Chem Soc Rev

Glycosyltransferases: mechanisms and applications in natural product development

Journal:	*Chemical Society Reviews*
Manuscript ID:	CS-REV-07-2015-000600
Article Type:	Review Article
Date Submitted by the Author:	31-Jul-2015

Complete List of Authors:	Liang, Dongmei; School of Chemical Engineering and Technology, Department of Pharmaceutical Engineering Liu, Jiaheng; School of Chemical Engineering and Technology, Department of Pharmaceutical Engineering Wu, Hao; School of Chemical Engineering and Technology, Department of Pharmaceutical Engineering Wang, Binbin; School of Chemical Engineering and Technology, Department of Pharmaceutical Engineering Zhu, Hongji; School of Chemical Engineering and Technology, Department of Pharmaceutical Engineering Qiao, Jianjun; School of Chemical Engineering and Technology, Tianjin University,

Glycosyltransferases: mechanisms and applications in natural product development

Dong-Mei Liang‡, Jia-Heng Liu‡, Hao Wu, Bin-Bin Wang, Hong-Ji Zhu and Jian-Jun Qiao*

Department of Pharmaceutical Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China;

Key Laboratory of Systems Bioengineering, Ministry of Education Tianjin 300072, China;

SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering , Tianjin 300072, China.

‡ Dong-Mei Liang and Jia-Heng Liu contributed equally to this work.

* *Corresponding author.* Tel: +86-022-8740-2107; Fax: +86-022-8740-2107. E-mail address: jianjunq@tju.edu.cn

Abstract

Glycosylation reactions mainly catalyzed by glycosyltransferases (Gts), occur almost everywhere in biosphere, and always play crucial roles in vital processes. In order to understand the full potential of Gts, the chemical and structural glycosylation mechanisms are systematically summarized in this review, including some new outlooks in inverting/retaining mechanisms and the overview of GT-C superfamily proteins as a novel Gt fold. Some special features of glycosylation and the evolutionary studies on Gts are also discussed to help us better understand the function and application potential of Gts. Natural product (NP) glycosylation and related Gts which play important roles in new drug development are emphasized in this paper. The recent advances of glycosylation pattern (particularly the rare C- and S-glycosylation), reversability, iterative catalysis and protein auxiliary of NP Gts are all summed up comprehensively. This review also presents the application of NP Gts and associated studies on synthetic biology, which may further broaden the mind and bring wider application prospects.

1. Introduction

Glycosylation reactions are widespread in nature, and involved in almost all vital processes. In general, glycosylation represents the saccharide polymerizations or the conjunctions of saccharides with other biomolecules including proteins, lipids, nucleic acids and natural small molecules (mainly referring to secondary metabolites). Glycosylated compounds directly exert a wide range of functions, including energy storage, maintenance of cell structural integrity, information storage and transfer,

molecular recognition, cell-cell interaction, cellular regulation, immune response, virulence and chemical defense etc. Glycoylation usually plays important roles in their physico-chemical properties and functional performance. Notably, the crucial contribution of glycosylation to natural products (NPs) provides numerous possibilities for new bioactive substance exploitation. Glycosylation also can promote the evolution of biological form and function,[1] especially for glycans, whose structural diversity may be crucial for evolution and speciation of different organisms.[2, 3] In prokaryotes, most of glycosylation reactions occur in the cytoplasm, at the plasma membrane and in periplasm, while in eukaryotes, they generally take place in the Golgi apparatus and endoplasmic reticulum.[4]

Glycosylated compounds are the most structurally diverse biomolecules, and their biosynthesis needs quite complex biological processes orchestrated by many enzyme systems. Current estimates suggest that up to 1% of the gene products of each organism are involved in glycosylation.[4, 5] Most glycosylation reactions are catalyzed by glycosyltransferases (Gts) (EC 2. 4. x. y), which mediate the region- and stereo-specific glycosidic bond formations between sugar moieties and a variety of important biomolecules. They can use diverse activated sugar donors, such as nucleotide-sugars, lipid-phospho sugar donors and sugar-1-phosphates. Currently, the universally accepted classification (GT family system) is mainly based on sequence similarity collected in the Carbohydrate Active Enzyme database (CAZy, http://www.cazy.org/).[6] Interestingly, despite the dramatic function differentiation and the great sequence diversity, the chemical reaction mechanisms (inversion or retention)

and the 3D structures (GT-A, GT-B and GT-C folds) of Gts do not show much diversification, shedding some light on the more rational categorization. Especially, the glycosylation on NPs has a series of attractive characteristics and promising applications in new drug development, making it a hotspot in NP biosynthesis and modification.

In this review, we firstly overview the glycosylation mechanisms from the perspective of chemical essence and protein structure respectively, with some new outlooks that may be a beneficial supplement to the classical theories. Moreover, the GT-C proteins are emphasized due to its significance and the rare discussion in former reviews. And then several interesting characteristics which may provide some new ideas about the functional studies of Gts are summarized, including substrate specificity and regioselectivity, oligomerization and protein complexes, and chaperones. Some evolutionary viewpoints on the basis of our former works are also discussed. Followed that, the paper concentrates on recent developments in NP glycosylations and Gts, especially focuses on the special features that are potentially useful for future applications. The representative studies on glycosylation pattern (particularly the rare C- and S-glycosylation), reversability, iterative catalysis, protein auxiliary, substrate specificity and the applications of NP Gts are all summed up systematically. Finally, more advances in associated studies on synthetic biology are included to inspire researchers to explore more new ideas on glycobiology.

2. Catalytic mechanisms of glycosylation

Glycochemists and glycobiologists have devoted to the studies on function and

catalytic mechanisms of Gts for decades, and made many efforts for rational classification. CAZy database is the most popolar Gt database. Until now, there are over 195834 Gt sequences comprising 97 families (three of which have been deleted) in CAZy database and 3519 non-classified sequences have been identified. Also, more than 150 Gt structures (24 involved in NP biosynthesis (Table 1)) were kept in the PDB database (http://www.rcsb.org/pdb/home/home.do). Each GT family contains at least one protein with an experimentally proved function, and then recruited by sequences with significant similarities. GT members in the same family are expected to have a similar 3D structure, as well as the stereoselectivity (Table 2). Though this classification can reveal the relationship between Gt function and evolution to some extent, some cases are still difficult to explain, such as sialyltransferases which unexpectedly fall into five different GT families (GT-29, -38, -42, -52, and -80), even adopting different stereochemical conformations and 3D structures. In order to better understand and investigate glycosylation reactions, in-depth discussion on glycosylation mechanisms (chemical and structural) is quite necessary.

2.1 The basic chemical glycosylation mechanisms

Based on the stereochemical difference of the glycosylation, the reaction mechanisms can be designated as inverting (e.g. NDP-α-sugar→β-glycoside) or retaining (e.g. NDP-α-sugar→α-glycoside) (Fig. 1 (a)). It has been proved that the stereochemistry of glycosylation was not directly tied to the overall fold of Gts (Table 2). The Gt structure provides suitable catalytic sites and special microenvironment for glycosylation reactions, which can properly position and orient sugar and acceptor

donors. Combined with these contributions, the region- and stereo- specific glycosylation reactions are completed.

Although the reaction mechanisms are mainly elucidated through structure analysis, theoretical studies which can provide more detailed information at the atomistic level are still necessary. In recent years, different glycosylation mechanisms have been gradually clarified using hybrid quantum mechanics/molecular mechanics (QM/MM) methods, together with associated experimental data. Hybrid QM/MM methods are now regularly used to investigate the catalytic mechanism of various enzymes. These methods generally model enzymatic reactions through quantum mechanics methods for the calculation of the electronic structure of the active site models, and treat the remaining enzyme environment with faster molecular mechanics methods.

2.1.1 The inverting glycosylation

Inverting Gts are supposed to utilize a direct displacement S_N2-like mechanism (Fig. 1(b)), supporting by both the theoretical[30-33] and experimental studies. An oxocarbenium-ion transition state (TS) forms with the help of a catalytic base usually provided by an active-site side chain (such as Asp, Glu or His) from Gts. The catalytic base abstracts a proton from OH-group of the acceptor, facilitating nucleophilic attack at the sugar anomeric C1, forming a glycosidic bond between the sugar donor and the acceptor with inversion of the configuration at C1. The theoretical works[30, 33] indicated that the nucleophilic addition and the breaking of the glycosidic bond were nearly simultaneous, accompanied by the proton transfer. During the formation of the

TS, the anomeric C1 atom moves towards the nucleophilic oxygen accompanied with the rotation of the diphosphate group, which is conducive to the glycosylation reaction. Structural studies showed that the catalytic bases usually located near the acceptor OH-group. Classically, the catalytic His residue in some GT-B members can form a hydrogen bond with an Asp residue, balancing the charge on the His after proton abstraction. In general, the negative charge on the phosphate group can be stabilized by a divalent metal ion (generally Mn^{2+} or Mg^{2+}) for most GT-A proteins or positive amino acids/helix dipole for GT-B proteins.

In addition, some researchers also proposed other possibility such as S_N1-like mechanism[34] (Fig. 1(c)), which were supported and elucidated by further experimental and theoretical studies recently[35]. In this mechanism, the cleavage of the glycosidic bond took place first, followed by the formation of an intimate oxocarbenium-phosphate ion pair. Due to the lack of candidate catalytic residue, the catalytic role of β-phosphate was proposed. Besides, another study[36] supposed that the α-phosphate (scarcely directly participates in catalysis) of UDP-GlcNAc served as the proton acceptor with the help of an essential Lys residue during the glycosylation catalyzed by human *O*-GluNAc transferase (OGT).

2.1.2 The retaining glycosylation

The exact reaction mechanism of retaining glycosylation is still a matter of much debate. Evidence is mounting that there is most likely not only one uniform mechanism for retaining Gts. A few of studies[37-39] supported the double-displacement mechanism (Fig. 1(d)) involving the formation of a covalent sugar–enzyme

intermediate, especially for the GT6 members which have a carboxylic residue in the proper position as a candidate nucleophile. This mechanism proposes that the sugar moiety first binds to a proper part of the Gt with an inverting configuration at sugar C1. Then it is transferred to the acceptor with the C1 atom reverting to its original configuration.

For most retaining Gts, the catalytic base could not be identified in the catalytic site, thus an alternative internal return (S_Ni-like) mechanism (Fig. 1(e)) was proposed. In this mechanism, the acceptor OH-group nucleophilic attacks the sugar anomeric C1 atom on the same side that the sugar group leaves the donor. The reaction involves the formation of an oxocarbenium ion TS that shielded on one face of the reaction center by the Gt, consequently protecting against nucleophilic attack from the opposite face and resulting in retention of C1 configuration.[40-43] Theoretical studies[42] indicated that the cleavage of the UDP–sugar bond took place first, followed by the formation of the donor-acceptor bond, allowing the formation of an intimate ion-pair intermediate. Both of the experimental[40, 44] and theoretical[43] studies suggested that the leaving phosphate group could function as the catalytic base to deprotonate the acceptor OH-group.

The controversy of retaining mechanisms has persisted for years. LgtC was initially recognized to prefer the double-displacement mechanism.[45] However, the subsequent theoretical studies revealed that it might be more inclined to adopt a dissociative S_Ni mechanism.[43] Furthermore, recent studies on GT6 retaining Gts[38, 46-48] indicated that the S_Ni-like and the double-displacement mechanisms were both

feasible in glycosylation catalyzed by theses special Gts.

2.2 Structure-based glycosylation mechanisms

Although the overall fold of Gts cannot determine the stereochemistry, it provides the proper microenvironment for glycosylation reactions. This "reaction container" imposes a correct positioning and orientation of substrates in the Michaelis complex, and involves all influencing factors of the reaction processes. Therefore, it is necessary to investigate glycosylation mechanisms from structural perspective. With the rapid development of protein crystallization and analytical techniques, the increasing numbers of Gt structures have been solved. In the past three years, ~50 new Gt structures were identified (Table 3), which occupy one third of total crystallized Gts, providing much possibility to discuss structure-function relationships and glycosylation mechanisms based on structure.

As mentioned above, most of natural Gts adopt three main structural topology designated as GT-A, GT-B and GT-C superfamilies. GT-A members possess a single domain fold (Fig. 2(a)) which composes of a seven stranded β-sheet core flanked by α-helices and a small anti-parallel β-sheet bridged via a DxD motif (or its variants, e.g. TDD, EDD, DxH, etc). GT-B members are comprised of two distinct Rossmann-like domains of several parallel β-sheet linked to α-helices. The N- and C-terminal domains are connected by a linker region, forming an interdomain cleft and creating the active sites. In the majority of GT-B structures, the C-terminal domain possesses a kinked α helix that extends out to interact with the N-terminal domain (Fig. 2(b)). Both the conventional GT-A or GT-B core can further combine with other domains,

including membrane-anchoring regions[79, 80], tetratricopeptide repeat (TPR) units for the interactions with other proteins[81, 82], all α-domain fold of the HMW1C-like proteins involved in the formation of a unique groove with potential to accommodate the acceptor protein[83], the SH3 domains[84] etc. Additionally, canonical GT-A or GT-B structures may also contain some special motifs, which are usually crucial for Gt functions. For example, a unique insert in the C-terminal domain of yeast glycogen synthase-2 (GT-3) forms a pair of long helices extending away from the core fold to construct the majority of the intersubunit interface of the tetramer[85]; a novel domain in the N-terminal domain of lipooligosaccharide sialyltransferase from NST (*Neisseria meningitides* serotype L1) provides hydrogen bonds and salt bridges across the dimer interface to consolidate the protein structure (GT-52)[80]; an unusual hairpin loop extension in *Sc*Mnn9 structure probably functions as a molecular ruler for restriction the length of a mannose backbone (GT-62)[69]; a protrusion domain in *Lp*GT (*Legionella pneumophila* glucosyltransferase) structure plays an important role in UDP-glucose binding (Gt-88)[86]; and a novel fold between the N- and C-terminal domains of a human OGT was characterized to pack exclusively against the C-terminal domain (GT-41)[81]. Interestingly, some non-Gt enzymes also adopt GT-A or GT-B folds.[87] We took the LgtC (GT-A fold, PDB: 1SS9) and GtfA (GT-B fold, PDB: 1PN3) (Fig. 2) as the query structures to search the structural similar proteins. The results showed that some enzymes responsible for transferring phosphorus-containing groups (EC: 2. 7. x. y) such as molybdenum cofactor biosynthesis protein MobA (e.g. 1E5K, 1HJJ), 3-deoxy-manno-octulosonate

cytidylyltransferase (e. g. 1H7T), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (e. g. 4KT7), ADP-glucose pyrophosphorylase (e. g. 1YP4), UDP-N-acetylglucosamine pyrophosphorylase (e. g. 1HV9, 4BMA) and inositol-1-phosphate cytidylyltransferase (e. g. 4JD0) (Fig. 4(a)) resembled the GT-A fold with a >9.0 Z-score, and UDP-N-acetylglucosamine 2-epimerase (e. g. 3OT5) (Fig. 4(b)) displayed an overall topological similarity (Z-score is 22.2) to GT-B proteins and bound the NDP-sugar in the similar region.

GT-C fold is predicted by multivariate sequence analysis[88] and identified in several recent structures, including EmbC (GT-51) and eight STT3/AglB/PglB proteins (GT-66). The GT-C superfamily contains a wide range of Gts that have 8–13 predicted transmembrane (TM) segments, and possesses a DxD (ExD, DxE, DDx, or DEx) motif in the first extracellular/lumenal loop that may be essential for catalysis. This motif is often followed by several hydrophobic amino acids as a part of the same loop.[29]

Additionally, Zhang *et al.* determined the crystal structures of an uncharacterized domain of unknown function (DUF1792) and confirmed that it functioned as glucosyltransferase participating in the biosynthesis of bacterial O-glycans, which has a completely different structure compared with the known Gts and was designated as a GT-D fold. It also contains a highly conserved metal-binding (DxE) motif.[51] GT51 also has a special structure which shows no similarity to any of above folds. Although its amino acid sequence is totally different with bacteriophage λ lysozyme, there is a remarkable similarity in secondary structure and key active site between the two

proteins. Therefore, this structure was defined as "lysozyme-type".[89, 90]

Regardless of individual topology, glycosylation reaction usually follows a sequential bi–bi mechanism. In this mechanism, the sugar donor and the aglycon are bound sequentially, followed by the sugar transfer. The glycosylated product is then released, followed by the nucleotide moiety release. The reactions inevitably involved the substrate recognization and binding which decided by some conserved regions in Gt structures, as well as the change of protein conformation. The mechanisms will be discussed in detail later according to the individual structure.

2.2.1 The GT-A Gts

In GT-A structure, the NDP-sugar-binding region contains a conserved DxD motif for divalent metal ion (usually Mn^{2+} or Mg^{2+}) binding. The metal ion can neutralize the negative charge on the phosphate group and induce the conformational change during catalysis. The Mn^{2+} was also shown to participate in formation and stabilization of the TS complex[91], or accelerate the hydrolysis of NDP-sugar presumably through electrophilic catalysis[92]. Both Asp residues of this motif in the majority of retaining Gts interact with Mn^{2+}, while in most of inverting Gts, only one of the Asp residues interacts with Mn^{2+}. Otherwise, a few of studies found that some DxD motifs were very important to enzyme activity, but not related to bivalent cation.[93] This particular motif always locates in a short loop linking two β-strands at the end of the nucleotide binding domain.[94] However, some GT-A enzymes lacking the DxD motif are commonly metal-ion-independent, including the members of GT-14, GT-29, GT-42 and the bacterial GT6 members involved in the synthesis of

Histo-blood group antigens[58].

As mentioned above, glycosylation mechanism for GT-A proteins follows a sequential bi–bi mechanism. The metal ion, if required for catalysis, binds initially to Gt interacting with Asp residue(s) in DxD motif and with oxygen atoms from the α-/β-phosphate of UDP. And then the sugar donor binds to a wide open active site cavity. The binding of metal ion and sugar donor is accompanied by local conformational changes of at least one flexible loop, consequently covering the sugar donor and creating the acceptor binding sites, simultaneously minimizing potential hydrolytic effects in a closed conformation. The flexible loop(s) usually locate in the vicinity of the sugar donor binding sites. After the sugar transfers to the acceptor, the glycosylated product is ejected, followed by the conformational reversal to the open form, accompanied with the release of metal ion and nucleotide (Fig. 4).

2.2.2 The GT-B Gts

The GT-B members contain two Rossmann-like folds. The N- and C-terminal domains are responsible for acceptor and sugar donor binding respectively (Fig. 3(b)). The C-terminal domain is always more conserved than the N-terminal domain. Most of the binding sites locate in the deep, interdomain cleft between the two domains. During the reaction, GT-B protein also needs to experience a series of conformational changes like GT-A protein (Fig. 4). Binding of the sugar donor triggers the change from open to closed conformation (usually the N-terminal domain rotates ~10° toward the C-terminal domain, while some special members might need >20° shift[95]), allowing the pyrophosphate to interact with both N- and C- terminal domains or

introducing direct interactions across the cleft, which may further stabilize the catalytically active conformation. The conformational closure also alters the shape of the binding pocket, accompanied with formation of the actual acceptor binding sites, which are stabilized by entropic effects, consistent with the induced-fit mechanism.

GT-B proteins are generally metal-ion-independent and lack related conserved motifs. Although it was reported that divalent cations were required for full activity of BGT (T4 phage β-Glucosyltransferase), their position in BGT structures indicated that the most probable role was not activating catalysis or stabilizing the leaving phosphate group, but facilitating the product release.[96] Recent studies on POFUT2 further elucidated the role of metal ions in product release.[74] Notably, despite the poor total sequence identity, many GT families adopting GT-B fold possess a conserved motif (commonly named Gly-rich motif due to containing several Gly residues in GT1 members) at the turn preceding the Cα4 helix which was presumed as an essential segment for NDP-sugar binding (Fig. 5). Comparative analysis indicates that these GT-B proteins using NDP-sugars as donors possess not only similar overall topology but also homologous donor binding manners, suggesting a magic evolutionary property (Fig. 5).

2.2.3 The GT-C Gts

To date, all of the identified GT-C proteins are inverting Gts (Table 2). The structure of GT-C members commonly comprises two domains: N-terminal TM domain and C-terminal globular domain with Gt activity. The TM domain is necessary both for substrate binding and catalysis. A divalent cation is usually

indispensable in glycosylation catalyzed by GT-C proteins. It may exert dual-function that orienting the acidic residues to interact with the acceptor and stabilizing the leaving group.[97] To date, most of GT-C proteins with solved crystal structures are in GT-66. These GT66 members are the catalytic subunit of oligosaccharyltransferase (OST) involved in protein N-glycosylation, and conserved in the three kingdoms, designated as STT3 in eukaryotes, AglB in archaea and PglB in eubacteria. Despite the poor sequence similarity, all STT3/AglB/PglB proteins contain a conserved WWDYG motif, in which the Asp residue is thought to be a catalytic site. Another two short motifs named DK motif (DxxK) and MI motif (MxxI) were also identified, which locate at spatially equivalent positions close to the WWDYG motif and were proposed to constitute the active site pocket with the WWDYG motif. All PglB and some AglB contain both DK and MI motifs, while the remaining AglB and all STT3 contain the DK motif only,[70] indicating two types of Ser/Thr-binding pockets. Additionally, there is a variable DxD motif in the TM region. Mutational studies indicated that the latter Asp residue might be essential for yeast growth. It has been also proved that the DK motif together with the DxD motif probably formed the binding site of the pyrophosphate group of the lipid-linked oligosaccharide, through a transient bound with Mn^{2+} or Mg^{2+}.[98]

All STT3/AglB/PglB proteins possess a central core structure, consisting of several α-helices and one three-stranded antiparallel β-sheet (Fig. 2(c)). The conserved WWDYG and DK/MI motifs all locate in this part.[72] Different proteins additionally contain other unique structural units, including insertion, peripheral 1 and

peripheral 2 (Fig. 2(c)). The structural studies on bacterial PglB revealed that the engagement or disengagement of EL5 (external loop in TM domain) was critical in the glycosylation process. Once the sequon substrate is bound, the C-terminal half of EL5 immediately pins the acceptor peptide against the periplasmic domain, accompanied by the formation of the catalytic site. Upon the formation of glycosidic bond, the nascent saccharides are tightly nestled on PglB, causing steric strain that can be released by disengagement of EL5, allowing the product release.[97]

3. Special reaction characteristics of Gts

3.1 Substrate specificity and regioselectivity

In general, Gts exhibit strict substrate specificity and regioselectivity that is mainly determined by some small motifs or several residues. Human GTA and GTB that produce ABO(H) blood group antigens are highly homologous enzymes differing in only four residues (R176G, G235S, L266M, and G268A).[99] The crystal studies showed that the latter two residues were essential for differentiating the donor substrates. Actually, only residue 266 was proved to be directly concern with the substrate specificity.[100] Meanwhile, during the study on domain swapping of quercetin Gts, an individual Asn residue (N142 in UGT74F1) was proved to significantly affect the regiospecificity toward the 4'-OH (Fig. 6(a)).[101] In the last decade, many studies on altering the substrate specificity or regiospecificity of Gts through site-directed mutation have been performed.[102-109] However, it is usually difficult to clarify the crucial sites for substrate specificity and regioselectivity. So protein crystallization[102, 109, 110], homologous modeling[104, 111] and other screening techniques[112] are introduced

to guide the site selection. Meanwhile, some high throughput screening methods have been explored to obtain the ideal mutants more efficiently.[103, 108, 113]

Nevertheless, many Gts with flexible substrate specificity or regiospecificity are also reported, such as a multifunctional sialyltransferase from *Pasteurella multocida* which can utilize several different sugar donors and acceptors[114], and UGT71G1 from *Medicago truncatula* which can transfer glucose to each of the five OH-groups of the flavonol quercetin[102]. These characteristics are most pronounced in NP Gts and made them potential tools for synthesizing diverse glycosides.

Furthermore, some Gts do not recognize the overall structure of the substrates, especially those for protein glycosylation. A recent study showed that a 19-amino-acid peptide of AIDA (adhesin-involved-in-diffuse-adherence), an autotransporter in *E. coli*, was necessary and sufficient for recognization and glycosylation by its associated heptosyltransferase (Aah). Aah may recognize a "short β-strand-short acceptor loop–short β-strand" motif formed by any sequence. The motif forms a loop starting with a Ser residue as the glycosylation site[115]. This characteristic is also possessed by other protein O-Gts, such as POFUT2[74] which can specifically recognize unique 3D structure of thrombospondin type 1 repeats, and the Gt for PSM (the porcine submaxillary mucin) tandem repeat glycosylation[116] which binds the acceptor peptide with a β-like conformation.

3.2 Gt oligomers and complexes

Gts are a special kind of enzymes due to their complex nature and various ways of functioning. Most Gts can work in monomer, nonetheless, a few particular Gts

must function through forming oligomers. Some functional complexes are even formed by different Gts to exert the collaborative effects.

The dimer is the major oligomer manner in Gts. The domain-swapped homodimer of lipooligosaccharide sialyltransferase from NST plays a potential acceptor activating role in regulating Gt function.[80] The dimer of yeast OST containing nine subunits is proposed to be required for effective association with the translocon dimer and for its allosteric regulation during co-translational glycosylation.[117] GM2 synthases likewise form very stable functional homodimers through disulfide bonds, resulting in an antiparallel orientation of the catalytic domains.[118] Disulfide bonds also play some other important roles in Gt structures. For instance, the two disulfide bonds in human FucT (fucosyltransferase) III can bring the N and C termini of the catalytic domain close together in space[119], and the disulfide bonds in *Ce*POFUT1 (*Saccharomyces cerevisiae* POFUT1) can both correctly position the sugar donor-binding site and limit structural flexibility[35]. Additionally, there are a few of different manners for combination of Gt monomers. SpnP is reported to homodimerize using Trp and His residues from each monomer forming π-stacking interactions. This homologous interaction of Trp residues is also observed in some other NP Gts.[13] Some other Gts are able to form more complex oligomers, including trimers[54], tetramers[52, 85, 120], and even dodecamers[77]. Structural studies revealed that most of the residues involved in oligomerization were conserved within the related Gts. They might be primarily hydrophobic and aromatic residues which form an extensive hydrophobic interface between the monomers.[120] However,

oligomerizations usually have no direct role in catalysis, because of the absence of active site at the interface between monomers.[121]

Some studies have described that Gt complex formation could improve the enzymatic activity of each of the partners[122-124], and Gt complexes were able to enhance glycosylation function.[125] Recent studies demonstrated that Gts might catalyze successive steps in the glycan biosynthesis through interacting with each other to form larger complexes, such as Alg7p, Alg13p, and Alg14p Gts catalyzing the early steps of N-glycosylation can form a functional multi-enzyme complex. These enzymes have similar structure and the spatial organization.[126] In the Golgi glycosylation pathways, Gt complexes are crucial for cell surface glycan synthesis. Homo- and heteromeric complexes were both described for Golgi Gts of the N- and O-glycosylation. Furthermore, the heteromeric complex formation was proved to be dependent on Golgi acidity.[127] Besides, there are also many examples for Gt complexes: ganglioside synthesis is involved in several distinct units, each of which formed by complexes of particular Gts, concentrating in different sub-Golgi compartments;[128] β3Gn (1,3-Nacetylglucosaminyltransferase)-T8 and β3Gn-T2 can form a complex involved in the elongation of specific branch structures of multiantennary N-glycans;[129] the core enzyme GtfA and co-activator GtfB in *Streptococcus pneumoniae* form an OGT complex to glycosylate the serine-rich repeat of adhesin PsrP (pneumococcal serine-rich repeat protein), and a β-meander add-on domain contributes to forming an active GtfA-GtfB complex.[55] Complexes are formed almost exclusively by the catalytic domains of the interacting enzymes. To

date, no complex composed of enzymes functioning in different pathways has been identified.

In addition, Gts are also able to form complexes with non-Gt proteins. The lactose synthase complex, which transfers Gal to Glc to produce lactose, is formed by the interactions of β-1, 4-galactosyltransferase-1 and α-lactalbumin. The interaction in the closed conformation only occurred in the presence of substrates. The complex prefers to hold Glc in the monosaccharide binding pocket for maximum contact, thereby improves its affinity for Glc.[130]

3.3 Gts with auxiliary proteins or chaperones

Some researchers have reported that a few Gts could not function without the particular auxiliary proteins or chaperones. A good example is the DesVII/ DesVIII –like protein pairs responsible for some antibiotic glycosylation, which will be discussed in the section of NP Gts.

Other Gts functioning with corresponding chaperones are also observed. The endoplasmic reticulum chaperone Cosmc is indispensable for the formation of active and stable T-synthase (core 1 β1-3-galactosyltransferase). It directly binds to the T-synthase to not only assist the folding correctly, but also regulate T-synthase biosynthesis.[131] Further studies demonstrated that a short peptide in the N-terminal stem region of the T-synthase was crucial for Cosmc recognition and binding.[132] Another recent study revealed that the two-protein enzyme complex, composed of Gtf1 and Gtf2, was required for glycosylation of a serine-rich repeat protein. Gtf2 functions as a chaperone to stabilize Gtf1, enhancing its enzymatic activity.[133]

4. Evolutionary studies of Gts

Considering the diversity of sequences and functions, elucidating the evolutionary origin and the evolutionary process of Gt families has caused much attention. Most of current evolutionary studies on Gts focus on certain species (mainly based on the whole genome data analysis[134-137]) or certain functions (exerting similar catalytic function or utilizing similar substrates[138-140]). Nevertheless, investigating the evolutionary process from the perspective of reaction mechanism and structure may have more universal significance. Some researchers proposed that the inverting GT2 and retaining GT4, containing about half of the total GT members, may be the ancestral families of the different stereochemistries.[26] Meanwhile, it remains a question for a long time whether GT-A and GT-B proteins derived from a common ancestor and which fold was the more ancient one. A typical view proposes that the GT-A fold is more ancient since GT-A members are the most representative Gts, and the GT-B fold is the gene duplication product from an ancestral GT-A protein. The similarity of two domains of GT-B fold further supports the viewpoint of gene duplication. An alternative view suggests that GT-B fold evolved first, because GT-B members are mostly responsible for the biosynthesis of core glycan structures while GT-A members for elongation and terminal decoration.[94] Yet some researchers believe that these two superfamilies have evolved independently.[89] Meanwhile, the remote homology between GT-B Gts and some non-Gt proteins such as UDP-N-acetylglucosamine 2-epimerase has been proposed based on the similarity of overall topology and the conserved domain in secondary structure. [141]

Additionally, these two superfamilies present somewhat evolutionary convergence in reaction mechanism. For inverting Gts, the common catalytic base is Asp or Glu residue, as well as His residue which is found almost exclusively in GT-B members. However, CstI (α-2, 3-sialyltransferase) and CstII (α-2, 3/2, 8-sialyltransferase) adopting GT-A fold unexpectedly employ a His to implement catalysis.[142] On the other hand, the phosphate leaving group is generally stabilized by divalent cation, coordinated by DxD motif for GT-A proteins. However, some members in GT-14 and GT-42 with GT-A fold lack divalent cation binding motif. The similar roles of neutralizing negative charges are probably played by Arg, Lys residues[143] and two Tyr residues[121] respectively. This phenomenon is reminiscent of the case in GT-B proteins, which are also independent on metal ions.

In the large Gt system, the evolution is not linear and regular totally. Some studies showed that horizontal gene transfers (HGTs) and other special evolution behaviors have contributed greatly to Gt evolution. Our research suggested that although most of GT1 antibiotic Gts derived from a common ancestor, polyene macrolide Gts displayed obvious unusual eukaryotic origin.[140] This result was further supported by the evolutionary analysis based on the protein structures.[144] Meanwhile, a recent study showed that genes coding UDP-glycosyltransferases (UGTs) in *Tetranychus urticae* genome may be also gained from bacteria by HGT.[145]

5. NP glycosylation

In nature, many NPs, especially the secondary metabolites of microbes and plants (e.g. antibiotics, flavonoids, terpenoids, alkaloids, etc.) are glycosylated (Fig. 6).

The regio- and stereo- specific attachments of sugar moieties usually result in better bioactivities and different physicochemical properties.[146, 147] Glycosylation is also a universal strategy for antibiotic resistance and elimination of harmful xenobiotics. Furthermore, glycosylation of secondary metabolites can also influence disease resistance and individual development in plants.

NP glycosylation especially for antibiotics generally occur in the late biosynthetic steps. Most of these sugar moieties for NPs are unique deoxy-hexoses and derived from the key intermediates of sugar metabolisms (such as NDP-4-keto-6-deoxyglucose).[148, 149] The vast majority of NP Gts utilize NDP-sugar as the donor, while 5-phosphoribose diphosphate was also reported to be used by a phosphoribosyltransferase, participating in the biosynthesis of the aminoglycoside antibiotic butirosin.[150] Most NP Gts fall into GT-1, belonging to the GT-B superfamily. Until now, many crystal structures of NP Gts have been illustrated which pave the way for further investigating their reaction mechanism (Table. 1).

The NP glycosylation reactions have some special characteristics in glycosylation pattern and substrate specificity. Some specific members of NP Gts even possess some peculiar functions, such as reversability, iterative catalysis, protein auxiliary, etc. Therefore, investigation of NP glycosylation and associated Gts has been a hot topic for many years. Table 4 exhibits NP Gts identified in recent five years.

5.1 Functions of NP glycosylation

5.1.1 Influence on physico-chemical properties and bioactivities

A significant function of glycosylation is to increase polarity and solubility which can raise the intracellular or extracellular concentration of NPs. Meanwhile, the chemical stability of the corresponding aglycones could be significantly improved. The glycosylation modification of macromolecules can even influence their overall topology.[194] The antibiotics with sugar moieties exhibit a mix of hydrophobic and hydrophilic surfaces, and the sugar moieties can offer the ability for hydrogen-bond formation that facilitates specific recognition by the biological targets. If the appended sugar(s) is or are removed, the bioactivities of these NPs are either completely lost or dramatically decreased.[194]

For most 16-membered macrolides (e. g. tylosin, Fig. 6(b)), the binding to the ribosome is primarily facilitated by the hydrophobic interactions of the C-5 sugar branch with the ribosome and its complementarity in shape to the binding site. However, orientations and conformations of the 14-membered macrolides (e. g. erythromycin, Fig. 6(c)) are different. The sugar moieties of macrolides are proposed to contribute notably to binding-free energy since they provide 1/2 to 2/3 of the interaction surface. The length of the oligopeptide synthesized by ribosome is determined to some extent by the substituents at the C5 position of the lactone ring in macrolides.[195] These sugar moieties were entrapped with ribosome through hydrogen bond during substrate binding. Although there is a significant difference among 14-, 15- and 16-membered macrolides, they all have a sugar moiety at the C5 position. The 2'-OH group of this sugar functions as an important binding site. The dimethylamino groups of sugar moiety in macrolides also play an important role for the antibiotic

effectiveness.[195] Table 5 lists some other findings in sugar-involved antibacterial mechanisms.

Sugar moieties can also serve as a binding domain of carotenoids allowing the proper folding and stacking of the thylakoid membrane in cyanobacterium.[205] In plants, O-glycosylation generally reduces flavonoid bioactivity, while C-glycosylation usually enhances the flavonoid benefits to human health, including antioxidant and anti-diabetic potential.[206]

5.1.2 Antibiotic resistance related to glycosylation

There is a kind of special NP Gts which could inactivate macrolides for self-protection from endogenous or exogenous antibiotics. The first Gt with this function was attained from *Streptomyces lividans* and was named MGT. MGT can transfer another sugar to the free 2'-OH group in the C-5 monosaccharide of 14- or 16-membered lactone ring (corresponding to C-3 in the 12-membered lactone ring), hiding the antibiotic-substrate binding site.[207] In general, a ternary complex of antibiotic, UDP-sugar and MGT is formed prior to the glycosyl transfer, and the antibiotic bound to MGT earlier than UDP-sugar. After the glycosylation, UDP is released first, followed by the glycosylated antibiotic.[208] Antibiotic inactivation through glycosylation is reported to be a universal resistant mechanism not only to macrolides. Some pathogenic nocardia show a strong capacity to inactivate many kinds of macrolides and ansamycins.[209] There is also an ORF encoding a special mannosyltransferase speculated to serve as a protectant in the biosynthetic gene cluster of glycopeptide antibiotic teicoplanin.[210] Considering the special function of

MGT, it always has broad substrate specificity and has been developed to be a powerful tool for glycorandomization.[211]

Interestingly, some macrolide producers utilize this mechanism to initially inactivate the endogenous antibiotics. Then the added sugar is hydrolyzed by their extracellular hydrolases and the bioactivity of antibiotics is recovered.[212] Besides MGT OleD, there is another Gt OleI in oleandomycin biosynthetic gene cluster, nearly specific for oleandomycin glycosylation rather than glycosylating a wide range of macrolides like OleD, indicating different functions. [213] Further studies displayed the donor and acceptor flexibility, and indicated that aromatic residues might play a crucial role in substrate recognition.[15] Asm25 together with Asm41 (glycosyl hydrolase) in ansamitocin biosynthetic gene cluster was also speculated to exert homologous function with OleI.[214]

Additionally, several other Gts have been identified or speculated to be MGTs (Table 4), such as GimA in a spiramycin producer (another MGT was proposed to exist but has not been identified)[215], Ses60310 which can use different NDP-sugars to rhamnosylate a series of phenolic compounds with a remarkable regioflexibility[176], and AcuGT3 with high similarity to MGT in aculeximycin biosynthesis gene cluster[152], etc.

5.2 Glycosylation patterns

The glycosylation patterns of NPs are quite diverse. The sugar moieties can be attached to the O-, N-, C- and even S-atoms of aglycons or other sugars. O-glycosylation is the most common pattern of NP glycosylation, while the N-, C-

and S-glycosylations are relatively rare. Rationally, the catalytic mechanism and evolutionary origin of related Gts are somewhat different. In nature, a few Gts can perform different catalytic functions, such as O-/N-glucosyltransferase UGT72B1 in *Arabidopsis thaliana*[216], UrdGT2 catalyzing both C- and O-glycosylation[217] and ThuS catalyzing both S-glycosylation of Cys and O-glycosylation of Ser to form glycopeptide bacteriocin thurandacin[218]. OleD from *Streptomyces antibioticus*[211] and UGT73AE1 from *Carthamus tinctorius*[184] can even glycosylate different acceptors to form O-, S-, and N-glycosidic bonds. The studies on UGT72B1 revealed that the N-glycosylation was probably determined by the ability of the candidate catalytic His to direct and orientate nucleophilic attack.[216] Comparison of LanGT2 (an O-Gt for landomycin (Fig. 6(d)) biosynthesis) and LanGT2S8Ac (an engineered C-GT, grafting ^{51}VATTDLPIRHFI^{62} of UrdGT2 into LanGT2 and an S8A mutation) demonstrated that the aglycon-binding pocket of LanGT2S8Ac is much smaller than that of LanGT2. In LanGT2, the C8-OH is closer to the NDP-sugar, facilitating the C-O glycosidic bond formation, while the orientation of acceptor caused by the spatial change brings the C9 atom closer to the sugar donor, thus resulting in a C-C coupling instead.[17] In short, precise positioning and orientation of the aglycon and the proper microenvironments in vicinity of the catalytic sites may determine the type of glycosidic bond and the reaction process.

For the vast majority of NPs containing natural C-C glycosidic bonds (such as urdamycin (Fig. 6(e)), Simocyclinone D8 (Fig. 6(f)), hedamycin, SF2575, gilvocarcin, etc), glycosylation exclusively occurs ortho and/or para to phenol OH-group.

Phylogenetic analysis displayed the obvious distance between C-Gts and O-Gts, and also indicated that these Gts may evolve via HGTs with whole antibiotic biosynthetic gene clusters.[218] Two possible C-glycosylation mechanisms (I and II) have been proposed. Mechanism I refers to the initial O-glycoside formation followed by an intramolecular rearrangement to an ortho-C-glycoside. Mechanism II is analogous to a direct Friedel-Crafts substitution reaction, involving in the attack of a resonance-stabilized phenolate anion at the anomeric C atom of the NDP-sugar (Fig. 7). In recent years, the increasing number of studies lent supports for Mechanism II, and provided more detailed reaction mechanism.[217, 219] After partial deprotonation of the phenol OH-group by a catalytic base, nucleophilic character is directly generated on O2, and then on the aromatic C3 through resonance. Subsequently, C-glycosylation is achieved by a single nucleophilic displacement at the anomeric C atom, probably involving formation of an oxocarbenium ion-like TS (Fig. 7).[219]

S-glycosylation is not universal in nature, and only a handful of S-Gts have been identified, including SunS[221] and ThuS[218] involved in S-linked glycopeptide bacteriocin biosynthesis; UGT74B1[222], UGT74C1[223], S-GT[224] participating in plant glucosinolate biosynthesis; and the trifunctional plant Gt UGT73AE1[184]. The reaction mechanisms and characteristics of S-glycosylation still remain to be understood. In addition, an angucycline-type antibiotic BE-7585A contains a thioether bond between a disaccharide unit and the anthraquinone core, but no S-Gt can be found in its biosynthetic gene cluster, indicating that the glycosylation may be a non-enzymatic reaction.[158]

5.3 Special functions of NP Gts

5.3.1 The reversibility and iterative catalysis

Gts are generally perceived as unidirectional catalysts for a long time. However, Zhang *et al.* discovered that some NP Gts were able to catalyze reversible, bidirectional reactions.[225] Based on the flexible substrate specificity of Gts, they applied the Gt (calicheamicin CalG1, CalG4 and vancomycin GtfD, GtfE) reversibility to get a series of unobtainable NDP-sugars and unnatural antibiotic derivatives, which provided more possibilities for developing antibiotic diversity. Other studies revealed that EryBV[226], VinC[227], AveBI[228], AmphDI, NysDI[229], CalG3[11], SpnP[13], AmiG[230], GT83F (UGT78G1) from *Medicago truncatula*[231] and UGT73AE1[184] also possess catalytic reversibility to form the corresponding sugar donor and aglycone.

This unique reaction characteristic may be related to the catabolism of secondary metabolites, but the reaction mechanism is still unclear. The site-directed mutation on *Medicago* Gt UGT85H2 revealed that a single Glu residue in the substrate binding pocket played a key role in the reversibility. This mutation may alter the size of substrate binding pocket, allowing the product to fit into the pocket with moderate flexibility.[232] Further study on UGT78G1 also indicated two acidic amino acid residues and a His residue played important roles in the reversibility.[24]

Unexpectedly, there is some mismatch between the numbers of sugar moieties and Gts in some particular antibiotics. Some early researchers suspected the unrevealed Gts might position out of the antibiotic biosynthetic gene clusters.

However, more and more recent studies have confirmed the existence of iterative Gts performing glycosylation more than once. Both of LanGT1 and LanGT4 can catalyze the linkage of two sugar moieties. Although LndGT1 (involved in landomycin biosynthesis in another strain) has 74.8% sequence similarity with LanGT1, it is not able to complete the iterative glycosylation. Meanwhile, similar to LanGT4, LndGT4 also catalyze the transfer of two L-rhodinoses. [233] Additionally, AveBI (avermectin)[228] and AknK (aclacinomycin (Fig. 6(g))[234] were identified to be responsible for the stepwise tandem assembly of the disaccharide. CosG, CosK and a third unidentified Gt may glycosylate at both sides of ring D of the cosmomycin aglycone.[235] MtmGIV can even recognize two distinct sugar donors in mithramycin biosynthetic pathway.[236] A few recent studies reported that SaqGT3, SaqGT4 in saquayamycin biosynthesis[237] and LobG3 in lobophorin pathway[169] also perform glycosylation iteratively. Surprisingly, PnxGT2 in FD-594 pathway even possesses continuous iterative glycosylation ability to form tri-, tetra- and penta-olivosides.[173] The particular feature of iterative Gts indicates that they may only recognize part of aglycone rather than the whole structure.

5.3.2 Glycosylation assisted by auxiliary proteins

As mentioned above, some NP Gts cannot function without the aid of an auxiliary protein. This reaction characteristic for Gts involved in antibiotic biosynthesis was first revealed in DesVII/DesVIII (Gt/auxiliary protein) in the methymycin/pikromycin pathway. DesVII is active only in the presence of DesVIII, and the auxiliary proteins may have a chaperone-like function to facilitate a one-time

conformational change.[238, 239] An increasing number of studies have found the DesVII/DesVIII homologues in other antibiotic biosynthetic pathways, including EryCIII/EryCII (erythromycin)[240], TylM2/TylM3 (tylosin)[241], MycB/MydC (mycinamicin)[241], MegCIII/MegCII and MegDI/MegDVI (megalomicin)[242] , AngMII/AngMIII (angolamycin)[243], Srm5/Srm6 and Srm29/Srm28 (spiramycin)[179], PnxGT1/PnxO5 (FD-594)[173] and AknS/AknT (aclacinomycin)[244]. All of above auxiliary proteins show somewhat sequence similarity to cytochrome P450 enzymes, but lack the strictly conserved Cys residues. Some studies revealed that the auxiliary proteins could affect neither the Gt expressions[245] nor the Gt specificities[242]. The unique interplay of Gts with the auxiliary proteins in spiramycin biosynthesis was also reported.[179] Srm28, one of the auxiliary proteins, is able to interact with both Gts (Srm5 and Srm29). Interestingly, genes encoding these auxiliary proteins are exclusively (with the exception of SpnP) located directly upstream of the corresponding Gt encoding genes.

Researchers made great efforts on elucidating this mechanism. Studies on aclacinomycin biosynthetic pathway indicated that AknT is a saturable activator for AknS, and the molar ratio of T/S is about 3:1. The two proteins are in rapid reversible equilibrium with the monomers.[244] Nevertheless, the studies on EryCIII/ EryCII may obtain a different conclusion. The studies on crystal structure displayed an almost linear organization of the heterotetrameric consisting of two EryCIII·EryCII heterodimers, and further supported the allosteric activator hypothesis (Fig. 8). The results also confirmed the important role of the auxiliary protein on stabilizing the

fold of its partner Gt.[12] Another structural study on SpnP identified a three helices motif as the common feature of Gts aided by auxiliary proteins and proposed the corresponding amino acid sequence to be an identification tag of these Gts.[13]

Hong *et al.* used complementation to verify that DesVIII and its homologues (TylMIII, EryCII, DnrQ and OleP1) were functionally exchangable in different degree. The results suggested that DnrQ and OleP1 might also exert the auxiliary function in antibiotic glycosylation.[239] In addition, several enzymes with sequence similarity to the auxiliary proteins have been identified in some other antibiotic biosynthetic gene clusters without the functional verification, such as StaN (assists for StaG, staurosporine (Fig. 6(h))[246], CosT (assists for CosG, cosmomycin)[235], StfPII (assists for StfG, steffimycin)[247], KstD4 (assists for KstD5, kosinostatin)[168], IdnO2 (assists for IdnS14, incednine)[166], Strop2219 (assists for Strop2220, lomaiviticin)[181], Pau14 (assists for Pau15, paulomycin)[172] and Arm11 and Arm32 (assist for Arm12 and Arm33 respectively, arimetamycin)[154].

5.4 NP Gt applications and synthetic biology

Gts especially that participate in NP biosynthesis have been proved to be useful for the chemoenzymatic synthesis or biosynthesis of functional compounds with new bioactivities. Attachment or change of sugar moieties in natural or synthetic compounds can greatly influence their biological properties. So in-depth knowledge of NP Gts is important for rational design of novel glycosylated compounds. The direct application of NP Gts is enzymatic glycosylation generally with precise regio/stereo-selectivity and moderate reaction conditions.[248-251] A few recent studies

extend this method to couple different Gts in "one-pot".[252]

Compared to Gts that involved in primary metabolism, NP Gts especially that involved in antibiotic glycosylations have more flexible substrate (sugar donors or aglycons) specificities (Table 6), making them powerful tools to catalyze various coupling reactions for obtaining novel glycoderivatives. It is the most common strategy to transfer different sugar moieties to various aglycons *in vivo* or *in vitro* by utilizing the broad substrate specificity of NP Gts.

Combinational biosynthesis (Fig. 9) is a typical engineering glycosylation method *in vivo*, through heterologous expression of partial or whole aglycon biosynthetic genes and/or genes encoding Gts in the strains that can generate various NDP-sugars[254, 263, 269], or expression of NDP-sugar biosynthetic gene cassettes (with or without genes encoding Gts) in the host strains of different antibiotics[243, 258, 261, 267, 270-273]. A recent study also reported the utilization of *E. coli* platform for combinational biosynthesis of antibiotics[274], consistent with the guideline of synthetic biology. These studies developed numerous new glycoderivatives, some of which may become valuable drug candidates. The well-known application of NP Gts *in vitro* is "*in vitro* glycorandomization" based on the flexibility of GtfD and GtfE. These two Gts were used on NDP-sugar libraries to generate glycorandomized NPs and then applied chemoselective ligation to produce monoglycosylated vancomycins. The obtained products varied considerably in terms of bioactivity and one of them notably improved the antibacterial property.[260] More studies based on this strategy have been highlighted in Table 6.

Although the flexible Gts can utilize different sugar donors or aglycons, the *in vitro* NDP-sugar library is difficult to obtain because of the synthetic difficulties of some special sugar donors. Fortunately, some NP Gts which can catalyze reversible reactions as mentioned above provide a novel strategy to generate uncommon NDP-sugars through cleavage of antibiotic glycoside bonds retransferring sugar moieties to NDP.[275] Based on the reversibility of Gts, the sugar/aglycon exchange method and a simpler one-pot strategy for one-off generation of copious glycosylated compounds have been performed.[225-228, 230] Nevertheless, the natural substrate promiscuity of Gts is always limited, inspiring researchers to further improve Gt promiscuity of diverse substrates. The most common strategies include the domain swapping to generate chimeric Gts[28, 101] and the directed evolution to change the crucial sites of Gts[103, 112, 211].

Combinational biosynthesis of natural glycoderivatives takes a page from the ideas of synthetic biology, which will broaden the mind of studies on glycosylation, and extend its applications. Synthetic biology has been driven by the development of systems biology and new powerful tools for DNA synthesis, sequencing and genome editing, which enables a new generation of microbial engineering for the biotechnological production of pharmaceuticals and other high-value chemicals[276]. The purpose of synthetic biology is to develop complex biological systems that represent novel biological functions by combining basic designed units, just like integrating circuits. The catalytic diversity of living systems[277] and the exploration of novel enzymes catalyzing unnatural reactions[278] can be developed for synthetic

devices or systems.

Besides bacteria, some studies have successfully designed and constructed novel biological systems of yeast to produce NPs in recent years. One of the most well-known examples is the biological production of artemisinic acid, a precursor of artemisinin, in engineered *Saccharomyces cerevisiae*.[279, 280] Glycosylated NPs are also popular targets for synthetic biology, especially plant triterpenes, flavonoids, and alkaloids with high bioactivities. Reconstitution of biosynthetic pathways in yeast is a promising strategy for rapid and inexpensive production of these complex molecules. Protopanaxadiol, the aglycon of several dammarane-type ginsenosides, was produced by *S. cerevisiae* through the introduction of dammarenediol-II synthase and protopanaxadiol synthase genes of *Panax ginseng*, together with a NADPH-cytochrome P450 reductase gene of *Arabidopsis thaliana*.[281] Then, 16 UGTs from *Panax ginseng* were expressed in this chassis yeast respectively, and compound K, the main functional component of ginsenosides, was successfully obtained in one recombinant strain.[282] Strictosidine, a glycosylated alkaloid, is the core scaffold of all known monoterpene indole alkaloids. It can also be produced in an engineered *S. cerevisiae* host through the introduction of 14 known monoterpene indole alkaloid pathway genes and other seven genes.[283]

In recent years, many expression hosts used in synthetic biology have been glyco-engineered, further strengthening the application of glycosylation in synthetic biology.[284] New tools provided by synthetic biology and the novel gene expression systems also create unprecedented opportunities for building a Gt-based

biomanufacturing platform.[285, 286] In a word, glycosylations have shown great potential for applications in synthetic biology and will certainly further promote its development.

6. Summary and outlook

Glycosylation is one of the most important physiological and biochemical reactions in nature, whose crucial roles in vital processes attracted numerous researchers to focus on glycosylation mechanisms and the characteristics of Gts, facilitating in-depth understanding and further application of glycosylation reactions. Despite the versatility of glycosylation processes and substrates, the basic mechanisms are not diverse. Orthodox researches on glycosylation mechanisms mainly focus on the individual stereochemistry (inverting or retaining) or Gt topology (GT-A, GT-B and GT-C), trying to designate a uniform mechanism in each type. Nevertheless, like for the retaining mechanism, which has been debated for more than a decade, it was demonstrated that the double-displacement and the S_Ni-like mechanisms were both possible by some recent works. Even for the inverting mechanism, although the S_N2-like mechanism was popular, a S_N1-like mechanism was proposed for some peculiar inverting Gts[35]. Gts adopting similar folds generally follow uniform reaction processes, consistent with the bi-bi mechanisms, while their catalytic sites and substrate binding patterns are diverse. Meanwhile, the increasing number of new Gt folds (such as GT-D[51] and lysozyme-type) and functional insertion motifs made the glycosylation reactions more complex. More and more experimental and theoretical studies indicated that the positioning and orientation of substrates as

well as the reaction microenvironments, but not the overall Gt topology, might be the determinants of stereo- and regio- selectivities and even the glycosylation patterns (O-, N-, C- or S-glycosylation).

Nevertheless, glycosylation reactions and Gts are still mystical due to their special characteristics. Why can a unique enzyme tolerate a wide range of substrates and even transfer sugar donors to the different positions of aglycons? How can Gts control the order and the direction of glycosylation reactions? Do structurally similar Gts with poor sequence identities derive from a common ancestor? Although great efforts have been made in functional and structural investigations of Gts, these questions remain unclear to date. With the flourish of protein crystallization and analytic techniques[287, 288] as well as site-directed mutation techniques[289, 290], the special features of glycosylation and Gts will be eventually interpreted and better utilized.

The crucial roles of sugar moieties in NP physico-chemical property and bioactivity made glycosylation a research hotspot in NP biosynthesis and modification. Despite experimental and clinical supports for the importance of sugar moieties, there are still great challenges that needed to be addressed for molecular mechanisms of the relationship between glycosylation and bioactivity. These studies will also lay a foundation for rational design of drug candidates with new glycosylation modification. Naturally, NP Gts, many of which have some applicable features such as substrate flexibility, catalytic reversibility or iterative capability, become useful tools in new drug development and are modified in various ways for better serviceability.

Consequently, in-depth understanding of the characteristics and mechanisms of Gts is essential for further application. During the past two decades, works on NP glycosylation modification emerged in an endless stream and many mature techniques have been developed including enzymatic glycosylation using whole cells or single Gts, *in vitro* glycorandomization, and combinational biosynthesis consistent with the idea of synthetic biology. Among these strategies, combinational biosynthesis is a useful and promising method but limited by the function research, DNA assembly techniques and the lack of biosynthetic elements. Fortunately, genome sequencing has become more economical, providing a lot of opportunities to mine numerous silent NP biosynthetic genes[291, 292]. Meanwhile, some new techniques arose in recent years such as CRISPR and TALENs made genome editing more efficient and accurate.[293]

Bringing the ideas of synthetic biology into glycosylation studies undoubtedly open a door to more extensive and more efficient application of glycosylation reactions. Exploration of new ideas and methods based on synthetic biology will also bring new vigor and vitality into glycobiology.

Acknowledgements

We thank Prof. Shuangjun Lin, Shanghai Jiaotong University, Dr. Bin Liu, Nankai University for many valuable comments and Dr. Jinghua Yang, Institute of Microbiology Chinese Academy of Science and Dr. Guangyu Yang, Shanghai Jiaotong University for critical reading of the manuscript. This work was funded by the National Key Technology Support Program (2014BAD02B00), the National

Natural Science Foundation of China (31270142), and Tianjin applied basic research projects and cutting-edge technology (Tianjin Natural Science Foundation, No. 14ZCZDNC0008).

References

1 A. Cortés, C. Muñoz-Antoli, J. Sotillo, B. Fried, J. G. Esteban and R. Toledo, *Parasite Immunol.*, 2015, 37, 32-42.
2 A. Varki, *Cell*, 2006, 126, 841-845.
3 J. R. Bishop and P. Gagneux, *Glycobiology*, 2007, 17, 23R-34R.
4 S. F. Hansen, E. Bettler, A. Rinnan, S. B. Engelsen and C. Breton, *Mol. Biosyst.*, 2010, 6, 1773-1781.
5 L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, *Annu. Rev. Biochem.*, 2008, 77, 521-555.
6 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic. Acids Res.*, 2014, 42, D490-D495.
7 A. M. Mulichak, H. C. Losey, W. Lu, Z. Wawrzak, C. T. Walsh and R. M. Garavito, *Proc. Natl. Acad. Sci. U S A.*, 2003, 100, 9238-9243.
8 A. M. Mulichak, H. C. Losey, C. T. Walsh and R. M. Garavito, *Structure*, 2001, 9, 547-557.
9 A. M. Mulichak, W. Lu, H. C. Losey, C. T. Walsh and R. M. Garavito, *Biochemistry*, 2004, 43, 5170-5180.
10 A. Chang, S. Singh, K. E. Helmich, R. D. Goff, C. A. Bingman, J. S. Thorson and G. N. Phillips Jr, *Proc. Natl. Acad. Sci. U S A.*, 2011, 108, 17649-17654.
11 C. Zhang, E. Bitto, R. D. Goff, S. Singh, C. A. Bingman, B. R. Griffith, C. Albermann, G. N. Phillips Jr and J. S. Thorson, *Chem. Biol.*, 2008, 15, 842-853.
12 M. C. Moncrieffe, M. J. Fernandez, D. Spiteller, H. Matsumura, N. J. Gay, B. F. Luisi and P. F. Leadlay, *J. Mol. Biol.*, 2012, 415, 92-101.
13 E.A. Isiorho, B. S. Jeon, N. H. Kim, H. W. Liu and A. T. Keatinge-Clay, *Biochemistry*, 2014, 53, 4292-4301.
14 E. A. Isiorho, H. W. Liu and A. T. Keatinge-Clay, *Biochemistry*, 2012, 51, 1213-1222.
15 D. N. Bolam, S. Robert, M. R. Proctor, J. P. Turkenburg, E. J. Dodson, C. Martinez-Fleites, M. Yang, B. G. Davis, G. J. Davies and H. J. Gilbert, *Proc. Natl. Acad. Sci. U S A.*, 2007, 104, 5336-5341.
16 M. Mittler, A. Bechthold and G. E. Schulz, *J. Mol. Biol.*, 2007, 372, 67-76.
17 H. K. Tam, J. Härle, S. Gerhardt, J. Rohr, G. Wang, J. S. Thorson, A. Bigot, M. Lutterbeck, W. Seiche, B. Breit, A. Bechthold and O. Einsle, *Angew. Chem. Int. Ed. Engl.*, 2015, 54, 2811-2815.
18 M. Claesson, V. Siitonen, D. Dobritzsch, M. Metsä-Ketelä and G. Schneider, *FEBS*

J., 2012, 279, 3251-3263.
19 F. Wang, M. Zhou, S. Singh, R. M. Yennamalli, C. A. Bingman, J. S. Thorson and G. N. Phillips Jr, *Proteins*, 2013, 81, 1277-1282.
20 T. Hiromoto, E. Honjo, T. Tamada, N. Noda, K. Kazuma, M. Suzuki and R. Kuroki, *J. Synchrotron. Radiat.*, 2013, 20, 894-898.
21 T. Hiromoto, E. Honjo, N. Noda, T. Tamada, K. Kazuma, M. Suzuki, M. Blaber and R. Kuroki, *Protein Sci.*, 2015, 24, 395-407.
22 H. Shao, X. He, L. Achnine, J. W. Blount, R. A. Dixon and X. Wang, *Plant Cell*, 2005, 17, 3141-3154.
23 L. Li, L. V. Modolo, L. L. Escamilla-Trevino, L. Achnine, R. A. Dixon and X. Wang, *J. Mol. Biol.*, 2007, 370, 951-963.
24 L. V. Modolo, L. Li, H. Pan, J. W. Blount, R. A. Dixon and X. Wang, *J. Mol. Biol.*, 2009, 392, 1292-1302.
25 W. Offen, C. Martinez-Fleites, M. Yang, E. Kiat-Lim, B. G. Davis, C. A. Tarling, C. M. Ford, D. J. Bowles, G. J. Davies, *EMBO J.*, 2006, 25, 1396-1405.
26 C. Martinez-Fleites, M. Proctor, S. Roberts, D. N. Bolam, H. J. Gilbert and G. J. Davies, *Chem. Biol.*, 2006, 13, 1143-1152.
27 M. C. Cavalier, Y. S. Yim, S. Asamizu, D. Neau, K. H. Almabruk, T. Mahmud and Y. H. Lee, *PLoS One.*, 2012, 7, e44934.
28 A. W. Truman, M. V. Dias, S. Wu, T. L. Blundell, F. Huang and J. B. Spencer, *Chem. Biol*., 2009, 16, 676-685.
29 J. Liu and A. Mushegian, *Protein Sci.*, 2003, 12, 1418-1431.
30 M. Krupicka and I. Tvaroska, *J. Phys. Chem. B*, 2009, 113, 11314-11319.
31 I Tvaroška, S. Kozmon, M. Wimmerová and J. Koča, *J. Am. Chem. Soc*., 2012, 134, 15563-15571.
32 I Tvaroška, S. Kozmon, M. Wimmerová and J. Koča, *Chemistry*, 2013, 19, 8153-8162.
33 S. Kozmon and I. Tvaroska, *J. Am. Chem. Soc*., 2006, 128, 16921-16927.
34 B. W. Murray, V. Wittmann, M. D. Burkart, S. C. Hung and C. H. Wong, *Biochemistry*, 1997, 36, 823-831.
35 E. Lira-Navarrete, J. Valero-González, R. Villanueva, M. Martínez-Júlvez, T. Tejero, P. Merino, S. Panjikar and R. Hurtado-Guerrero, *PLoS One*, 2011, 6, e25365.
36 M. Schimpl, X. Zheng, V. S. Borodkin, D. E. Blair, A. T. Ferenbach, A. W. Schüttelkopf, I. Navratilova, T. Aristotelous, O. Albarbarawi, D. A. Robinson, M. A. Macnaughtan and D. M. van Aalten, *Nat. Chem. Biol.*, 2012, 8, 969-974.
37 A. Monegal and A. Planas, *J. Am. Chem. Soc*., 2006, 128, 16030-16031.
38 N. Soya, Y. Fang, M. M. Palcic and J. S. Klassen, *Glycobiology*, 2011, 21, 547-552.
39 V. Rojas-Cervellera, A. Ardèvol, M. Boero, A. Planas and C. Rovira, *Chemistry*, 2013, 19, 14018-14023.
40 J. C. Errey, S. S. Lee, R. P. Gibson, C. Martinez Fleites, C. S. Barry, P. M. Jung, A. C. O'Sullivan, B. G. Davis and G. J. Davies, *Angew. Chem. Int. Ed. Engl*., 2010, 49, 1234-1237.
41 S. S. Lee, S. Y. Hong, J. C. Errey, A. Izumi, G. J. Davies and B. G. Davis, *Nat.*

Chem. Biol., 2011, 7, 631-638.
42 A. Ardèvol and C. Rovira, *Angew. Chem. Int. Ed. Engl*., 2011, 50, 10897-10901.
43 H. Gómez, I. Polyak, W. Thiel, J. M. Lluch and L. Masgrau, *J. Am. Chem. Soc.*, 2012, 134, 4743-4752.
44 M. O. Ziegler, T. Jank, K. Aktories and G. E. Schulz, *J. Mol. Biol*., 2008, 377, 1346-1356.
45 L. L. Lairson, C. P. Chiu, H. D. Ly, S. He, W. W. Wakarchuk, N. C. Strynadka and S. G. Withers, *J. Biol. Chem.*, 2004, 279, 28339-28344.
46 H. Gómez, J. M. Lluch and L. Masgrau, *J. Am. Chem. Soc.*, 2013, 135, 7053-7063.
47 A. Bobovská, I. Tvaroška and J. Kóňa, *Glycobiology*, 2015, 25, 3-7.
48 I. Tvaroška, *Carbohydr. Res.*, 2015, 403, 38-47.
49 R. W. Wheatley, R. B. Zheng, M. R. Richards, T. L. Lowary and K. K. Ng, *J. Biol. Chem.*, 2012, 287, 28132-28143.
50 J. L. Morgan, J. Strumillo and J. Zimmer, *Nature*, 2013, 493, 181-186.
51 H. Zhang, F. Zhu, T. Yang, L. Ding, M. Zhou, J. Li, S. M. Haslam, A. Dell, H. Erlandsen and H. Wu, *Nat. Commun.*, 2014, 5, 4339.
52 E. Zeqiraj, X. Tang, R. W. Hunter, M. García-Rocha, A. Judd, M. Deak, A. von Wilamowitz-Moellendorff, I. Kurinov, J. J. Guinovart, M. Tyers, K. Sakamoto and F. Sicheri, *Proc. Natl. Acad. Sci. USA*, 2014, 111, E2831-2840.
53 S. Sobhanifar, L. J. Worrall, R. J. Gruninger, G. A. Wasney, M. Blaukopf, L. Baumann, E. Lameignere, M. Solomonson, E. D. Brown, S. G. Withers and N. C. Strynadka, *Proc. Natl. Acad. Sci. USA*, 2015, 112, E576-E585.
54 C. Koç, D. Gerlach, S. Beck, A. Peschel, G. Xia and T. Stehle, *J. Biol. Chem.*, 2015, 290, 9874-9885.
55 W. W. Shi, Y. L. Jiang, F. Zhu, Y. H. Yang, Q. Y. Shao, H. B. Yang, Y. M. Ren, H. Wu, Y. Chen and C. Z. Zhou, *J. Biol. Chem.*, 2014, 289, 20898-20907.
56 J. A. Cuesta-Seijo, M. M. Nielsen, L. Marri, H. Tanaka, S. R. Beeren and M. M. Palcic, *Acta Crystallogr. D Biol. Crystallogr.*, 2013, 69, 1013-1025.
57 M. Momma and Z. Fujimoto, *Biosci. Biotechnol. Biochem.*, 2012, 76, 1591-1595.
58 N. Thiyagarajan, T. T. Pham, B. Stinson, A. Sundriyal, P. Tumbale, M. Lizotte-Waniewski, K. Brew and K. R. Acharya, *Sci. Rep.*, 2012, 2, 940.
59 T. T. Pham, B. Stinson, N. Thiyagarajan, M. Lizotte-Waniewski, K. Brew and K. R. Acharya, *J. Biol. Chem.*, 2014, 289, 8041-8050.
60 Y. Tsutsui, B. Ramakrishnan and P. K. Qasba, *J. Biol. Chem.*, 2013, 288, 31963-31970.
61 K. Brown, S. C. Vial, N. Dedi, J. Westcott, S. Scally, T. D. Bugg, P. A. Charlton and G. M. Cheetham, *Protein Pept. Lett.*, 2013, 20, 1002-1008.
62 B. Kuhn, J. Benz, M. Greif, A. M. Engel, H. Sobek and M. G. Rudolph, *Acta Crystallogr. D Biol. Crystallogr.*, 2013, 69, 1826-1838.
63 L. Meng, F. Forouhar, D. Thieker, Z. Gao, A. Ramiah, H. Moniz, Y. Xiang, J. Seetharaman, S. Milaninia, M. Su, R. Bridger, L. Veillon, P. Azadi, G. Kornhaber, L. Wells, G. T. Montelione, R. J. Woods, L. Tong and K. W. Moremen, *J. Biol. Chem.*, 2013, 288, 34680-34698.
64 H. Schmidt, G. Hansen, S. Singh, A. Hanuszkiewicz, B. Lindner, K. Fukase, R.W. Woodard, O. Holst, R. Hilgenfeld, U. Mamat and J. R. Mesters, *Proc. Natl. Acad Sci.*

U S A, 2012, 109, 6253-6258.
65 E. C. O'Neill, A. M. Rashid, C. E. M. Stevenson, A. C. Hetru, A. P. Gunning, M. Rejzek, S. A. Nepogodiev, S. Bornemann, D. M. Lawson and R. A. Field, *Chem. Sci.*, 2014, 5, 341-350.
66 R. N. Pruitt, N. M. Chumbler, S. A. Rutherford, M. A. Farrow, D. B. Friedman, B. Spiller and D. B. Lacy, *J. Biol. Chem.*,2012, 287, 8013-8020.
67 N. D'Urzo, E. Malito, M. Biancucci, M. J. Bottomley, D. Maione, M. Scarselli and M. Martinelli, *FEBS J.*, 2012, 279, 3085-3097.
68 J. H. Jeong, Y. S. Kim, C. Rojviriya, S. C. Ha, B. S. Kang and Y. G. Kim, *Antimicrob. Agents Chemother*, 2013, 57, 3507-3512.
69 A. Striebeck, D. A. Robinson, A. W. Schüttelkopf and D. M. van Aalten, *Open Biol.*, 2013, 3, 130022.
70 S. Matsumoto, M. Igura, J. Nyirenda, M. Matsumoto, S. Yuzawa, N. Noda, F. Inagaki and D. Kohda, *Biochemistry*, 2012, 51, 4157-4166.
71 J. Nyirenda, S. Matsumoto, T. Saitoh, N. Maita, N. N. Noda, F. Inagaki and D. Kohda, *Structure*, 2013, 21, 32-41.
72 S. Matsumoto, A. Shimada and D. Kohda, *BMC Struct. Biol.*, 2013, 13, 11.
73 S. Matsumoto, A. Shimada, J. Nyirenda, M. Igura, Y. Kawano and D. Kohda, *Proc. Natl. Acad. Sci. U S A,* 2013, 110, 17868-17873.
74 C. I. Chen, J. J. Keusch, D. Klein, D. Hess, J. Hofsteenge and H. Gut, *EMBO J.*, 2012, 31, 3183-3197.
75 K. Schmölzer, T. Czabany, C. Luley-Goedl, T. Pavkov-Keller, D. Ribitsch, H. Schwab, K. Gruber, H. Weber and B. Nidetzky, *Chem. Commun. (Camb.)*, 2015, 51, 3083-3086.
76 N. Huynh, Y. Li, H. Yu, S. Huang, K. Lau, X. Chen and A. J. Fisher, *FEBS Lett.*, 2014, 588, 4720-4729.
77 Q. Yao, Q. Lu, X. Wan, F. Song, Y. Xu, M. Hu, A. Zamyatina, X. Liu, N. Huang, P. Zhu and F. Shao, *Elife*, 2014, 3, e03714.
78 T. Jank, X. Bogdanović, C. Wirth, E. Haaf, M. Spoerner, K. E. Böhmer, M. Steinemann, J. H. Orth, H. R. Kalbitzer, B. Warscheid, C. Hunte and K. Aktories, *Nat. Struct. Mol. Biol.*, 2013, 20, 1273-1280.
79 F. V. Rao, J. R. Rich, B. Rakić, S. Buddai, M. F. Schwartz, K. Johnson, C. Bowe, W. W. Wakarchuk, S. Defrees, S. G. Withers and N. C. Strynadka, *Nat. Struct. Mol. Biol.*, 2009, 16, 1186–1188.
80 L.Y. Lin, B. Rakic, C. P. Chiu, E. Lameignere, W. W. Wakarchuk, S. G. Withers and N. C. Strynadka, *J. Biol. Chem.*, 2011, 286, 37237-37248.
81 M. B. Lazarus, Y. Nam, J. Jiang, P. Sliz and S. Walker, *Nature*, 2011, 469, 564–567.
82 A. J. Clarke, R. Hurtado-Guerrero, S. Pathak, A. W. Schüttelkopf, V. Borodkin, S. M. Shepherd, A. F. Ibrahim and D. M. van Aalten, *EMBO J.*, 2008, 27, 2780-2788.
83 F. Kawai, S. Grass, Y. Kim, K. J. Choi, J. W. St Geme 3rd and H. J. Yeo, *J. Biol. Chem.*, 2011, 286, 38546-38557.
84 H. Ihara, Y. Ikeda, S. Toma, X. Wang, T. Suzuki, J. Gu, E. Miyoshi, T. Tsukihara, K. Honke, A. Matsumoto, A. Nakagawa and N. Taniguchi, *Glycobiology*, 2007, 17,

455-466.
85 S. Baskaran, P. J. Roach, A. A. DePaoli-Roach and T. D. Hurley, *Proc. Natl. Acad. Sci. U S A.*, 2010, 107, 17563–17568.
86 R. Hurtado-Guerrero, T. Zusman, S. Pathak, A. F. Ibtahim, S. Shepherd, A. Prescott, G. Segal and D. M. van Aalten, *Biochem. J.*, 2010, 426, 281-292.
87 C. Breton, H. Heissigerová, C. Jeanneau, J. Moravcová and A. Imberty, *Biochem. Soc. Symp.*, 2002, 69, 23-32.
88 M. L. Rosén, M. Edman, M. Sjöström and A. Wieslander, *J. Biol. Chem.*, 2004, 279, 38683-38692.
89 A. L. Lovering, L. H. de Castro, D. Lim and N. C. Strynadka, *Science*, 2007, 315, 1402-1405.
90 Y. Yuan, D. Barrett, Y. Zhang, D. Kahne, P. Sliz and S. Walker, *Proc. Natl. Acad. Sci. U S A.*, 2007, 104, 5348-5353.
91 B. Ramakrishnan, E. Boeggeman and P. K. Qasba, *Biochemisry*, 2004, 43, 12512-12522.
92 Q. Lu, Q. Yao, Y. Xu, L. Li, S. Li, Y. Liu, W. Gao, M. Niu, M. Sharon, G. Ben-Nissan, A. Zamyatina, X. Liu, S. Chen and F. Shao, *Cell Host Microbe.*, 2014, 16, 351-363.
93 J. S. Klutts, S. B. Levery and T. L. Doering, *J. Biol. Chem.*, 2007, 282, 17890-17899.
94 C. Breton, S. Fournel-Gigleux and M. M. Palcic, *Curr. Opin. Struct. Biol.*, 2012, 22, 540-549.
95 L. Ni, M. Sun, H. Yu, H. Chokhawala, X. Chen and A. J. Fisher, *Biochemistry*, 2006, 45, 2139-2148.
96 L. Larivière, V. Gueguen-Chaignon and S. Moréra, *J. Mol. Biol.*, 2003, 330, 1077-1086.
97 C. Lizak, S. Gerber, S. Numao, M. Aebi and K. P. Locher, *Nature*, 2011, 474, 350-355.
98 K. Hese, C. Otto, F. H. Routier and L. Lehle, *Glycobiology*, 2009, 19, 160-171.
99 F. Yamamoto, H. Clausen, T. White, J. Marken and S. Hakomori, *Nature*, 1990, 345, 229-233.
100 S. I. Patenaude, N. O. Seto, S. N. Borisova, A. Szpacenko, S. L. Marcus, M. M. Palcic and S. V. Evans, *Nat. Struct. Biol.*, 2002, 9, 685-690.
101 A. M. Cartwright, E. K. Lim, C. Kleanthous and D. J. Bowles, *J. Biol. Chem.*, 2008, 283, 15724-15731.
102 X. Z. He, X. Wang and R. A. Dixon, *J. Biol. Chem.*, 2006, 281, 34441-34447.
103 G. J. Williams, C. Zhang and J. S. Thorson, *Nat. Chem. Biol.*, 2007, 3, 657-662.
104 T. Sayama, E. Ono, K. Takagi, Y. Takada, M. Horikawa, Y. Nakamoto, A. Hirose, H. Sasama, M. Ohashi, H. Hasegawa, T. Terakawa, A. Kikuchi, S. Kato, N. Tatsuzaki, C. Tsukamoto and M. Ishimoto, *Plant Cell*, 2012, 24, 2123-2138.
105 A. Noguchi, M. Horikawa, Y. Fukui, M. Fukuchi-Mizutani, A. Iuchi-Okada, M. Ishiguro, Y. Kiso, T. Nakayama and E. Ono, *Plant Cell*, 2009, 21, 1556-1572.
106 A. Ramos, C. Olano, A. F. Braña, C. Méndez and J. A. Salas, *J. Bacteriol*, 2009, 191, 2871-2875.

107 E. Ono, Y. Homma, M. Horikawa, S. Kunikane-Doi, H. Imai, S. Takahashi, Y. Kawai, M. Ishiguro, Y. Fukui and T. Nakayama, *Plant Cell*, 2010, 22, 2856-2871.
108 G. Yang, J. R. Rich, M. Gilbert, W. W. Wakarchuk, Y. Feng and S. G. Withers, *J. Am. Chem. Soc*., 2010, 132, 10570-10577.
109 G. Sugiarto, K. Lau, J. Qu, Y. Li, S. Lim, S. Mu, J. B. Ames, A. J. Fisher and X. Chen, *ACS Chem. Biol*., 2012, 7, 1232-1240.
110 B. Ma, L. H. Lau, M. M. Palcic, B. Hazes and D. E. Taylor, *J. Biol. Chem.*, 2005, 280, 36848-36856.
111 A. Kubo, Y. Arai, S. Nagashima and T. Yoshikawa, *Arch. Biochem. Biophys.*, 2004, 429, 198-203.
112 G. J. Williams, R. D. Goff, C. Zhang and J. S. Thorson, *Chem. Biol.*, 2008, 15, 393-401.
113 R. W. Gantt, P. Peltier-Pain, S. Singh, M. Zhou and J. S. Thorson, *Proc. Natl. Acad. Sci. U S A*., 2013, 110, 7648-7653.
114 H. Yu, H. Chokhawala, R. Karpel, H. Yu, B. Wu, J. Zhang, Y. Zhang, Q. jia and X. Chen, *J. Am. Chem. Soc.*, 2005, 127, 17618-17619.
115 M. È. Charbonneau, J. P. Côté, M. F. Haurat, B. Reiz, S. Crépin, F. Berthiaume, C. M. Dozois, M. F. Feldman and M. Mourez, *Mol. Microbiol.*, 2012, 83, 894-907.
116 T. A. Gerken, C. L. Owens and M. Pasumarthy, *J. Biol. Chem.*, 1997, 272, 9709-9719.
117 M. Chavan, Z. Chen, G. Li, H. Schindelin, W. J. Lennarz and H. Li, *Proc. Natl. Acad. Sci. U S A.*, 2006, 103, 8947-8952.
118 J. Li, T. Y. Yen, M. L. Allende, R. K. Joshi, J. Cai, W. M. Pierce, E. Jaskiewicz, D. S. Darling, B. A. Macher and W. W. Jr Young, *J. Biol. Chem.*, 2000, 275, 41476-41486.
119 E. H. Holmes, T. Y. Yen, S. Thomas, R. Joshi, A. Nguyen, T. Long, F. Gallet, A. Maftah, R. Julien and B. A. Macher, *J. Biol. Chem.*, 2000, 275, 24237-24245.
120 C. P. Chiu, L. L. Lairson, M. Gilbert, W. W. Wakarchuk, S. G. Withers and N. C. Strynadka, *Biochemistry*, 2007, 46, 7196-7204.
121 C. P. Chiu, A. G. Watts, L. L. Lairson, M. Gilbert, D. Lim, W. W. Wakarchuk, S. G. Withers and N. C. Strynadka, *Nat. Struct. Mol. Biol.*, 2004, 11, 163-170.
122 C. G. Giraudo, J. L. Daniotti and H. J. Maccioni, *Proc. Natl. Acad. Sci. U S A.*, 2001, 98, 1625-1630.
123 E. Bieberich, S. MacKinnon, J. Silva, D. D. Li, T. Tencomnao, L. Irwin, D. Kapitonov and R. K. Yu, *Biochemistry*, 2002, 41, 11479-11487.
124 T. Kouno, Y. Kizuka, N. Nakagawa T. Yoshihara, M. Asano and S. Oka, *J. Biol. Chem.*, 2011, 286, 31337-31346.
125 W. Spessott, P. M. Crespo, J. L. Daniotti and H. J. Maccioni, *FEBS Lett.*, 2012, 586, 2346-2350.
126 C. Noffz, S. Keppler-Ross and N. Dean, *Glycobiology*, 2009, 19, 472-478.
127 A. Hassinen, F. M. Pujol, N. Kokkonen, C. Pieters, M. Kihlström, K. Korhonen and S. Kellokumpu, *J. Biol. Chem.*, 2011, 286, 38329-38340.
128 C. G. Giraudo and H. J. Maccioni, *J. Biol. Chem.*, 2003, 278, 40262-40271.
129 A. Seko and K. Yamashita, *Glycobiology*, 2005, 15, 943-951.

130 B. Ramakrishnan, E. Boeggeman, V. Ramasamy and P. K. Qasba, *Curr. Opin. Struct. Biol.*, 2004, 14, 593-600.
131 R. P. Aryal, T. Ju and R. D. Cummings, *J. Biol. Chem.*, 2012, 287, 15317-15329.
132 R. P. Aryal, T. Ju and R. D. Cummings, *J. Biol. Chem.*, 2014, 289, 11630-11641.
133 R. Wu and H. Wu, *J. Biol. Chem*., 2011, 286, 34923-34931.
134 L. Caputi, M. Malnoy, V. Goremykin, S. Nikiforova and S. Martens, *Plant J.*, 2012, 69, 1030-1042.
135 K. Yonekura-Sakakibara and K. Hanada, *Plant J.*, 2011, 66, 182-193.
136 E. K. Lim, S. Baldauf, Y. Li, L. Elias, D. Worrall, S. P. Spencer, R. G. Jackson, G. Taguchi, J. Ross and D. J. Bowles, *Glycobiology*, 2003, 13, 139-145.
137 K. Hashimoto, T. Tokimatsu, S. Kawano, A. C. Yoshizawa, S. Okuda, S. Goto and M. Kanehisa, *Carbohydr. Res.*, 2009, 344, 881-887.
138 I. Martinez-Duncker, R. Mollicone, J. J. Candelier, C. Breton and R. Oriol, *Glycobiology*, 2003, 13, 1C-5C.
139 P. L. DeAngelis, *Anat. Rec.*, 2002, 268, 317-326.
140 D. Liang and J. Qiao, *J. Mol. Evol.*, 2007, 64, 342-353.
141 J. O. Wrabl and N. V. Grishin, *J. Mol. Biol.*, 2001, 314, 365-374.
142 P. H. Chan, L. L. Lairson, H. J. Lee, W. W. Wakarchuk, N. C. Strynadka, S. G. Withers and L. P. Mclntosh, *Biochemistry*, 2009, 48, 11220-11230.
143 J. E. Pak, P. Arnoux, S. Zhou, P. Sivarajah, M. Satkunarajah, X. Xing and J. M. Rini, *J. Bio. Chem.*, 2006, 281, 26693-26701.
144 Y. Sun, F. Zeng, W. Zhang and J. Qiao, *Gene*, 2012, 499, 288-296.
145 S. J. Ahn, W. Dermauw, N. Wybouw, D. G. Heckel and T. Van Leeuwen, *Insect Biochem. Mol. Biol.*, 2014, 50, 43-57.
146 V. Kren and T. Rezanka, *FEMS Microbiol. Rev.*, 2008, 32, 858-889.
147 D. J. Newman and G. M. Cragg, *J. Nat. Prod*., 2012, 75, 311-335.
148 Y. Ogasawara, K. Katayama, A. Minami, M. Otsuka, T. Eguchi and K. Kakinuma, *Chem. Biol.*, 2004, 11, 79-86.
149 C. Fischer, F. Lipata and J. Rohr, *J. Am. Chem. Soc.*, 2003, 125, 7818-7819.
150 F. Kudo, T. Fujii, S. Kinoshita and T. Eguchi, *Bioorg. Med. Chem.*, 2007, 15, 4360-4368.
151 N. Stephens, B. Rawlings and P. Caffrey, *Biosci. Biotechnol. Biochem*., 2013, 77, 880-883.
152 Y. Rebets, B. Tokovenko, I. Lushchyk, C. Rückert, N. Zaburannyi, A. Bechthold, J. Kalinowski and A. Luzhetskyy, *BMC Genomics*, 2014, 15, 885
153 Y. Du, D. K. Derewacz, S. M. Deguire, J. Teske, J. Ravel, G. A. Sulikowski and B. O. Bachmann, *Tetrahedron*., 2011, 67, 6568-6575.
154 H. S. Kang and S. F. Brady, *Angew. Chem. Int. Ed. Engl*., 2013, 52, 11063-11067.
155 H. S. Kang and S. F. Brady, *J. Am. Chem. Soc*., 2014, 136, 18111-18119.
156 H. S. Kang and S. F Brady, *ACS Chem. Biol*., 2014, 9, 1267-1272.
157 G. Yim, L. Kalan, K. Koteva, M. N. Thaker, N. Waglechner, I. Tang and G. D. Wright, *Chembiochem*, 2014, 15, 2613-2623.
158 E. Sasaki, Y. Ogasawara and H. W. Liu, *J. Am. Chem. Soc*., 2010, 132, 7405-7417.

159 M. K. Kharel, S. E. Nybo, M. D. Shepherd and J. Rohr, *Chembiochem*., 2010, 11, 523-532.
160 K. Amagai, R. Takaku, F. Kudo and T. Eguchi, *Chembiochem*., 2013, 14, 1998-2006.
161 Y. Mast, T. Weber, M. Gölz, R. Ort-Winklbauer, A. Gondran, W. Wohlleben and E. Schinko, *Microb. Biotechnol*., 2011, 4, 192-206.
162 L.Kaysser, E. Wemakor, S. Siebenberg, J. A. Salas, J. K. Sohng, B. Kammerer and B. Gust, *Appl. Environ. Microbiol*., 2010, 76, 4008-4018.
163 S. H. Kim, J. H. Kim, B. Y. Lee and P. C. Lee, *Appl. Microbiol. Biotechnol*., 2014, 98, 9993-10003.
164 D. Chen, J. Feng, L. Huang, Q. Zhang, J. Wu, X. Zhu, Y. Duan and Z. Xu, *PLoS One*, 2014, 9, e108129.
165 Y. Zhang, H. Huang, Q. Chen, M. Luo, A. Sun, Y. Song, J. Ma and J. Ju, *Org. Lett.*, 2013, 15, 3254-3257.
166 M. Takaishi, F. Kudo and T. Eguchi, *J. Antibiot*. (Tokyo), 2013, 66, 691-699.
167 J. R. Lohman, S. X. Huang, G. P. Horsman, P. E. Dilfer, T. Huang, Y. Chen, E. Wendt-Pienkowski and B. Shen, *Mol. Biosyst*., 2013, 9, 478-491.
168 H. M. Ma, Q. Zhou, Y. M. Tang, Z. Zhang, Y. S. Chen, H. Y. He, H. X. Pan, M. C. Tang, J. F. Gao, S. Y. Zhao, Y. Igarashi and G. L. Tang, *Chem. Biol.*, 2013, 20, 796-805.
169 S. Li, J. Xiao, Y. Zhu, G. Zhang, C. Yang, H. Zhang, L. Ma and C. Zhang, *Org. Lett*., 2013, 15, 1374-1377.
170 Y. Ding, Y. Yu, H. Pan, H. Guo, Y. Li and W. Liu, *Mol. Biosyst*., 2010, 6, 1180-1185.
171 A. W. Truman, M. J. Kwun, J. Cheng, S. H. Yang, J. W. Suh and H. J. Hong, *Antimicrob. Agents Chemother*, 2014, 58, 5687-5695.
172 J. Li, Z. Xie, M. Wang, G. Ai and Y. Chen, *PLoS One*, 2015, 10, e0120542.
173 F. Kudo, T. Yonezawa, A. Komatsubara, K. Mizoue and T. Eguchi, *J. Antibiot. (Tokyo)*, 2011, 64, 123-132.
174 P. M. Flatt, X. Wu, S. Perry and T. Mahmud, *J. Nat. Prod*., 2013, 76, 939-946.
175 J. S. Chen, Y. X. Wang, L. Shao, H. X. Pan, J. A. Li, H. M. Lin, X. J. Dong and D. J. Chen, *Biotechnol. Lett.*, 2013, 35, 1501-1508.
176 T. Strobel, Y. Schmidt, A. Linnenbrink, A. Luzhetskyy, M. Luzhetska, T. Taguchi, E. Brötz, T. Paululat, M. Stasevych, O. Stanko, V. Novikov and A. Bechthold, *Appl. Environ. Microbiol.*, 2013, 79, 5224-5232.
177 H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel and R. Müller, *Chembiochem*., 2010, 11, 1840-1849.
178 T. Li, Y. Du, Q. Cui, J. Zhang, W. Zhu, K. Hong and W. Li, *Mar. Drugs*, 2013, 11, 466-488.
179 H. C. Nguyen, F. Karray, S. Lautru, J. Gagnat, A. Lebrihi, T. D. Huynh and J. L. Pernodet, *Antimicrob. Agents Chemother*, 2010, 54, 2830-2839.
180 Z. Guo, J. Li, H. Qin, M. Wang, X. Lv, X. Li and Y. Chen, *Angew. Chem. Int. Ed. Engl.*, 2015, 54, 5175-5178.
181 R. D. Kersten, A. L. Lane, M. Nett, T. K. Richter, B. M. Duggan, P. C. Dorrestein

and B. S. Moore, *Chembiochem*., 2013, 14, 955-962.
182 Y. Xiao, S. Li, S. Niu, L. Ma, G. Zhang, H. Zhang, G. Zhang, J. Ju and C. Zhang, *J. Am. Chem. Soc*., 2011, 133, 1092-1105.
183 J. Ren, Y. Cui, F. Zhang , H. Cui, X. Ni, F. Chen, L. Li and H. Xia, *Microbiol. Res*., 2014, 169, 602-608.
184 K. Xie, R. Chen, J. Li, R. Wang, D. Chen, X. Dou and J. Dai, *Org. Lett.*, 2014, 16, 4874-4877.
185 J. M. Augustin, S. Drok, T. Shinoda, K. Sanmiya, J. K. Nielsen, B. Khakimov, C. E. Olsen, E. H. Hansen, V. Kuzina, C. T. Ekstrøm, T. Hauser and S. Bak, *Plant Physiol*., 2012, 160, 1881-1895.
186 M. A. Naoumkina, L. V. Modolo, D. V. Huhman, E. Urbanczyk-Wochniak, Y. Tang, L. W. Sumner and R. A. Dixon, *Plant Cell*, 2010, 22, 850-866.
187 L.Dai, C. Liu, Y. Zhu, J. Zhang, Y. Men, Y. Zhang and Y. Sun, *Plant Cell Physiol.*, 2015, 56, 1172-1182.
188 A. Owatworakit, B. Townsend, T. Louveau, H. Jenner, M. Rejzek, R. K. Hughes, G. Saalbach, X. Qi, S. Bakht, A. D. Roy, S. T. Mugford, R. J. Goss, R. A. Field and A. Osbourn, *J. Biol. Chem.*, 2013, 288, 3696-3704.
189 K. Ghose, K. Selvaraj, J. McCallum, C. W. Kirby, M. Sweeney-Nixon, S. J. Cloutier, M. Deyholos, R. Datla and B. Fofana, *BMC Plant Biol.*, 2014, 14, 82.
190 Ruby, R. J. Santosh Kumar, R. K. Vishwakarma, S. Singh and B. M. Khan, *Mol. Biol. Rep.*, 2014, 41, 4675-4688.
191 M. Nagatoshi, K. Terasaka, A. Nagatsu and H. Mizukami, *J. Biol. Chem*., 2011, 286, 32866-32874.
192 W. W. Bhat, N. Dhar, S. Razdan, S. Rana, R. Mehra, A. Nargotra, R. S. Dhar, N. Ashraf, R. Vishwakarma and S. K. Lattoo, *PLoS One*, 2013, 8, e73804.
193 I. N. Van Bogaert, K. Holvoet, S. L. Roelants, B. Li, Y. C. Lin, Y. Van de Peer and W. Soetaert, *Mol. Microbiol*., 2013, 88, 501-509.
194 X. M. He and H. W. Liu, *Annu. Rev. Biochem.*, 2002, 71, 701-754.
195 J. L. Hansen, J. A. Ippolito, N. Ban, P. Nissen, P. B. Moore and T. A. Steitz, *Mol. Cell*, 2002, 10, 117-128.
196 F. Schlünzen , R. Zarivach, J. Harms, A. Bashan, A. Tocilj, R. Albrecht, A. Yonath and F. Franceschi, *Nature*, 2001, 413, 814-821.
197 Q. Vicens and E. Westhof, *J. Mol. Biol.*, 2003, 326, 1175-1188.
198 E. A. Owen, G. A. Burley, J. A. Carver, G. Wickham and M. A. Keniry, *Biochem. Biophys. Res. Commun.*, 2002, 290, 1602-1608.
199 R. E. Hibbs and E. Gouaux, *Nature*, 2011, 474, 54-60.
200 J. Lu, S. Patel, N. Sharma, S. M. Soisson, R. Kishii, M. Takei, Y. Fukuda, K. J. Lumb and S. B. Singh, *ACS Chem. Biol.*, 2014, 9, 2023-2031.
201 C. Temperini, L. Messori, P. Orioli, C. Di Bugno, F. Animati and G. Ughetto, *Nucleic Acids Res.*, 2003, 31, 1464-1469.
202 C. Temperini, M. Cirilli, M. Aschi and G. Ughetto, *Bioorg. Med. Chem.*, 2005, 13, 1673-1679.
203 R. J. Lewis, O. M. Singh, C. V. Smith, T. Skarzynski, A. Maxwell, A. J. Wonacott and D. B. Wigley, *EMBO J.*, 1996, 15, 1412-1420.

204 R. Pfoh, H. Laatsch and G. M. Sheldrick, *Nucleic Acids Res.*, 2008, 36, 3508-3514.
205 H. E. Mohamed, A. M. van de Meene, R. W. Roberson and W. F. Vermaas, *J. Bacteriol.*, 2005, 187, 6883-6892.
206 J. Xiao, T. S. Muzashvili and M. I. Georgiev, *Biotechnol. Adv.*, 2014, 32, 1145-1156.
207 C. Vilches, C. Hernandez C. Mendez and J. A. Salas, *J. Bacteriol.*, 1992, 174, 161-165.
208 L. M. Quirós, R. J. Carbajo, A. F. Braña and J. A. Salas, *J. Biol. Chem.*, 2000, 275,11713-11720.
209 N. Morisaki, Y. Hashimoto, K. Furihata, K. Yazawa, M. Tamura and Y. Mikami, *J. Antibiot. (Tokyo)*, 2001, 54, 157-165.
210 C. Walsh, C. L. Freel Meyers and H. C. Losey, *J. Med. Chem.*, 2003, 46, 3425-3436.
211 R. W. Gantt, R. D. Goff, G. J. Williams and J. S. Thorson, *Angew. Chem. Int. Ed. Engl.*, 2008, 47, 8889-8892.
212 L. Zhao, N. J. Beyer, S. A. Borisova and H. W. Liu, *Biochemistry*, 2003, 42, 14794-14804.
213 L. M. Quirós, I. Aguirrezabalaga, C. Olano, C. Méndez and J. A. Salas, *Mol. Microbiol.*, 1998, 28, 1177-1185.
214 T. W. Yu, L. Bai, D. Clade, D. Hoffmann, S. Toelzer, K. Q. Trinh, J. Xu, S. J. Moss, E. Leistner and H. G. Floss, *Proc. Natl. Acad. Sci. U S A*, 2002, 99, 7968-7973.
215 A. Gourmelen, M. H. Blondelet-Rouault, J. L. Pernodet, *Antimicrob. Agents Chemother*, 1998, 42, 2612-2619.
216 M. Brazier-Hicks, W. A. Offen, M. C. Gershater, T. J. Revett, E. K. Lim, D. J. Bowles, G. J. Davies and R. Edwards, *Proc. Natl. Acad. Sci. U S A*, 2007, 104, 20238-20243.
217 C. Dürr, D. Hoffmeister, S. E. Wohlert, K. Ichinose, M. Weber, U. Von Mulert, J. S. Thorson and A. Bechthold, *Angew. Chem. Int. Ed. Engl.*, 2004, 43, 2962-2965.
218 H. Wang, T. J. Oman, R. Zhang, C. V. Garcia De Gonzalo, Q. Zhang, W. A. van der Donk, *J. Am. Chem. Soc.*, 2014, 136, 84-87.
219 L. Li, P. Wang and Y. Tang, *J. Antibiot. (Tokyo)*, 2014, 67, 65-70.
220 A. Gutmann and B. Nidetzky, *Angew. Chem. Int. Ed. Engl.*, 2012, 51, 12879-12883.
221 T. J. Oman, J. M. Boettcher, H. Wang, X. N. Okalibe and W. A. van der Donk, *Nat. Chem. Biol.*, 2011, 7, 78-80.
222 C. D. Grubb, B. J. Zipp, J. Ludwig-Müller, M. N. Masuno, T. F. Molinski and S. Abel, *Plant J.*, 2004, 40, 893-908.
223 C. D. Grubb, B. J. Zipp, J. Kopycki, M. Schubert, M. Quint, E. K. Lim, D. J. Bowles, M. S. Pedras and S. Abel, *Plant J.*, 2014, 79, 92-105.
224 E. F. Marillia, J. M. MacPherson, E. W. Tsang, K. Van Audenhove, W. A. Keller and J. W. GrootWassink, *Physiol. Plant*, 2001, 13, 176-184.
225 C. Zhang, B. R. Griffith, Q. Fu, C. Albermann, X Fu, I. K. Lee, L. Li and J. S. Thorson, *Science*, 2006, 313, 1291-1294.

226 C. Zhang , Q. Fu, C. Albermann, L. Li and J. S. Thorson, *Chembiochem*., 2007, 8, 385-390.
227 A. Minami, K. Kakinuma and T. Eguchi, *Tetrahedron Lett*., 2005, 46, 6187-6190.
228 C. Zhang, C. Albermann, X. Fu and J. S. Thorson, *J. Am. Chem. Soc*., 2006, 128, 16420-16421.
229 C. Zhang, R. Moretti, J. Jiang and J. S. Thorson, *Chembiochem*., 2008, 9, 2506-2514.
230 R. Chen, H. Zhang, G. Zhang, S. Li, G. Zhang, Y. Zhu, J. Liu and C. Zhang, *J. Am. Chem. Soc.*, 2013, 135, 12152-12155.
231 L. V. Modolo, J. W. Blount, L. Achnine, M. A. Naoumkina, X. Wang and R. A. Dixon, *Plant Mol. Biol.*, 2007, 64, 499-518.
232 L. V. Modolo, L. L. Escamilla-Treviño, R. A. Dixon and X. Wang, *FEBS Lett.*, 2009, 583, 2131-2135.
233 C. Krauth, M. Fedoryshyn, C. Schleberger, A. Luzhetskyy and A. Bechthold, *Chem. Biol.*, 2009, 16, 28-35.
234 C. Leimkuhler, M. Fridman, T. Lupoli, S. Walker, C. T. Walsh and D. Kahne, *J. Am. Chem. Soc.*, 2007, 129, 10546-10550.
235 L. M. Garrido, F. Lombó, I. Baig, M. Nur-E-Alam, R. L. Furlan, C. C. Borda, A. Braña, C. Méndez, J. A. Salas, J. Rohr and G. Padilla, *Appl. Microbiol. Biotechnol.*, 2006, 73, 122-131.
236 G. Wang, P. Pahari, M. K. Kharel, J. Chen, H. Zhu, S. G. Van Lanen and J. Rohr, *Angew. Chem. Int. Ed. Engl.*, 2012, 51, 10638-10642.
237 A. Erb, A. Luzhetskyy, U. Hardter and A. Bechthold, *Chembiochem*., 2009, 10, 1392-1401.
238 S. A. Borisova, L. Zhao, C. E. Melançon III, C. L. Kao and H. W. Liu, *J. Am. Chem. Soc.*, 2004, 126, 6534-6535.
239 J. S. Hong, S. J. Park, N. Parajuli, S. R. Park, H. S. Koh, W. S. Jung, C. Y. Choi and Y. J. Yoon, *Gene*, 2007, 386, 123-130.
240 Y. Yuan, H. S. Chung, C. Leimkuhler, C. T. Walsh, D. Kahne and S. Walker, *J. Am. Chem. Soc.*, 2005, 127, 14128-14129.
241 C. E. 3rd Melançon, H. Takahashi and H. W. Liu, *J. Am. Chem. Soc.*, 2004, 126, 16726-16727.
242 M. Useglio, S. Peirú, E. Rodríguez, G. R. Labadie, J. R. Carney and H. Gramajo, *Appl. Environ. Microbiol.*, 2010, 76, 3869-3877.
243 U. Schell, S. F. Haydock, A. L. Kaja, I. Carletti, R. E. Lill, E. Read, L. S. Sheehan, L. Low, M. J. Fernandez, F. Grolle, H. A. Mcarthur, R. M. Sheridan, P. F. Leadlay, B. Wilkinson and S. Gaisser, *Org. Biomol. Chem.*, 2008, 6, 3315-3327.
244 W. Lu, C. Leimkuhler, G. J. Jr Gatto, R. G. Kruger, M. Oberthür, D. Kahne and C. T. Walsh, *Chem. Biol.*, 2005, 12, 527-534.
245 E. Rodríguez, S. Peirú, J. R. Carney and H. Gramajo, *Microbiology*, 2006, 152, 667-673.
246 H. Onaka, S. Taniguchi, Y. Igarashi and T. Furumai, *J. Antibiot. (Tokyo)*, 2002, 55, 1063-1071.
247 S. Gullón, C. Olano, M. S. Abdelfattah, A. F. Braña, J. Rohr, C. Méndez and J. A.

Salas, *Appl. Environ. Microbiol.*, 2006, 72, 4172-4183.
248 M. Zhou and J. S. Thorson, *Org. Lett.*, 2011, 13, 2786-2788.
249 P. Peltier-Pain, K. Marchillo, M. Zhou, D. R. Andes and J. S. Thorson, *Org. Lett.*, 2012, 14, 5086-5089.
250 L. Bungaruang, A. Gutmann and B. Nidetzky, *Adv. Synth. Catal.*, 2013, 355, 2757-2763.
251 S. H. Choi, M. Ryu, Y. J. Yoon, D. M. Kim and E. Y. Lee, *Biotechnol. Lett.*, 2012, 34, 499-505.
252 A. Gutmann, C. Krump, L. Bungaruang and B. Nidetzky, *Chem. Commun. (Camb)*, 2014, 50, 5465-5468.
253 W. Lu, C. Leimkuhler, M. Oberthür, D. Kahne and C. T. Walsh, *Biochemistry*, 2004, 43, 4548-4558.
254 A. Luzhetskyy, A. Mayer, J. Hoffmann, S. Pelzer, M. Holzenkämper, B. Schmitt, S. E. Wohlert, A. Vente and A. Bechthold, *Chembiochem.*, 2007, 8, 599-602.
255 W. Qin, Y. Liu, P. Ren, J. Zhang, H. Li, L. Tian and W. Li, *Chembiochem.*, 2014, 15, 2747-2753.
256 S. A. Borisova, L. Zhao, D. H. Sherman and H. W. Liu, *Org. Lett.*, 1999, 1, 133-136.
257 L. Rodríguez, I. Aguirrezabalaga, N. Allende, A. F. Braña, C. Méndez and J. A. Salas, *Chem. Biol.*, 2002, 9, 721-729.
258 M. D. Shepherd, T. Liu, C. Méndez, J. A. Salas and J. Rohr, *Appl. Environ. Microbiol*., 2011, 77, 435-441.
259 M. Oberthür, C. Leimkuhler, R. G. Kruger, W. Lu, C. T. Walsh and D. Kahne, *J. Am. Chem. Soc.*, 2005, 127, 10747-10752.
260 X. Fu, C. Albermann, J. Jiang, J. Liao, C. Zhang and J. S. Thorson, *Nat. Biotechnol*., 2003, 21, 1467-1469.
261 I. Baig, M. Perez, A. F. Braña, R. Gomathinayagam, C. Damodaran, J. A. Salas, C. Méndez and J. Rohr, *J. Nat. Prod.*, 2008, 71, 199-207.
262 C. L. Freel Meyers, M. Oberthür, J. W. Anderson, D. Kahne and C. T. Walsh, *Biochemistry*, 2003, 42, 4179-4189.
263 E. De Poire, N. Stephens, B. Rawlings and P. Caffrey, *Appl. Environ. Microbiol.*, 2013, 79, 6156-6159.
264 M. Doumith, R. Legrand, C. Lang, J. A. Salas and M. C. Raynal, *Mol. Microbiol.*, 1999, 34, 1039-1048.
265 C. Zhang, C. Albermann, X. Fu, N. R. Peters, J. D. Chisholm, G. Zhang, E. J. Gilbert, P. G. Wang, D. L. Van Vranken and J. S. Thorson, *Chembiochem.*, 2006, 7, 795-804.
266 M. Kopp, C. Rupprath, H. Irschik, A. Bechthold, L. Elling and R. Müller, *Chembiochem.*, 2007, 8, 813-819.
267 C. Olano, M. S. Abdelfattah, S. Gullón , A. F. Braña, J. Rohr, C. Méndez and J. A. Salas, *Chembiochem.*, 2008, 9, 624-633.
268 A. Minami, R. Uchida, T. Eguchi and K. Kakinuma, *J. Am. Chem. Soc*., 2005, 127, 6148-6149.
269 T. Moses, J. Pollier, L. Almagro, D. Buyst, M. Van Montagu, M. A. Pedreño, J. C.

Martins, J. M. Thevelein and A. Goossens, *Proc. Natl. Acad. Sci. U S A*, 2014, 111, 1634-1639.
270 A. R. Han, P. B. Shinde, J. W. Park, J. Cho, S. R. Lee, Y. H. Ban, Y. J. Yoo, E. J. Kim, E. Kim, S. R. Park, B. G. Kim, D. G. Lee and Y. J. Yoon, *Appl. Microbiol. Biotechnol*., 2012, 93, 1147-1156.
271 L. E. Núñez, S. E. Nybo, J. González-Sabín, M. Pérez, N. Menéndez, A. F. Braña, K. A. Shaaban, M. He, F. Morís, J. A. Salas, J. Rohr and C. Méndez, J. Med. Chem., 2012, 55, 5813-5825.
272 M. Pérez, F. Lombó, I. Baig, A. F. Braña, J. Rohr, J. A. Salas and C. Méndez, *Appl. Environ. Microbiol.*, 2006, 72, 6644-6652.
273 P. B. Shinde, A. R. Han, J. Cho, S. R. Lee, Y. H. Ban, Y. J. Yoo, E. J. Kim, E. Kim, M. C. Song, J. W. Park, D. G. Lee and Y. J. Yoon, *J. Biotechnol*., 2013, 168, 142-148.
274 M. Jiang, H. Zhang, S. H. Park, Y. Li and B. A. Pfeifer, *Metab. Eng.*, 2013, 20, 92-100.
275 J. Zhang, S. Singh, R. R. Hughes, M. Zhou, M. Sunkara, A. J. Morris and J. S. Thorson, *Chembiochem*., 2014, 15, 647-652.
276 Y. F. Yao, C. S. Wang, J. Qiao and G. R. Zhao, *Metab. Eng*., 2013, 19, 79-87.
277 B. W. Thuronyi and M. C. Chang, *Acc. Chem. Res*., 2015, 48, 584-592.
278 L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. 3rd Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, *Science*, 2008, 319, 1387–1391.
279 D. K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, N. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Chang, S. T. Withers, Y. Shiba, R. Sarpong and J. D. Keasling, *Natrue*, 2006, 440, 940-943.
280 C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, *et al.*, *Nature*, 2013, 496, 528-532.
281 Z. Dai, Y. Liu, X. Zhang, M. Shi, B. Wang, D. Wang, L. Huang and X. Zhang *Metabolic. Eng.*, 2013, 20,146-156.
282 X. Yan, Y. Fan, W. Wei, P. Wang, Q. Liu, Y. Wei, L. Zhang, G. Zhao, J. Yue and Z. Zhou, *Cell Res.*, 2014, 24, 770-773.
283 S. Brown, M. Clastre, V. Courdavault and S. E. O'Connor, *Proc. Natl. Acad. Sci. U S A*, 2015, 112, 3205-3210.
284 A. Loos and H. Steinkellner, *Arch. Biochem. Biophys.*, 2012, 526, 167-173.
285 A. Loos and H. Steinkellner, *Front Plant Sci.*, 2014, 5, 523.
286 T. Vogl, F. S. Hartner and A. Glieder, *Curr. Opin. Biotechnol.*, 2013, 24, 1094-1101.
287 L. Ronda, S. Bruno, S. Bettati, P. Storici and A. Mozzarelli, *Front Mol. Biosci.*, 2015, 2, 12.
288 J. A. Delmar, J. R. Bolla, C. C. Su and E. W. Yu, *Methods Enzymol.*, 2015, 557, 363-392.
289 H. Liu, R. Ye and Y. Y. Wang, *Genet. Mol. Res.*, 2015, 14, 3466-3473.
290 G. P. Flowers, A. T. Timberlake, K. C. McLean, J. R. Monaghan and C. M. Crews, *Development*, 2014, 141, 2165-1271.
291 U. R. Abdelmohsen, T. Grkovic, S. Balasubramanian, M. S. Kamel, R. J. Quinn

and U. Hentschel, *Biotechnol. Adv.*, 2015, doi:10.1016/j.biotechadv.2015.06.003.
292 B. O. Bachmann, S. G. Van Lanen and R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2014, 41, 175-184.
293 J. Lozano-Juste and S. R. Cutler, *Trends Plant Sci*., 2014, 19, 284-287.

Abbreviations

Terms	Abbreviations	Terms	Abbreviations
natural product	NP	protein O-fucosyltransferase	POFUT
glycosyltransferase	Gt	UDP-glycosyltransferase	UGT
quantum mechanics / molecular mechanics	QM/MM	*O*-GlcNAc transferase	OGT
transition state	TS	horizontal gene transfer	HGT
transmembrane	TM	oligosaccharyltransferase	OST

Tables

Table 1 NP Gts with solved 3D structures

Gt	NPs	Organism	PDB number	GT Family
GtfA[7]	chloroeremomycin	*Amycolatopsis orientalis*	1PN3,1PNV	GT-1
GtfB[8]			1IIR	
GtfD[9]	vancomycin		1RRV	
CalG1[10]	calicheamicin	*Micromonospora echinospora*	3OTG, 3OTH	GT-1
CalG2[10]			3IAA, 3RSC	
CalG3[10, 11]			3D0Q, 3D0R, 3OTI	
CalG4[10]			3IA7	
EryCIII[12]	erythromycin D	*Saccharopolyspora erythraea*	2YJN	GT-1
SpnP[13]	spinosyn	*Saccharopolyspora spinosa*	4LDP, 4LEI	GT-1
SpnG[14]			3TSA, 3UYK, 3UYL	
OleD[15]	oleandomycin	*Streptomyces antibioticus*	2IYF, 4M60, 4M7P, 4M83	GT-1
OleI[15]			2IYA	
UrdGT2[16]	urdamycin	*Streptomyces fradiae*	2P6P	GT-1
LanGT2[17]	landomycin	*Streptomyces cyanogenus*	4RIE, 4RIF	GT-1
VinC	vicenistatin	*Streptomyces halstedii*	3WAD, 3WAG	GT-1
SnogD[18]	nogalamycin	*Streptomyces nogalater*	4AMB, 4AMG, 4AN4	GT-1
SsfS6[19]	SF2575	*Streptomyces* sp. SF2575	4FZR, 4G2T	GT-1
UGT78K6[20, 21]	ternatin	*Clitoria ternatea*	3WC4, 4REM, 4REL, 4REN, 4WHM	GT-1
UGT71G1[22]	triterpene / flavonoid	*Medicago truncatula*	2ACV,2ACW	GT-1
UGT85H2[23]	(iso)flavonoids	*Medicago truncatula*	2PQ6	GT-1
UGT78G1[24]	(iso)flavonoids	*Medicago truncatula*	3HBF,3HBJ	GT-1
VvGT1[25]	anthocyanins	*Vitis vinifera*	2C1X, 2C1Z, 2C9Z	GT-1
AviGT4[26]	avilamycin A	*Streptomyces viridochromogenes*	2IUY, 2IV3	GT-4
VldE[27]	validamycin A	*Streptomyces hygroscopicus*	3T5T, 3T7D, 3VDM, 3VDN, 4F96, 4F97, 4F9F	GT-20
GtfAH1[28]	vancomycin /	--	3H4I, 3H4T	Chimeric Gt

LanGT2S8Ac[17]	teicoplanin landomycin	--	4RIG, 4RIH, 4RII	Chimeric Gt

Table 2 The classification of Gts

GT Family members	Inverting	Retaining
GT-A	2, 7, 12, 13, 14, 16, 21, 25, 29, 31, 40, 42, 43, 49, 82, 84	6, 8, 15, 24, 27, 34, 44, 45, 55, 60, 62, 64, 78, 81, 88
GT-B	1, 9, 10, 17, 19, 23, 26, 28, 30, 33, 41, 47, 52, 56, 63, 65, 68, 70, 80	3, 4, 5, 20, 32, 35, 52, 72
GT-C	22, 39, 48, 50, 53, 57, 58, 59, 66, 83, 85, 87	
Others	51 (Lysozyme-type)	
Unknown	11, 18, 37, 38, 54, 61, 67, 73, 74, 75, 76, 90, 92, 93, 94, 97	69, 71, 77, 79, 89, 95, 96

Note: The underlined numbers represent GT families with at least one solved 3D structure. Other families in GT-A/GT-B/GT-C were structurally predicted by Liu and Mushegian[29] and/or CAZy database. Notably, currently available information showed that members in GT52 were not uniform with inverting or retaining stereochemical conformations, and both the structure and stereochemistry of GT91 were unknown.

Table 3 New Gt structures identified in the last three years

Gt	Function	Organism	GT family	Structure	PDB number
EryCIII[12]	α-mycarosyl erythronolide B desosaminyltransferase	*Saccharopolyspora erythraea*	GT-1	GT-B	2YJN
SpnP[13]	spinosyn β-D-forosaminyltransferase	*Saccharopolyspora spinosa*	GT-1	GT-B	4LDP, 4LEI
SpnG[14]	spinosyn 9-O-α-L-rhamnosyltransferase				3TSA, 3UYK, 3UYL
LanGT2[17]	8-O-tetrangulol β-D-olivosyltransferase	*Streptomyces cyanogenus*	GT-1	GT-B	4RIE, 4RIF,
LanGT2S 8Ac[17]	Chimeric Gt	--	--	GT-B	4RIG, 4RIH, 4RII
VinC	vicenilactam β-vicenisaminyltransferase	*Streptomyces halstedii*	GT-1	GT-B	3WAD, 3WAG
SnogD[18]	nogalaminyltransferase	*Streptomyces nogalater*	GT-1	GT-B	4AMB, 4AMG, 4AN4
SsfS6[19]	tetracycline β-olivosyltransferase	*Streptomyces* sp. SF2575	GT-1	GT-B	4FZR, 4G2T
UGT78K6 [20, 21]	anthocyanidin 3-O-glucosyltransferase	*Clitoria ternatea*	GT-1	GT-B	3WC4, 4REM, 4REL, 4REN, 4WHM
GlfT2[49]	UDP-galactofuranosyl transferase	*Mycobacterium tuberculosis*	GT-2	GT-A	4FIX, 4FIY
BcsA[50]	cellulose synthase subunit A	*Rhodobacter sphaeroides*	GT-2	GT-A	4HG6
GalT1[51]	glucosyltransferase	*Streptococcus parasanguinis*	GT-2	GT-A	4PHR, 4PHS, 4PFX
CeGS[52]	glycogen synthase	*Caenorhabditis elegans*	GT-3	GT-B	4QLB
TarM[53]	teichoic acid α-N-acetylglucosaminyltransferase	*Staphylococcus aureus*	GT-4	GT-B	4X7M,4X6L, 4X7R,4X7P
TarM[54]	teichoic acid α-N-acetylglucosaminyltransferase	*Staphylococcus aureus*	GT-4	GT-B	4WAC, 4WAD

GtfA[55]	protein [Serine] α-N-acetylglucosaminyltransferase	*Streptococcus pneumoniae*	GT-4	GT-B	4PQG
SSI[56]	starch synthase I	*Hordeum vulgare*	GT-5	GT-B	4HLN
Gbss1[57]	starch synthase	*Oryza sativa Japonica* Group	GT-5	GT-B	3VUE, 3VUF
BoGT6a[58, 59]	2'-fucosyl lactose α-N-acetylgalactosaminyltransferase	*Bacteroides ovatus*	GT-6	GT-A	4AYJ, 4AYL, 4CJ8, 4CJB, 4CJC
GalT-I[60]	xylosylprotein β-1,4-galactosyltransferase-I / VII	*Homo sapiens*	GT-7	GT-A	4IRP, 4IRQ
Gyg2	glycogenin 2	*Homo sapiens*	GT-8	GT-A	4UEG
MurG[61]	N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	*Pseudomonas aeruginosa*	GT-28	GT-B	3S2U
ST6Gal-I[62]	β-galactoside α-2,6-sialyltransferase	*Homo sapiens*	GT-29	GT-A variant	4JS1, 4JS2
ST6Gal1[63]	β-galactoside α-2,6-sialyltransferase I	*Rattus norvegicus*	GT-29	GT-A variant	4MPS
KdtA	α-3-deoxy-D-manno-octulosonic-acid transferase	*Acinetobacter baumannii*	GT-30	GT-B	4BFC
KdtA[64]	α-3-deoxy-D-manno-octulosonic-acid transferase	*Aquifex aeolicus*	GT-30	GT-B	2XCI, 2XCU
GlgP	maltodextrin phosphorylase	*Streptococcus mutans*	GT-35	GT-B	4L22
AtPHS2[65]	α-glucan phosphorylase	*Arabidopsis thaliana*	GT-35	GT-B	4BQE, 4BQF, 4BQI
TcdA[66, 67]	toxin A / Rap2A α-glucosyltransferase	*Clostridium difficile*	GT-44	GT-A	3SRZ, 3SS1, 4DMV, 4DMW
Pbp4[68]	bifunctional glycosyltransferase/ acyltransferase penicillin-binding protein 4	*Listeria monocytogenes*	GT-51	Lysozyme-type	3ZG7, 3ZG8, 3ZG9, 3ZGA
ScMnn9[69]	mannan polymerase complexes subunit Mnn9	*Saccharomyces cerevisiae*	GT-62	GT-A	3ZF8
AglB-S1[70]	oligosaccharyltransferas	*Archaeoglobu*	GT-66	GT-C	3VGP

	e	*s fulgidus*			
AglB-S2[71]			GT-66	GT-C	3VU0
AglB-L[72, 73]	protein O-fucosyltransferase	*Pyrococcus horikoshii*	GT-66	GT-C	3WAI, 3WAJ, 3WAK
AglB-L[71]			GT-66	GT-C	3VU1
AglB-L		*Pyrococcus abyssi*	GT-66	GT-C	3WOV
POFUT2[74]		*Homo sapiens*	GT-68	GT-B	4AP5, 4AP6
PdST[75]	CMP-Neu5Ac α-2,3-sialyltransferase	*Pasteurella dagmatis*	GT-80	GT-B	4V2U, 4V38, 4V39, 4V3B, 4V3C
Bst[76]	β-galactoside α-2,6-sialyltransferase	*Photobacterium damselae*	GT-80	GT-B	4R83, 4R84, 4R9V
TibC[77]	autotransporter O-heptosyltransferase	*Escherichia coli*	NC	GT-B variant	4RAP, 4RB4
PaToxG[78]	Rho [tryrosine] N-acetylglucosaminyltransferase	*Photorhabdus asymbiotica*	NC	GT-A	4MIX

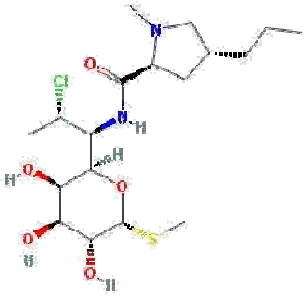
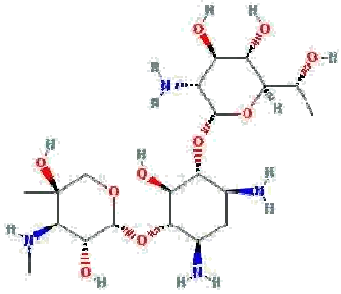
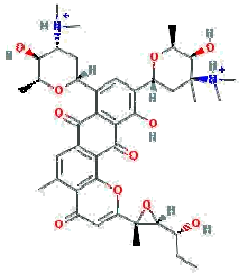
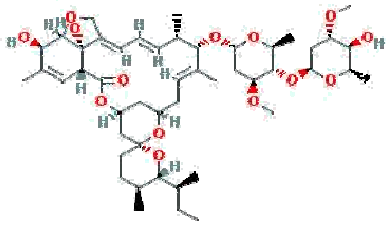
Table 4 NP Gts identified in recent five years

Gts	NPs	Organism	Experiment/Homology
AceDI, PegA[151]	67-121C	*Couchioplanes caeruleus*	E
AcuGT1-8 (AcuGT3[#])[152]	Aculeximycin	*Kutzneria albida* DSM 43870	H
ApoGT1, GT2 & GT3[153]	Apoptolidin	*Nocardiopsis sp*. FU40	H
Arm12, 25 & 33[154]	Arimetamycin A	uncultured bacterium	H
Arn11 & 14[155]	Arenimycin	uncultured bacterium	H
Arx3 & 9[156]	Arixanthomycin	uncultured bacterium	H
Auk10, 11 & 12[157]	UK-68,597	*Actinoplanes sp*. ATCC 53533	E
BexG1 & G2[158]	BE-7585A	*Amycolatopsis orientalis subsp. vinearia* BA-07585	E
ChryGT[159]	Chrysomycin	*Streptomyces albaduncus*	H
Clx40[155]	Calixanthomycin	uncultured bacterium	H
CmiM5[160]	Cremimycin	*Streptomyces sp*. MJ635-86F5	H
Cpp14, 27 and 29[161]	Unknown	*Streptomyces pristinaespiralis*	H
Cpz31[162]	Caprazamycin	*Streptomyces sp. MK730-62F2*	E
CrtX[163]	Astaxanthin dideoxyglycoside	*Sphingomonas sp*. PB304	H
EryBV* & CIII*[164]	Erythromycin	*Actinopolyspora erythraea*	H
GcnG1, G2 & G3[165]	Grincamycin	*Streptomyces lusitanus*	H
IdnS4 & S14[166]	Incednine	*Streptomyces sp*. ML694-90F3	H
KedS6 & S10[167]	Kedarcidin	*Streptoalloteichus sp*. ATCC 53650	H
KstD5[168]	Kosinostatin	*Micromonospora sp*. TP-A0468	H
LobG1, G2 & G3[169]	Lobophorin	*Streptomyces sp*. SCSIO 01127	E
NocS1[170]	Nocathiacin I	*Nocardia sp*. ATCC 202099	H
Orf16,17,18,20, 22 & 34[171]	Ristocetin	*Amycolatopsis lurida*	H
Pau15 & 25[172]	paulomycin	*Streptomyces paulus*	H
PnxGT1 & GT2[173]	FD-594	*Streptomyces sp*. TA-0256	H, E
PrlH[174]	Pyralomicin	*Nonomuraea spiralis* IMC	E

		A-0156	
Ramo-orf29[175]	Ramoplanin	*Actinoplanes sp.* ATCC 33076	E
RavGT[159]	Ravidomycin	*Streptomyces ravidus*	H
Ses60310[#176]	Different phenolic compounds[#]	*Saccharothrix espanaensis*	E
SpcG[178]	Indolocarbazole	*Streptomyces sanyensis* FMA	E
Srm5, 29 and 38[179]	Spiramycin	*Streptomyces ambofaciens*	E
StnG[180]	Streptothricin	*Streptomyces sp.* TP-A0356	E
Strop2213 & 2220[181]	Lomaiviticin	*Salinispora tropica* CNB-440	H
TiaG1 & G2[182]	Tiacumicin B	*Dactylosporangium aurantiacum subsp. hamdenensis*	E
TtmK[183]	Tetramycin	*Streptomyces ahygroscopicus*	H
UGT73AE1[184]	phloretin and resveratrol	*Carthamus tinctorius*	E
UGT73C10, C11, C12 & C13[185]	Sapogenins oleanolic acid and hederagenin	*Barbarea vulgaris*	E
UGT73F2 & F4[104]	Triterpenoid saponins	*soybean*	E
UGT73F3[186]	Hederagenin	*Medicago truncatula*	E
UGT74AC1[187]	Mogrosides	*Siraitia grosvenorii*	E
UGT74H5[188]	Avenacin	*Oat*	E
UGT74S1[189]	Secoisolariciresinol	*flax*	E
UGT74W1[190]	Sophoricoside	*Bacopa monniera*	E
UGT85A24[191]	Iridoid	*Gardenia jasminoides*	E
UGT86C4 & 94F4[192]	Iridoid	*Picrorhiza kurrooa*	E
UgtA1 & B1[193]	Sophorolipid	*Starmerella bombicola*	H

[#] may be not involved in the biosynthesis of secondary metabolites but the glycosylation of endogenous or exogenous natural products for self-protection

Table 5 Antibacterial mechanisms related to sugar moieties

Antibiotics	Antibacterial mechanisms
Clindamycin 	The three hydroxyl groups in the sugar moiety can form stable hydrogen-bonds with peptidyl transferase substrate.[196]
Geneticin 	The hydroxyl groups and the ammonium groups of the two sugar rings can form direct hydrogen bonds to base atoms and phosphate oxygen atoms of the A site (decoding aminoacyl-tRNA site) on 16S rRNA.[197]
Hedamycin 	The anglosamine and N, N-dimethylvancosamine both contact with substrate DNA, and the orientation and location of the N, N-dimethylvancosamine in the minor groove may determine the preference of binding sites.[198]
Ivermectin 	It stabilizes the open state of the ion channel through contacts between the disaccharide moiety and substrate.[199]
Kibdelomycin	One of the sugar moieties penetrates into the well-known ATP-binding site and is anchored to the lower binding site by three pairs of hydrogen bonds. The other sugar moiety makes contact with a surface area consisting of helix α4 and the flexible loop connecting

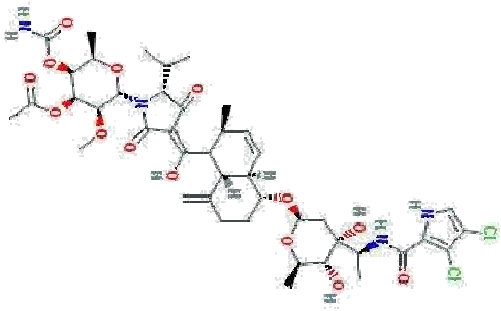
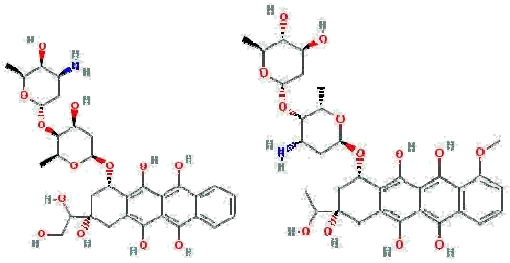
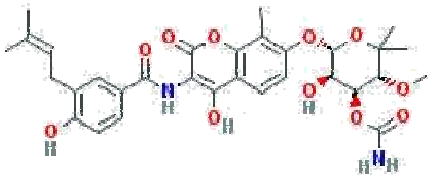
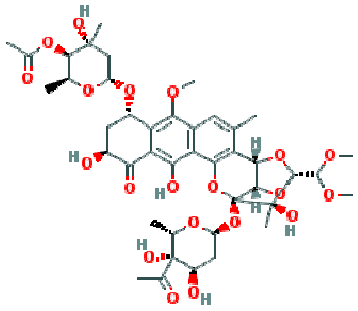
	helices α3 and α4.[200]
MEN 10755 and MAR70	The first sugar is an important substrate binding site, and the existence of amino group can enhance DNA binding affinity. The second sugar may govern the interactions with topoisomerases and other cellular targets to facilitate the formation of the ternary complex drug–DNA–topoisomerase.[201, 202]
	
Novobiocin	The 2'-hydroxyl group, 3'-carbamoyl group, 4'-methoxy methyl group and 5',-5'-dimethyl group can interact with the ATP binding site of the GyrB subunit either with hydrogen bonds or with hydrophobic contacts.[203]
	
Trioxacarcin A	The 3'-OH of the 4-sugar can form hydrogen bond with substrate DNA in the minor groove, and the 3''-OH of 13-sugar forms internal hydrogen bond with aglycon hydroxyl group to stabilize it.[204]
	

Table 6 Substrate specificities of flexible antibiotic Gts

Gts	Sugar specificity	Aglycon specificity
AknK[253]	certain sugar flexibility	monoglycosylated anthracyclines
AmiG[230]	moderate sugar flexibility	some aglycon tolerance
AmphDI[229]	some tolerance to aglycon structural diversity	stringent GDP-sugar specificity
AngMII[243]	certain sugar flexibility	structurally related aglycones
AraGT[254]	different NDP D-and L-sugars	--
AveBI*[228]	broad sugar flexibility	certain aglycon flexibility
BmmGT1[255]	--	certain aglycon flexibility
CalG1[225]	flexible to diverse TDP- D-and L-sugar donors	certain aglycon flexibility
CalG3[11]	broad sugar flexibility	--
CalG4[225]	flexible to diverse TDP- D-and L-sugar donors	certain aglycon flexibility
CosG[235]	some sugar flexibility	--
CosK[235]	some sugar flexibility	--
DesVII*[256]	some sugar flexibility	broad macrolide flexibility
ElmGt[257]	broad sugar flexibility	--
EryCIII[243]	certain sugar flexibility	structurally related aglycones
EryBV[226]	a small set of TDP-β-L-sugars	C5/C6 aglycon modifications
GilGT[258]	moderate sugar flexibility	stringent aglycon specificity
GtfA[259]	only utilizes donors closely related to its natural sugar substrate	vancomycin analogues
GtfC[259]	moderate sugar flexibility	vancomycin analogues
GtfD*[259]	broad sugar flexibility	vancomycin analogues
GtfE*[260]	moderate sugar flexibility	vancomycin analogues
MtmGIII[261]	some sugar donor tolerance	--
NovM[262]	broad sugar flexibility	a range of planar bicyclic aromatic compounds
NypY[263]	--	some aglycon flexibility
NysDI[229]	some tolerance to aglycon structural diversity	stringent GDP-sugar specificity
OleD*[15]	some sugar flexibility	broad aglycon flexibility
OleG2[264]	some sugar flexibility	some aglycon flexibility
OleI[15]	some sugar flexibility	only glycosylates oleandomycin
PegA[263]	--	some aglycon flexibility
RavGT[159]	broad sugar flexibility	coumarin-based polyketide derived backbone
RebG[265]	--	a set of indolocarbazole surrogates
Ses60310[176]	some sugar flexibility	a variety of phenolic compounds
SorF[266]	broad sugar flexibility	--

SpcG[178]	certain sugar flexibility	--
StaG[246]	certain sugar flexibility	--
StfG[267]	moderate sugar flexibility	stringent aglycon specificity
Thus[218]	certain sugar flexibility	certain promiscuity towards peptide substrates
TiaG1 & G2[182]	some sugar flexibility	--
TylMII[243]	certain sugar flexibility	structurally related aglycones
UrdGT2[217]	certain sugar flexibility	certain aglycon flexibility
VinC*[268]	utilize diverse NDP-sugars	structurally related and non-related compounds

*were utilized in "*in vitro* glycorandomization" strategy to generate diverse glycosylated antibiotics.

Fig. 1

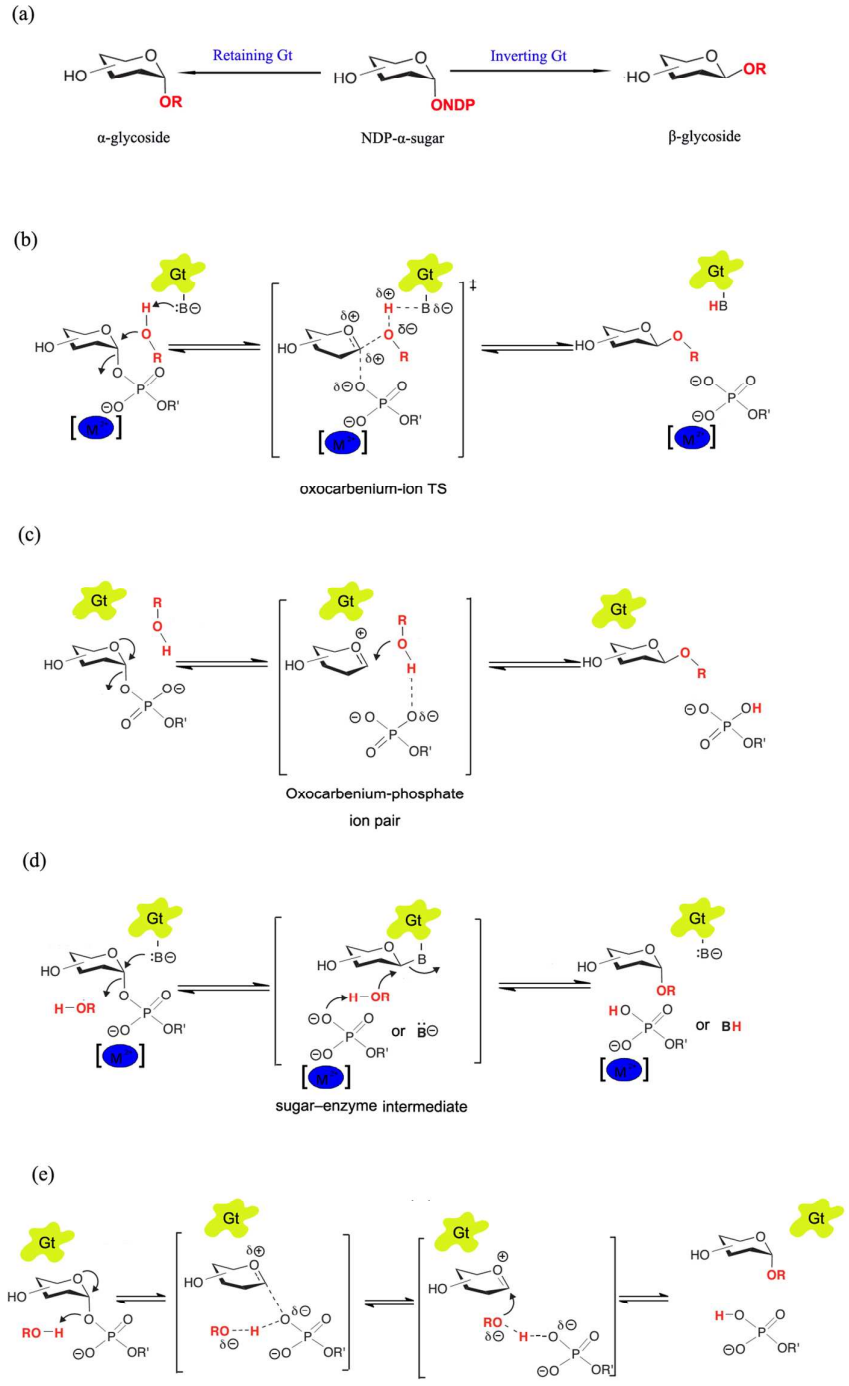


Fig. 1 The basic chemical reaction mechanisms of glycosylation. (a) Reaction schemes of inverting and retaining Gts. (b) Classical S_N2-like mechanism for inverting glycosylation. "B" represents the catalytic base provided by Gts, and the square brackets mean the divalent cations which are necessary for some inverting Gts

to stabilize the negative charge on leaving group, but not for others. (c) S_N1-like mechanism that may be utilized by inverting protein O-fucosyltransferase 1 (POFUT1) involving the formation of an oxocarbenium-phosphate ion pair. (d) Proposed double-displacement mechanism for retaining Gts. The proton abstraction from the acceptor OH-group may be performed by the leaving group phosphate or another catalytic base. (e) Proposed internal return (S_Ni-like) mechanism for retaining Gts.

Fig. 2

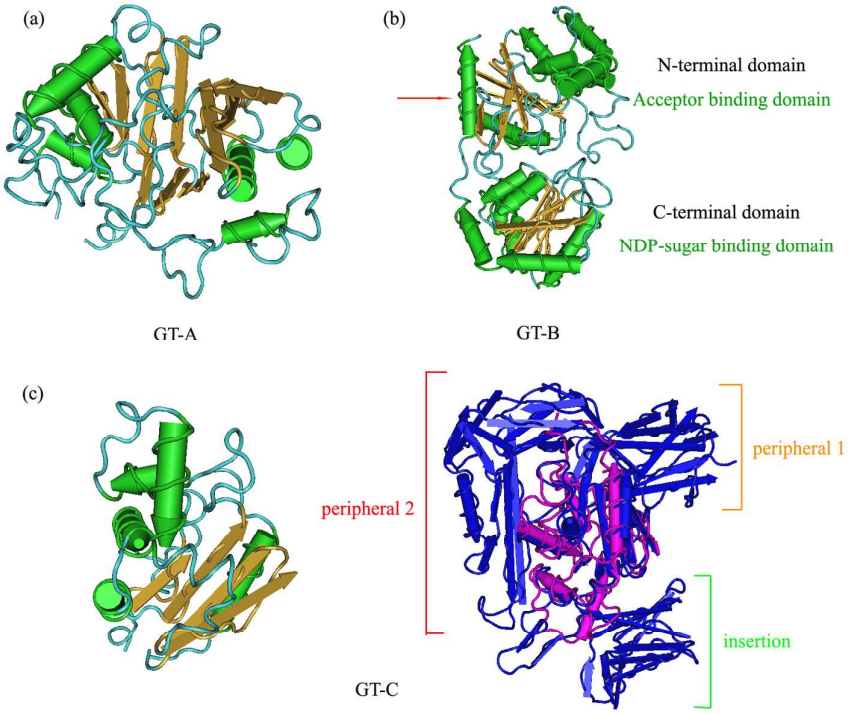


Fig. 2 Topology of GT-A, GT-B and GT-C proteins. (a) Schematic diagram of representative crystal structure of GT-A protein, which is represented by lipopolysaccharyl-α-1,4-galactosyltransferase C (LgtC) (GT-8, PDB number: 1SS9). (b) Schematic diagram of representative crystal structure of GT-B protein, which is represented by a NP Gt GtfA involved in the biosynthesis of chloroeremomycin (GT-1, PDB number: 1PN3). The region directed by a red arrow shows the C-terminal kinked α helix that extends out to interact with the N-terminal domain. (c) Schematic diagram of representative crystal structures of GT-C protein. The left is the 3D structure of AglB-S1 from *Archaeoglobus fulgidus* (GT-66, PDB number: 3VGP) with the shortest amino acid sequence and the simplest structure of STT3/AglB/PglB proteins, which can mainly be considered as the common structural unit; the right shows the structural alignment of *Archaeoglobus fulgidus* AglB-S1 (violet) and *Pyrococcus furiosus* AglB (blue), which has additional structural units (insertion, peripheral 1 and peripheral 2). All drawings were created using Cn3D.

Fig. 3

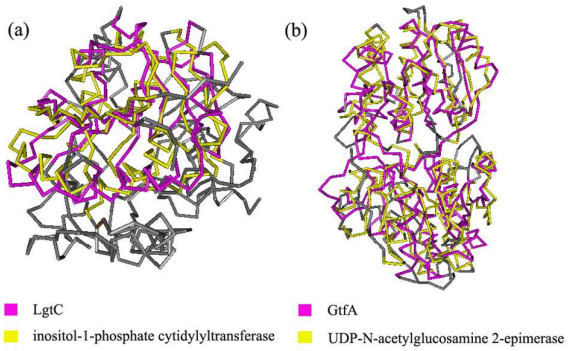


Fig. 3 Structural alignments of Gts and non-Gts with similar folds. (a) The structure of LgtC (GT-A fold, PDB: 1SS9) is partially matched by inositol-1-phosphate cytidylyltransferase (PDB: 4JD0). (b) The structure of UDP-N-acetylglucosamine 2-epimerase (PDB: 3OT5) have an overall topological similarity to GtfA (GT-B fold, PDB: 1PN3). The structural alignments were performed by Vector Alignment Search Tool (VAST, http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml), and the drawings were created using Cn3D.

Fig. 4

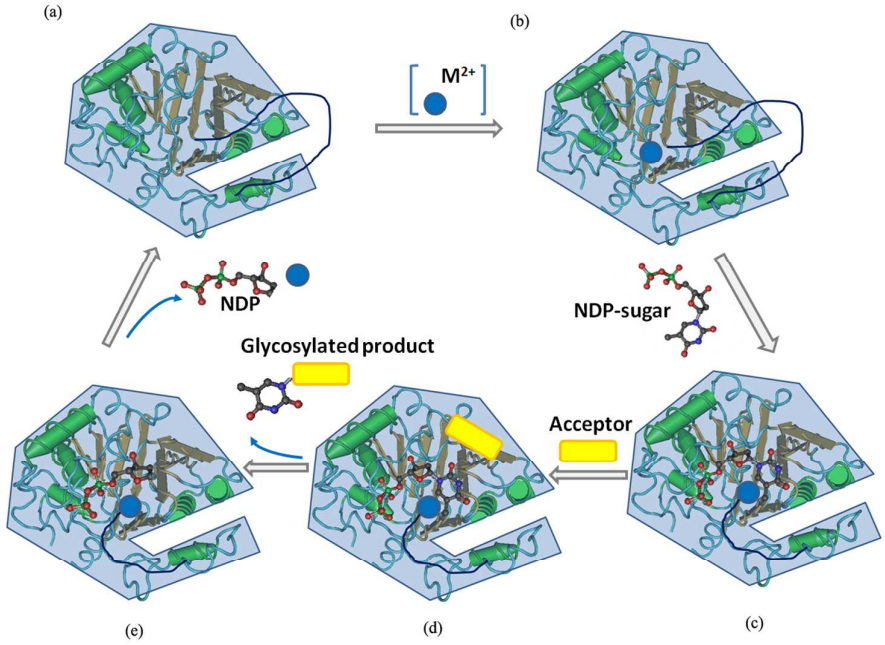


Fig. 4 The sequential bi–bi mechanism of glycosylation reactions with corresponding conformational changes. (a) and (b) shows the open conformation, and (c)-(e) means the closed conformation due to the movement of the flexible loop(s). The square brackets mean the divalent cations are not necessary for all Gts. The divalent cations will bind to Gts first when they are needed for reactions, while to Gts which do not need divalent cations, the first step will be the sugar donor binding (c).

Fig. 5

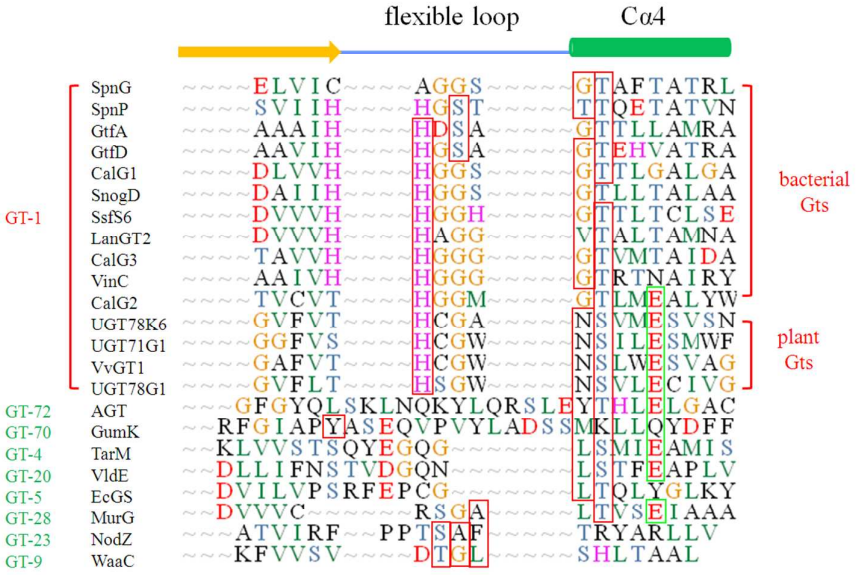


Fig. 5 The conserved "Gly-rich motif" in GT-B proteins. The Gts represented here are all crystallized in complex with NDP or NDP-sugar, and the residues forming hydrogen bonds with α/β-phosphate are highlighted in red boxes, as well as the residues interacting with ribose ring in green boxes. The flexible loop preceding the Cα4 helix is usually highly variable during the Gt conformational changes between open and closed conformations.

Fig. 6

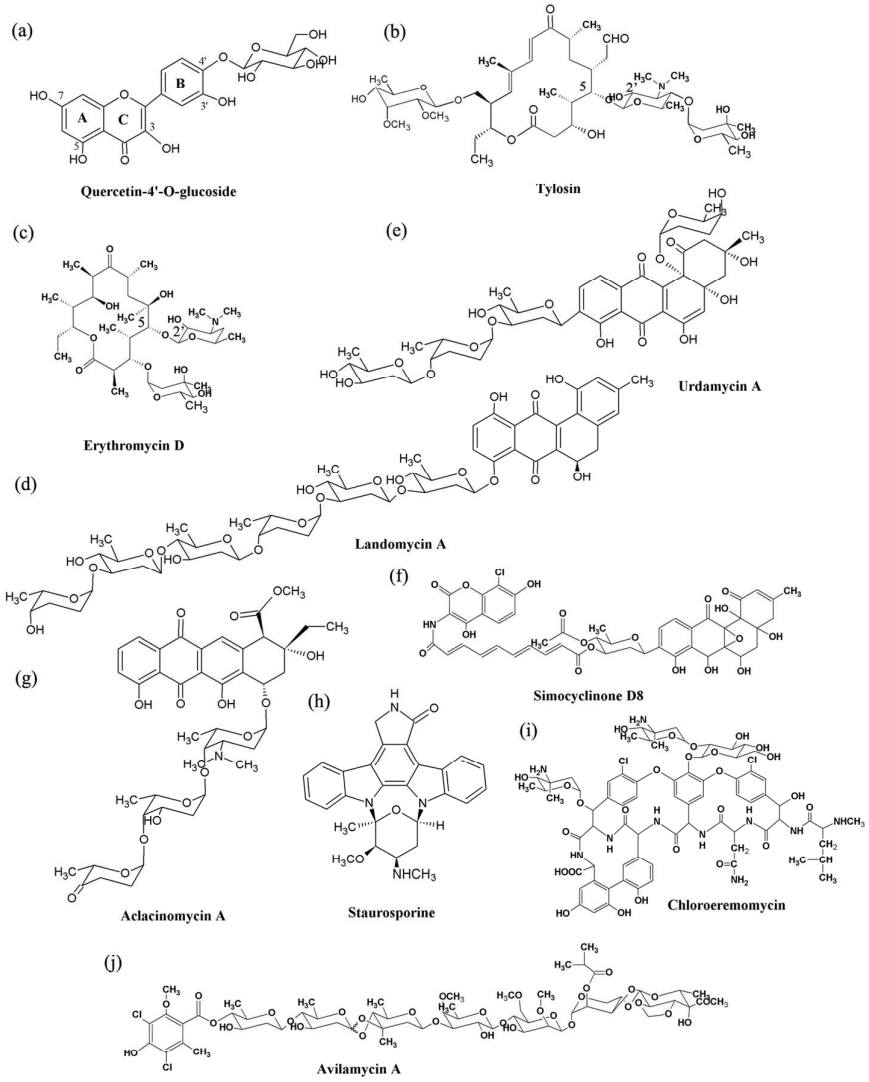


Fig. 6 The structurally diverse glycosylated NPs.

Fig. 7

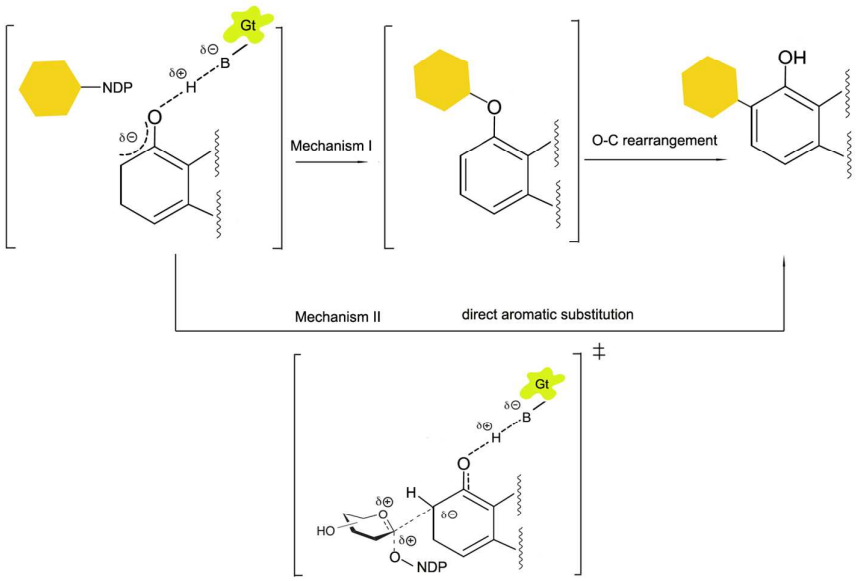


Fig. 7 The proposed C-glycosylation mechanisms.

Fig. 8

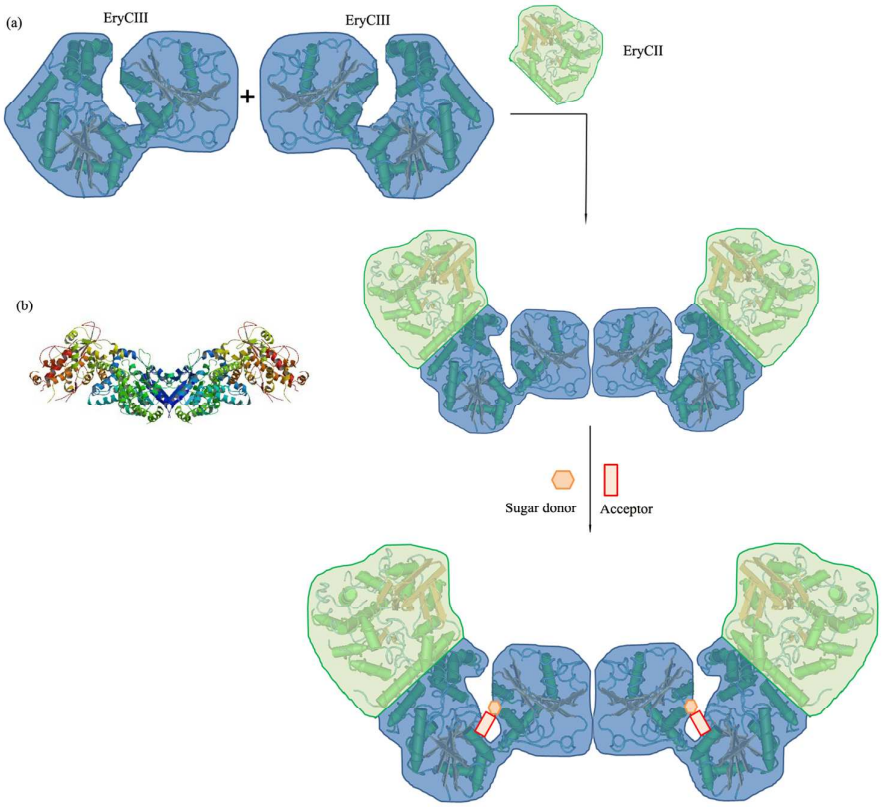


Fig. 8 The structure and proposed mechanism of EryCIII/ EryCII protein pair. (a) The possible reaction process of EryCII activated EryCIII glycosylation according to the structural studies. (b) The crystal structure of the EryCIII·EryCII complex.

Fig. 9

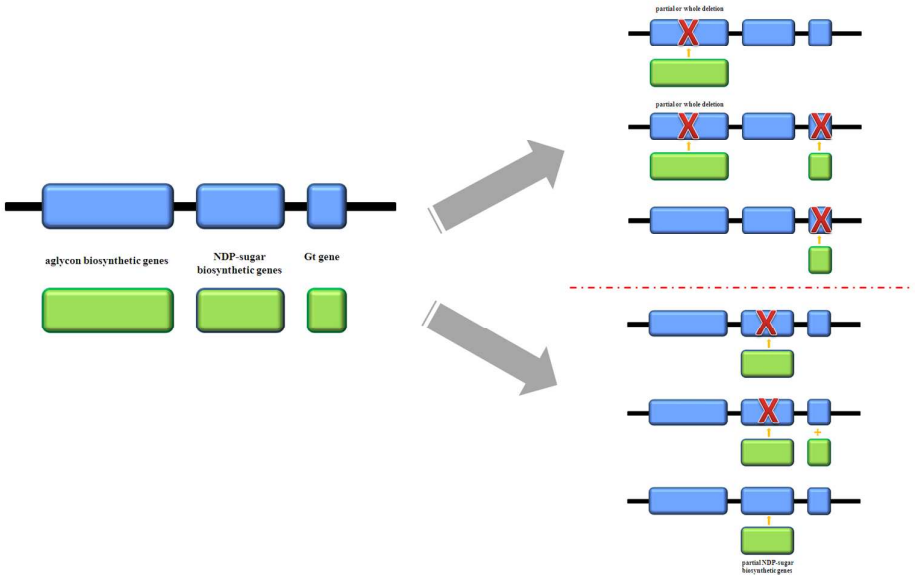


Fig. 9 The common strategies for combinational biosynthesis of glycosylated antibiotics. The upper half displays the methods based on the remaining of endogenous NDP-sugar biosynthetic genes, and the exogenous aglycon biosynthetic genes and genes encoding Gts can be heterologous expressed partially or totally; The lower half shows the methods based on the remaining of endogenous aglycon biosynthetic genes, and the exogenous NDP-sugar biosynthetic gene cassettes (with or without genes encoding Gts) can be also heterologous expressed.