "Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands.," *Biochemistry*, vol. 43, no. 18, pp. 5170–80, May 2004.

[58] A. Chang, S. Singh, K. E. Helmich, R. D. Goff, C. A. Bingman, J. S. Thorson, and G. N. Phillips, "Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regiospecificity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 43, pp. 17649–54, Oct. 2011.

[59] W. Offen, C. Martinez-Fleites, M. Yang, E. Kiat-Lim, B. G. Davis, C. a Tarling, C. M. Ford, D. J. Bowles, and G. J. Davies, "Structure of a flavonoid

glucosyltransferase reveals the basis for plant natural product modification.," *The EMBO journal*, vol. 25, no. 6, pp. 1396–405, Mar. 2006.

[60] L. Holm and P. Rosenström, "Dali server: conservation mapping in 3D.," *Nucleic acids research*, vol. 38, no. Web Server issue, pp. W545–9, Jul. 2010.

[61] T. D. H. Bugg, M. Ahmad, E. M. Hardiman, and R. Rahmanpour, "Pathways for degradation of lignin in bacteria and fungi.," *Natural product reports*, vol. 28, no. 12, pp. 1883–96, Nov. 2011.

[62] J. V. D. Brink and R. P. D. Vries, "Fungal enzyme sets for plant polysaccharide degradation," pp. 1477–1492, 2011.

[63] D. Singh and S. Chen, "The white-rot fungus Phanerochaete chrysosporium: conditions for the production of lignin-degrading enzymes.," *Applied microbiology and biotechnology*, vol. 81, no. 3, pp. 399–417, Dec. 2008.

[64] D. W. S. Wong, "Structure and action mechanism of ligninolytic enzymes.," *Applied biochemistry and biotechnology*, vol. 157, no. 2, pp. 174–209, May 2009.

[65] R. ten Have and P. J. Teunissen, "Oxidative mechanisms involved in lignin degradation by white-rot fungi.," *Chemical reviews*, vol. 101, no. 11, pp. 3397–413, Nov. 2001.

[66] K. Koths, R. Halenbeck, and M. Moreland, "Synthesis of the antibiotic cortalcerone from D-glucose using pyranose 2-oxidase and a novel fungal enzyme, aldos-2-ulose dehydratase.," *Carbohydrate research*, vol. 232, no. 1, pp. 59–75, Jul. 1992.

[67] B. M. Hallberg, C. Leitner, D. Haltrich, and C. Divne, "Crystal structure of the 270 kDa homotetrameric lignin-degrading enzyme pyranose 2-oxidase.," *Journal of molecular biology*, vol. 341, no. 3, pp. 781–96, Aug. 2004.

[68] R. Baute, M.-A. Baute, and G. Deffieux, "Proposed pathway to the pyrones cortalcerone and microthecin in fungi," *Phytochemistry*, vol. 26, no. 5, pp. 1395–1397, Apr. 1987.

[69] J. Gabriel, J. Volc, P. Sedmera, G. Daniel, and E. Kubátová, "Pyranosone dehydratase from the basidiomycete Phanerochaete chrysosporium: improved purification, and identification of 6-deoxy-D-glucosone and D-xylosone reaction products.," *Archives of microbiology*, vol. 160, no. 1, pp. 27–34, Jan. 1993.

[70] R. Baute, M.-A. Baute, G. Deffieux, and M.-J. Filleau, "Cortalcerone, a new antibiotic induced by external agents in Corticium caeruleum," *Phytochemistry*, vol. 15, no. 11, pp. 1753–1755, Jan. 1976.

[71] A. Broberg, L. Kenne, and M. Pederse, "4-Deoxy-glycero-hexo-2 , 3-diulose in Algae Using Gas Chromatography – Mass Spectrometry in Selected Ion Monitoring Mode," vol. 42, pp. 35– 42, 1999.

[72] S. Yu, “Enzymatic description of the anhydrofructose pathway of glycogen degradation II. Gene identification and characterization of the reactions catalyzed by aldos-2-ulose dehydratase that converts 1,5-anhydro-D-fructose to microthecin with ascopyrone M as the,” *Biochimica et biophysica acta*, vol. 1723, no. 1–3, pp. 63–73, May 2005.

[73] R. Fiskesund, L. V. Thomas, M. Schobert, I. Ernberg, I. Lundt, and S. Yu, “Inhibition spectrum studies of microthecin and other anhydrofructose derivatives using selected strains of Gram-positive and -negative bacteria, yeasts and moulds, and investigation of the cytotoxicity of microthecin to malignant blood cell lines.,” *Journal of applied microbiology*, vol. 106, no. 2, pp. 624–33, Feb. 2009.

[74] S. Yu, C. Refdahl, and I. Lundt, “Enzymatic description of the anhydrofructose pathway of glycogen degradation; I. Identification and purification of anhydrofructose dehydratase, ascopyrone tautomerase and alpha-1,4-glucan lyase in the fungus Anthracobia melaloma.,” *Biochimica et biophysica acta*, vol. 1672, no. 2, pp. 120–9, May 2004.

[75] S. Yu, M. Andreassen, and I. Lundt, “Enzymatic production of microthecin by aldos-2-ulose dehydratase from 1,5-anhydro- D -fructose and stability studies of microthecin,” *Biocatalysis and Biotransformation*, vol. 26, no. 1–2, pp. 169–176, Jan. 2008.

[76] M. Sakuma, S. Kametani, and H. Akanuma, “Purification and Some Properties of a Hepatic NADPH-Dependent Reductase That Specifically Acts on 1 , 5-Anhydro-D-Fructose,” vol. 123, no. 1, pp. 189–193, 1998.

[77] A. Kühn, S. Yu, and F. Giffhorn, “Catabolism of 1,5-anhydro-D-fructose in Sinorhizobium morelense S-30.7.5: discovery, characterization, and overexpression of a new 1,5-anhydro-D-fructose reductase and its application in sugar analysis and rare sugar synthesis.,” *Applied and environmental microbiology*, vol. 72, no. 2, pp. 1248–57, Feb. 2006.

[78] S. Yu, M. Andreassen, and I. Lundt, “Enzymatic production of microthecin by aldos-2-ulose dehydratase from 1,5-anhydro- D -fructose and stability studies of microthecin,” *Biocatalysis and Biotransformation*, vol. 26, no. 1–2, pp. 169–176, Jan. 2008.

[79] R. A. Field and J. H. Naismith, “Current Topics Structural and Mechanistic Basis of Bacterial Sugar Nucleotide-Modifying,” *Current*, vol. 42, no. 25, 2003.

[80] C. Waldron, P. Matsushima, P. R. Rosteck, M. C. Broughton, J. Turner, K. Madduri, K. P. Crawford, D. J. Merlo, and R. H. Baltz, “Cloning and analysis of the spinosad biosynthetic gene cluster of Saccharopolyspora spinosa 1,” vol. 8, pp. 487–499, 2001.

[81] E. P. Patallo, G. Blanco, C. Fischer, a F. Brana, J. Rohr, C. Mendez, and J. a Salas, “Deoxysugar methylation during biosynthesis of the antitumor polyketide elloramycin by Streptomyces olivaceus. Characterization of three methyltransferase genes.,” *The Journal of biological chemistry*, vol. 276, no. 22, pp. 18765–74, Jun. 2001.

[82] J. Ahlert, E. Shepard, N. Lomovskaya, E. Zazopoulos, A. Staffa, B. O. Bachmann, K. Huang, L. Fonstein, A. Czisny, R. E. Whitwam, C. M. Farnet, and J. S. Thorson, “The calicheamicin gene cluster and its iterative type I enediyne PKS.,” *Science (New York, N.Y.)*, vol. 297, no. 5584, pp. 1173–6, Aug. 2002.

[83] A. Trefzer, D. Hoffmeister, E. Künzel, S. Stockert, G. Weitnauer, L. Westrich, U. Rix, J. Fuchser, K. U. Bindseil, J. Rohr, and A. Bechthold, “Function of glycosyltransferase genes involved in urdamycin A biosynthesis.,” *Chemistry & biology*, vol. 7, no. 2, pp. 133–42, Feb. 2000.

[84] H. Nakano, Y. Matsuda, K. Ito, S. Ohkubo, M. Morimoto, and F. Tomita, “Gilvocarcins, new antitumor antibiotics. 1. Taxonomy, fermentation, isolation and biological activities.,” *The Journal of antibiotics*, vol. 34, no. 3, pp. 266–70, Mar. 1981.

[85] H. Schmitz, K. E. Crook, and J. A. Bush, “Hedamycin, a new antitumor antibiotic. I. Production, isolation, and characterization.,” *Antimicrobial agents and chemotherapy*, vol. 6, pp. 606–12, Jan. 1966.

[86] K. Ichinose, D. J. Bedford, D. Tornus, A. Bechthold, M. J. Bibb, W. P. Revill, H. G. Floss, and D. A. Hopwood, “The granaticin biosynthetic gene cluster of Streptomyces violaceoruber Tü22: sequence analysis and expression in a heterologous host.,” *Chemistry & biology*, vol. 5, no. 11, pp. 647–59, Nov. 1998.

[87] S. Gräslund, J. Sagemark, H. Berglund, L.-G. Dahlgren, A. Flores, M. Hammarström, I. Johansson, T. Kotenyova, M. Nilsson, P. Nordlund, and J. Weigelt, “The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins.,” *Protein expression and purification*, vol. 58, no. 2, pp. 210–21, Apr. 2008.

[88] C. Aslanidis and P. J. de Jong, “Ligation-independent cloning of PCR products (LIC-PCR).,” *Nucleic acids research*, vol. 18, no. 20, pp. 6069–74, Oct. 1990.

[89] R. K. Knaust and P. Nordlund, “Screening for soluble expression of recombinant proteins in a 96-well format.,” *Analytical biochemistry*, vol. 297, no. 1, pp. 79–85, Oct. 2001.

[90] F. W. Studier, “Protein production by auto-induction in high density shaking cultures.,” *Protein expression and purification*, vol. 41, no. 1, pp. 207–34, May 2005.

[91] C. Zhang, B. R. Griffith, Q. Fu, C. Albermann, X. Fu, I.-K. Lee, L. Li, and J. S. Thorson, “Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions.,” *Science (New York, N.Y.)*, vol. 313, no. 5791, pp. 1291–4, Sep. 2006.

[92] C. Zhang, B. R. Griffith, Q. Fu, C. Albermann, X. Fu, I.-K. Lee, L. Li, and J. S. Thorson, “Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions.,” *Science (New York, N.Y.)*, vol. 313, no. 5791, pp. 1291–4, Sep. 2006.

[93] C. Zhang, C. Albermann, X. Fu, and J. S. Thorson, “The in vitro characterization of the iterative avermectin glycosyltransferase AveBI reveals reaction reversibility and sugar nucleotide flexibility.,” *Journal of the American Chemical Society*, vol. 128, no. 51, pp. 16420–1, Dec. 2006.

[94] C. Zhang, Q. Fu, C. Albermann, L. Li, and J. S. Thorson, “The in vitro characterization of the erythronolide mycarosyltransferase EryBV and its utility in macrolide diversification.,” *Chembiochem : a European journal of chemical biology*, vol. 8, no. 4, pp. 385–90, Mar. 2007.

[95] G. J. Williams, C. Zhang, and J. S. Thorson, “Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution.,” *Nature chemical biology*, vol. 3, no. 10, pp. 657–62, Oct. 2007.

[96] S. P. Sim, B. Gatto, C. Yu, a a Liu, T. K. Li, D. S. Pilch, E. J. LaVoie, and L. F. Liu, “Differential poisoning of topoisomerases by menogaril and nogalamycin dictated by the minor groove-binding nogalose sugar.,” *Biochemistry*, vol. 36, no. 43, pp. 13285–91, Oct. 1997.

[97] E. H. Fischer, A. B. Kent, E. R. Snyder, and E. G. Krebs, “THE REACTION OF SODIUM BOROHYDRIDE WITH MUSCLE PHOSPHORYLASE,” *J. Am. Chem. Soc*, vol. 80, no. 11, pp. 2906–2907, 1958.

[98] G. E. Means and R. E. Feeney, “Reductive alkylation of amino groups in proteins.,” *Biochemistry*, vol. 7, no. 6, pp. 2192–201, Jun. 1968.

[99] T. S. Walter, C. Meier, R. Assenberg, K.-F. Au, J. Ren, A. Verma, J. E. Nettleship, R. J. Owens, D. I. Stuart, and J. M. Grimes, “Lysine methylation as a routine rescue strategy for protein crystallization.,” *Structure (London, England: 1993)*, vol. 14, no. 11, pp. 1617–22, Nov. 2006.

[100] Y. Kim, P. Quartey, H. Li, L. Volkart, C. Hatzos, C. Chang, B. Nocek, M. Cuff, J. Osipiuk, K. Tan, Y. Fan, L. Bigelow, N. Maltseva, R. Wu, M. Borovilos, E. Duggan, M. Zhou, T. A. Binkowski, R. Zhang, and A. Joachimiak, “Large-scale evaluation of protein reductive methylation for improving protein crystallization.,” *Nature methods*, vol. 5, no. 10, pp. 853–4, Oct. 2008.

[101] T. Bergfors, “Seeds to crystals,” *Journal of Structural Biology*, vol. 142, no. 1, pp. 66–76, Apr. 2003.

[102] M. G. Rossmann, D. Moras, and K. W. Olsen, “Chemical and biological evolution of a nucleotide-binding protein,” *Nature*, vol. 250, no. 5463, pp. 194–199, Jul. 1974.

[103] D. N. Bolam, S. Roberts, M. R. Proctor, J. P. Turkenburg, E. J. Dodson, C. Martinez-Fleites, M. Yang, B. G. Davis, G. J. Davies, and H. J. Gilbert, “The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 13, pp. 5336–41, Mar. 2007.

[104] Y. Hu, L. Chen, S. Ha, B. Gross, B. Falcone, D. Walker, M. Mokhtarzadeh, and S. Walker, “Crystal structure of the MurG: UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 3, p. 845, Feb. 2003.

[105] E. a Isiorho, H.-W. Liu, and A. T. Keatinge-Clay, “Structural Studies of the Spinosyn Rhamnosyltransferase, SpnG.,” *Biochemistry*, vol. 51, no. 6, pp. 1213–22, Feb. 2012.

[106] F. Wang, M. Zhou, S. Singh, C. A. Bingman, J. S. Thorson, and G. N. Phillips Jr., “Crystal Structure of SsfS6, Streptomyces sp. SF2575 glycosyltransferase - To be Published,” 2012.

[107] L. V. Modolo, L. Li, H. Pan, J. W. Blount, R. a Dixon, and X. Wang, “Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids.,” *Journal of molecular biology*, vol. 392, no. 5, pp. 1292–302, Oct. 2009.

[108] C. Zhang, E. Bitto, R. D. Goff, S. Singh, C. a Bingman, B. R. Griffith, C. Albermann, G. N. Phillips, and J. S. Thorson, “Biochemical and structural insights of the early glycosylation steps in calicheamicin biosynthesis.,” *Chemistry & biology*, vol. 15, no. 8, pp. 842–53, Aug. 2008.

[109] A. M. Mulichak, W. Lu, H. C. Losey, C. T. Walsh, and R. M. Garavito, “Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands.,” *Biochemistry*, vol. 43, no. 18, pp. 5170–80, May 2004.

[110] T. Bililign, C. Hyun, J. S. Williams, A. M. Czisny, and J. S. Thorson, “The Hedamycin Locus Implicates a Novel Aromatic PKS Priming Mechanism,” vol. 11, pp. 959–969, 2004.

[111] J. Härle, S. Günther, B. Lauinger, M. Weber, B. Kammerer, D. L. Zechel, A. Luzhetskyy, and A. Bechthold, “Rational design of an aryl-C-glycoside catalyst from a natural product O-glycosyltransferase.,” *Chemistry & biology*, vol. 18, no. 4, pp. 520–30, Apr. 2011.

[112] K. Ichinose, T. Tagauchi, Y. Ebizuka, and D. A. Hopwood, “Biosynthetic Gene Clusters of Benzoisochromanequinone Antibiotics in Streptomyces spp. – Identification of Genes Involved in Post-PKS Tailoring Steps–,” *Actinomycetologica*, vol. 12, no. 2, pp. 99–109, 1998.

[113] S. a Borisova, L. Zhao, C. E. Melançon III, C.-L. Kao, and H.-W. Liu, “Characterization of the glycosyltransferase activity of desVII: analysis of and implications for the biosynthesis of macrolide antibiotics.,” *Journal of the American Chemical Society*, vol. 126, no. 21, pp. 6534–5, Jun. 2004.

[114] S. a Borisova and H.-W. Liu, “Characterization of glycosyltransferase DesVII and its auxiliary partner protein DesVIII in the methymycin/picromycin biosynthetic pathway.,” *Biochemistry*, vol. 49, no. 37, pp. 8071–84, Sep. 2010.

[115] Y. Yuan, H. S. Chung, C. Leimkuhler, C. T. Walsh, D. Kahne, and S. Walker, “In vitro reconstitution of EryCIII activity for the preparation of unnatural macrolides.,” *Journal of the American Chemical Society*, vol. 127, no. 41, pp. 14128–9, Oct. 2005.

[116] M. C. Moncrieffe, M.-J. Fernandez, D. Spiteller, H. Matsumura, N. J. Gay, B. F. Luisi, and P. F. Leadlay, “Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII.,” *Journal of molecular biology*, vol. 415, no. 1, pp. 92–101, Jan. 2012.

[117] W. Lu, C. Leimkuhler, G. J. Gatto, R. G. Kruger, M. Oberthür, D. Kahne, and C. T. Walsh, “AknT is an activating protein for the glycosyltransferase AknS in L-aminodeoxysugar transfer to the aglycone of aclacinomycin A.,” *Chemistry & biology*, vol. 12, no. 5, pp. 527–34, May 2005.

[118] D. Martinez, L. F. Larrondo, N. Putnam, M. D. S. Gelpke, K. Huang, J. Chapman, K. G. Helfenbein, P. Ramaiya, J. C. Detter, F. Larimer, P. M. Coutinho, B. Henrissat, R. Berka, D. Cullen, and D. Rokhsar, “Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78.,” *Nature biotechnology*, vol. 22, no. 6, pp. 695–700, Jun. 2004.

[119] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen, “SignalP 4.0: discriminating signal peptides from transmembrane regions.,” *Nature methods*, vol. 8, no. 10, pp. 785–6, Jan. 2011.

[120] W. Montfort, J. E. Villafranca, A. F. Monzingo, S. R. Ernst, B. Katzin, E. Rutenber, N. H. Xuong, R. Hamlin, and J. D. Robertus, “The three-dimensional structure of ricin at 2.8 A.,” *The Journal of biological chemistry*, vol. 262, no. 11, pp. 5398–403, Apr. 1987.

[121] C. K.-M. Chen, N.-L. Chan, and A. H.-J. Wang, “The many blades of the β-propeller proteins: conserved but versatile.,” *Trends in biochemical sciences*, vol. 36, no. 10, pp. 553–61, Oct. 2011.

[122] J. M. Dunwell, a Culham, C. E. Carter, C. R. Sosa-Aguirre, and P. W. Goodenough, “Evolution of functional diversity in the cupin superfamily.,” *Trends in biochemical sciences*, vol. 26, no. 12, pp. 740–6, Dec. 2001.

[123] J. M. Dunwell, A. Purvis, and S. Khuri, “Cupins: the most functionally diverse protein superfamily?,” *Phytochemistry*, vol. 65, no. 1, pp. 7–17, Jan. 2004.

[124] S. Andersen, I. Lundt, and J. Marcussen, “1,5-Anhydro-D-Fructose: Stereoselective Conversions to 1,5-Anhydroalditols and Deoxy/Amino Substituted Analogues,” *Journal of Carbohydrate Chemistry*, vol. 19, no. 6, pp. 717–725, 2000.

[125] D. Cobessi, R. Dumas, V. Pautre, C. Meinguet, J.-L. Ferrer, and C. Alban, “Biochemical and structural characterization of the Arabidopsis bifunctional enzyme dethiobiotin synthetase-diaminopelargonic acid aminotransferase: evidence for substrate channeling in biotin synthesis.,” *The Plant cell*, vol. 24, no. 4, pp. 1608–25, Apr. 2012.

[126] S. J. J. Brouns, T. R. M. Barends, P. Worm, J. Akerboom, A. P. Turnbull, L. Salmon, and J. van der Oost, “Structural insight into substrate binding and catalysis of a novel 2-keto-3-deoxy-D-arabinonate dehydratase illustrates common mechanistic features of the FAH superfamily.,” *Journal of molecular biology*, vol. 379, no. 2, pp. 357–71, May 2008.