



Reconstitution of antibiotics glycosylation by domain exchanged chimeric glycosyltransferase

Sung-Hee Park^a, Hyung-Yeon Park^b, Byung-Kwan Cho^a, Yung-Hun Yang^c, Jae Kyung Sohng^d, Hee Chan Lee^d, Kwangkyoung Liou^d, Byung-Gee Kim^{a,c,*}

^a Institute of Molecular Biology and Genetics, Interdisciplinary Program for Bioengineering, Seoul National University, Sillim-dong, Gwanak-gu, Seoul 151-742, Republic of Korea

^b Department of Chemistry, Inha University, Incheon 402-751, Republic of Korea

^c Institute of Bio Engineering, School of Chemical and Biological Engineering, Seoul National University, Sillim-dong, Gwanak-gu, Seoul 151-742, Republic of Korea

^d Institute of Biomolecule Reconstruction, SunMoon University, Chungnam 336-708, Republic of Korea

ARTICLE INFO

Article history:

Received 8 October 2008

Received in revised form 11 March 2009

Accepted 11 March 2009

Available online 24 March 2009

Keywords:

Antibiotics

Natural product

GT-B

Glycosyltransferase

Domain exchanging

Molecular evolution

ABSTRACT

Glycosyltransferase (GT), which catalyzes the transfer of nucleotide-activated sugar to aglycone, is the key enzyme for the preparation of a variety of glycosylated natural products. In this study, we report that the substrate specificity of GTs can be altered via domain exchanging accompanied by rationally designed peptide linkers. Comprehensive similarity searches were first conducted based upon the amino acid sequences of antibiotic GTs. The phylogenetic tree displays a distinct separation of 25 antibiotic GTs into three groups according to their aglycone substrates. Two different groups of GTs, KanF (aminoglycoside) and GtFE (glycopeptide) were employed for domain exchanging of GT-B fold enzyme. In order to determine linker peptide sequences, the secondary structures of KanF and GtFE were carefully evaluated via 3D-PSSM. The N-terminal domain of KanF and the C-terminal domain of GtFE were fused with rationally designed linker sequences. The hybrid GT, referred to as KE chimera, showed hybrid activity toward dTDP-D-glucose and 2-deoxystreptamine. Moreover, the key catalytic amino acid residues of KE chimera were investigated via protein modeling and docking simulations.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Natural products harboring sugar moieties are common sources of pharmaceutical or agricultural drug lead compounds [1]. Among them, therapeutically relevant antibiotics, including aminoglycosides, polyketides and glycopeptides harbor deoxy- or amino-sugars that are attached to the core aglycone structure [2]. The sugar units often participate in the specific recognition of their biological targets, and help to exhibit their inherent functions such as increasing the solubility, detoxification and antibacterial activity [3]. Therefore, there are burgeoning interests in the development of chemical or enzymatic techniques to alter glycosylation patterns in the natural products [4–11]. Glycosyltransferase (GT), which catalyzes the transfer of nucleotide-activated sugar onto aglycone, has been recognized as the key enzyme for the synthesis of a variety of glycosylated natural products [12]. Although combinatorial biosynthesis utilizing GTs with broad substrate specificities has shown

examples of remarkable success in the development of a variety of natural products, there still remain substantial limitations in resolving all the relevant cases depending on the substrate specificities of GTs [13].

Although molecular evolution approaches have been accomplished to enhance the specificity of GTs via random mutagenesis, directed evolution and domain swapping, GTs engineering remains a significant challenge due to the lack of high-throughput screening methods and crystallographic structure data. In the case of urdamycin GT, chimera construction was proceeded between UrdGT1b and UrdGT1c [14]. These two genes are highly homologous (>90%) in their amino acid sequences, so that they could determine the 13 crossover points at non-homologous regions. However, the glycosylated products were not newly generated, but the ratio of urdamycin G to urdamycin A, which are the authentic products, could be only altered. In the study of glycosylation of phenolic compounds with UGT72B1 [15], the crystal structure of UGT72B1 was first determined and used to generate 40 mutants via point mutation. The point mutation method is powerful, but requires the exact crystal structure of target enzyme. Moreover, only glucose was used as their sugar substrates. In a recent study of oleandomycin GT (OleD) [16], the evolution was conducted via EP-PCR, which requires an efficient HTS screening method. They used small fluo-

* Corresponding author at: Institute of Bio Engineering, School of Chemical and Biological Engineering, Seoul National University, Sillim-dong, Gwanak-gu, Seoul 151-742, Republic of Korea. Tel.: +82 2 880 6774; fax: +82 2 876 8945.

E-mail address: byungkim@snu.ac.kr (B.-G. Kim).

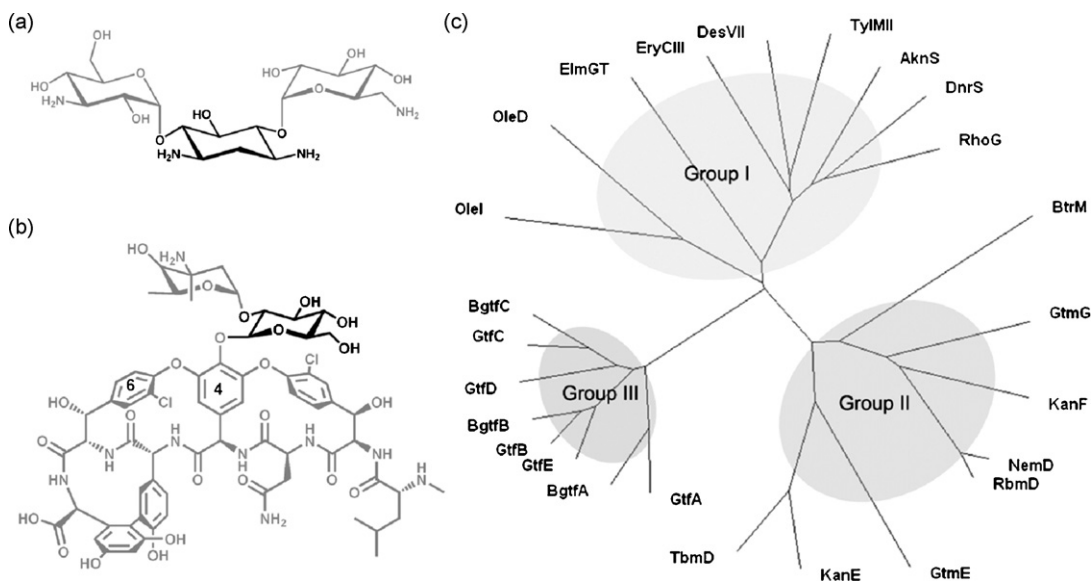


Fig. 1. (a) Structures of 2-deoxystreptamine (bold) in kanamycin, (b) glucose moiety (bold) in vancomycin, (c) phylogenetic analysis of antibiotic glycosyltransferases. (Polyketide Group I): Ole I (oleandomycin), OleD (oleandomycin), ElmGT (elloramyacin), EryCIII (erythromycin), DesVII (pikromycin), TyIMII (tylosin), AknS (aclerubicin), DnrS (doxorubicin) and RhoG (rhodomycin); (Aminoglycoside Group II): BtrM (butyrosin), GtmG (gentamycin), KanF (kanamycin), NemD (neomycin), RbmD (ribostamycin), GtmE (gentamycin), KanE (kanamycin), TbmD (tobramycin); (Glycopeptide Group III): GtfA (chloroeremomycin), BgtfA (balhimycin), GtfE (vancomycin), GtfB (chloroeremomycin), BgtfB (balhimycin), GtfD (vancomycin), GtfC (chloroeremomycin), BgtfC (balhimycin).

rescent compound as aglycone substrate for the HTS via quenching fluorescence. It is quite useful to screen active mutants, however, somewhat limited to generate hybrid oleandomycin with diverse sugars, because the aglycone of oleandomycin does not show fluorescence.

According to the structural and functional diversities of the natural products, GTs generated a marked diversity in their amino acid levels [17]. Indeed, there are more than 65 different families of GTs on the basis of their sequence homologies [18]. GTs are frequently classified into two major super-families due to their fold differences, GT-A and GT-B folds [18]. Almost all secondary metabolite synthetic GTs are classified into the GT family 1 of GT-B fold, which comprises two distinctive Rossmann-like domains linked by a flexible linker loop domain [19]. The C-terminal domain of GT-B folds harbors a binding site of the nucleotide-activated sugar donor located at the bottom of the interacting cleft, and the N-terminal domain performs an important function in aglycone acceptor binding [20,21]. The multiple alignments of a variety of antibiotic GTs revealed that the observed low homologies seem to arise from the changes in N-terminal domain sequences [20]. The hyper-variable motifs in the N-terminal sequences were known to be involved in accepting aglycone substrates. The C-terminal domain sequences harbor some highly conserved motifs, which contain key amino acids in the NDP-sugar binding catalysis. In result, we could postulate that the antibiotic GTs are classified according to their aglycone families on the basis of their sequence homology. Based on this special structural information of GTs, the idea of domain exchanging between the two different GTs has been hypothesized [12]. Then, to create various hybrid antibiotics by changing glycosylation patterns only, an optional approach is mix-and-match of the N- and C-terminal domains of GTs.

Here, two different groups of GTs, i.e. aminoglycoside KanF [22] and glycopeptide GtfE [23], were employed for domain exchanging in order to exchange the acceptor specificity of the newly designed enzyme. KanF is a putative GT in kanamycin (Fig. 1a) synthetic pathway from *Streptomyces kanamyceticus* and accepts 2-deoxystreptamine (2-DOS) as an aglycone substrate [24]. The sugar substrate of KanF has not been clearly verified yet; however, UDP-

N-actylglucosamine may be served as donor substrate, deduced from the structure of kanamycin and biosynthetic gene cluster [25]. GtfE is a well-known vancomycin GT (Fig. 1b), which utilizes dTDP-glucose as a sugar donor and heptapeptide acceptor of vancomycin [23,26]. In this study, we report that the domain exchanging of GT-B fold antibiotic GT can be possible by the use of rationally designed peptide linker. The N-terminal domain of KanF and the C-terminal domain of GtfE were fused with a rationally designed three-linker sequence. Using the most effective hybrid GT, referred to as the KE chimera, we attempted to determine the key amino acid residues for hybrid activity of domain-swapped GT-B enzyme via ligand docking simulation.

2. Materials and methods

2.1. Materials

Streptomyces kanamyceticus (ATCC 12853) and *Amycolatopsis orientalis* (ATCC19795) were used as the source strains for KanF and GtfE, respectively. *E. coli* strains DH5 α and BL21 (DE3) were used as the host strains for gene cloning and expression, respectively. All the restriction endo-nucleases and T4 DNA ligase were acquired from New England Biolabs. The pET28a expression vector was purchased from Novagen. Pfu DNA polymerase was purchased from Genenmed (Seoul, Korea). The DNA primers were obtained from the Custom Oligonucleotide Synthesis Manufacture Office (COSMO) (Seoul, Korea). dTDP-glucose, UDP-glucose and 2-deoxystreptamine were purchased from Genechem Inc. (Deajeon, Korea). All other materials were obtained from Sigma. The mass spectra were acquired using MS/MS LCQ DECA XP ion trap (ThermoFinnigan) equipped with Agilent 1100 HPLC. For the measurements of the UV absorbance of pH indicators, we used Multiskan Spectrum (Thermo Labsystems).

2.2. Sequence analysis

The antibiotic GT sequences in this study were aligned and calculated using CLUSTALX v1.81 [37] acquired from Université Louis Pasteur, Strasbourg. A neighbor-joining (NJ) phylogenetic analysis

of amino acid alignment was conducted using the modules of the PHYLIP software package v3.65.

2.3. 3D-PSSM

The protein folding structures of antibiotic GTs were analyzed using 3D-PSSM web server v 2.6.0 [30]. 3D-PSSM is a profile-based method relying on both multiple sequence and structural alignments. The key to this method is the “Three Dimensional Position Specific Scoring Matrix” (3D-PSSM), which combines the data from multiple sequence profiles obtained by PSI-BLAST with the structure-based profiles taking into account its secondary structure and solvent accessibility. The fold library for 3D-PSSM is predicated by the SCOP database [38], and included 9864 structures at the time when this study was undertaken. The 25 antibiotic GT sequences were submitted to the server by running a Perl script. 3D-PSSM scans submitted query sequences against its fold library and potential homologues were suggested. The results were sent to the indicated e-mail address and downloaded automatically. All the hits with *E*-values of less than 0.05 should be regarded as confident at 95% of confidence level.

2.4. Protein modeling

The SYBYL 6.4/COMPOSER module (TRIPOS Inc.) was utilized for structure modeling [39]. Knowledge-based modeling attempts to use what is already known about protein structures in general to predict specific cases. This approach includes searching techniques which attempt to fulfill selected criteria via the retrieval of protein fragments from a structural database. The docking of ligands into the active site of hybrid enzymes was conducted with FlexX (TRIPOS Inc.). The instance with the best FlexX score was

selected as the optimal conformational pose [40]. Dynamic hydrogen bonds were generated between the best-docked pose of a ligand and the amino acid residues in the active pocket of the protein.

2.5. KE chimera gene manipulation and enzyme preparation

A fragment harboring the N-domain of the *kanF* gene was amplified via PCR with the forward primer AGCGTTgaattcATGCGAGGTACAGATCTG and the reverse primer AGCGTTaagcttGCCCCGATCGGGTCCGAC. A fragment harboring the C-domain of the *gtfE* gene was amplified via PCR with the forward primer AAATACaagcttGAG-CGGCCGCTTCCCCGGAGATG and the reverse primer TGGTGActcgagTC-AGGCGGAACGGCGGGCTTCTC. The PCR conditions were as follows: 95 °C, 10 min for denaturation, 30 cycles of 95 °C, 1 min, 61 °C, 1 min, 72 °C 2 min for extension of DNA. Each PCR product was subcloned into T-vector to generate pTkanFN and pTgtfEC. After the confirmation of both sequences, an EcoRI-HindIII fragment derived from pTkanFN was inserted into pET28a to acquire pkanFN28. A HindIII-XhoI fragment derived from pTgtfEC was introduced into the HindIII-XhoI site of pkanFN28 in order to acquire the pKEchimera.

The plasmid pKEchimera harboring the hybrid GT gene was transformed into *E. coli* BL21 (DE3). The transformant was incubated in 300 ml LB medium containing selective marker (50 ppm kanamycin) at 37 °C. Expression was induced by the addition of 0.1 mM IPTG at OD₆₀₀ 0.6 for 4 h. The cells were harvested by 20 min of centrifugation at 4000 rpm at 4 °C and washed with 50 mM phosphate buffer (pH 7.0). The cells were suspended in 5 ml of 50 mM phosphate buffer (pH 7.5) and then disrupted via sonicator. After the removal of cell debris via 20 min of centrifugation at 15,000 rpm at 4 °C, the supernatant was loaded onto an Ni-NTA column and eluted

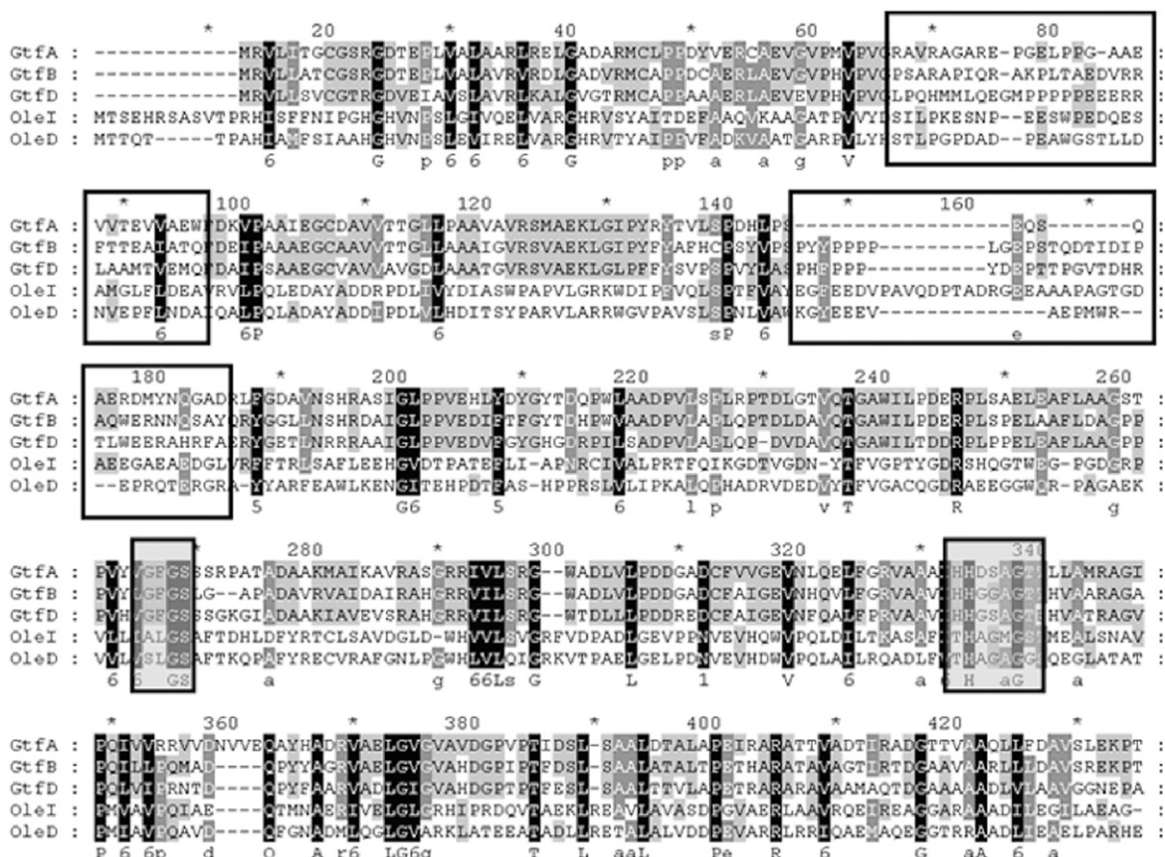


Fig. 2. Sequence alignment of glycosyltransferases. (□) Hyper-variable regions in the N-domain and (■) glycine-rich nucleotide binding motif in the C-domain.

with 200 mM imidazole-containing buffer (pH 7.5) after washing with 20 mM imidazole buffer.

For the longest linker loop (KanF+GtfE), the forward primer ACTGCTgaattc ATGCAGGTACAGATCCTG and the reverse primer TTGTCCAagctt CGCGCCTCGGCCGATC were utilized for the PCR amplification of the N-domain fragment and the forward primer TTGTCCAagcttATCCTCCCCGACGAGCGG and the reverse primer ACTGCTctcagTCAGGCGGGAACGGCGGG were utilized for the C-domain fragment. For the shortest linker design (GtfE), the forward primer ACTGCTgaattc ATGCAGGTACAGATCCTG and the reverse primer TTGTCCAagcttCACGTGATCACCTCTG were used for N-domain fragment and the forward primer TTGTCCAagcttATCCTCCCCGACGAGCGG and the reverse primer ACTGCTctcagTCAGGCGGGAACGGCGGG were used for the amplification of the C-domain fragment. Plasmid construction, transformation and protein preparation were conducted as described above.

2.6. GT assay

The calibration curve demonstrates the relationship between proton production and the absorbance of the pH indicator, cresol red. In a 1 mM phosphate buffer (pH 8.0) containing 0.05 mM cresol red, 10 mM dTDP-glucose, 6 µg purified enzymes and different quantities of hydrochloric acid were added to final concentrations of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1 mM, respectively, and absorbance at 436 nm was determined using Multiskan Spectrum (Thermo Labsystems). A quantitative linear relationship between the proton concentrations and absorbance was observed. The calibration curve was generated with the same enzyme mixture as that used in the actual assay only except the acceptor substrate. The activities of enzymes correlated with the proton concentration were calculated from the calibration curve. The initial velocities were fitted to the Michaelis–Menten equation, and kinetic parameters were calculated via Lineweaver–Burk plot of the velocity versus the substrate concentration. To compare activity, 0.05 mM cresol red, 10 mM dTDP-glucose and purified enzymes were mixed with phosphate buffer (50 mM, pH 8.0). The reaction was started by adding 2 mM 2-DOS. The absorbance at 436 nm was recorded for each sample in 96-well plate. All measurements were repeated in triplicate and averaged.

2.7. Product separation and identification

The pH of reaction mixture was adjusted to pH 2.0 with HCl, and it was loaded onto the OASIS MCX (Waters) SPE cleanup column which was pre-activated with 3 ml methanol and 3 ml water. The column was rewashed with 3 ml water and 3 ml methanol. The glucosylated products were extracted with 0.5 ml of 5% NH₄OH solution in methanol (v/v) and evaporated by vacuum centrifugation. The product was reconstituted with water and analyzed by HPLC-ESI-MS/MS.

3. Results

3.1. Classification of 25 antibiotic glycosyltransferases

The multiple alignment results of OleI, OleD, GtfA, GtfB and GtfD indicated that the C-terminal domain sequences were highly conserved, whereas the N-terminal sequences comprise a large variable region (Fig. 2). Two glycine-rich motifs conserved in the C-terminal domains were identified as the NDP-recognizing site [19] of the glycone substrate on the basis of their crystal structures and were well preserved in the majority of glyco-enzymes. Hyper-variability in the N-terminal domains recognizing aglycone substrate suggests that the antibiotic GTs may have diverged or co-evolved according to their aglycone moieties involved. Twenty-five antibiotic GTs

Table 1

3D-PSSM fold recognition result. (PDB ID: 1F6D), UDP-N-acetylglucosamine 2-epimerase in *E.coli*; (PDB ID: 1PNV), dTDP-epi-vancosaminyltransferase of chloroeremomycin; (PDB ID: 1IIR), UDP-glucosyltransferase of vancomycin; (PDB ID: 1RRV), dTDP-L-vancosaminyltransferase of vancomycin.

Class	GT	Antibiotics	3D-PSSM Hit (PDB code)	E-value
Polyketide	AknS	Aklarubicin	1RRV	2.93e–16
	ElmGT	Elloramycin	1IIR	2.44e–12
	RhoG	Rhodomyacin	1IIR	2.17e–12
	EryCIII	Erythromycin	1RRV	4.83e–16
	TylMI	Tylosin	1RRV	1.11e–17
	DnrS	Daunorubicin	1RRV	1.92e–15
	DesVII	Pikromycin	1RRV	8.40e–17
	OleG1	Oleandomycin	1RRV	4.95e–15
Aminoglycoside	BtrM	Butyrocycin	1F6D	1.64e–07
	NemD	Neomycin	1F6D	8.07e–06
	GtmG	Gentamycin	1F6D	1.11e–05
	GtmE	Gentamycin	1F6D	9.22e–06
	KanE	Kanamycin	1F6D	3.73e–07
	KanF	Kanamycin	1F6D	1.65e–06
	RbmD	Ribostamycin	1F6D	5.66e–06
Glycopeptide	GtfA	Chloroeremomycin	1PNV ^a	5.37e–50
	GtfB	Chloroeremomycin	1IIR ^a	
	GtfC	Chloroeremomycin	1RRV	
	GtfD	Vancomycin	1RRV ^a	4.15e–42
	GtfE	Vancomycin	1RRV	
	BgtfA	Balhumycin	1PNV	
	BgtfB	Balhumycin	1RRV	
	BgtfC	Balhumycin	1RRV	

^a GtfA, GtfB and GtfD are equal to 1PNV, 1IIR and 1RRV, respectively.

were classified into three groups using phylogenetic tree analysis. Nine polyketide, eight glycopeptide and eight aminoglycoside GT sequences were combined to visualize the evolutionary relationships between antibiotic GTs (Fig. 1c). Interestingly, the GTs were also classified in accordance with their aglycone structures of various antibiotics such as polyketide (Group I), aminoglycoside (Group II), and glycopeptide (Group III). Glycopeptide (Group III) GTs can be divided into few subgroups to reflect regio-specificity. GtfE, GtfB and BgtfB transferring glucose to the same position of 4-OH–Phegly₄ (Fig. 1b) in aglycone were clustered together [23,27,28]. GtfA and BgtfA evidenced identical regio-specificity on β–OH–Tyr₆ (Fig. 1b) and GtfD, GtfC and BgtfC shown to form an α–1,2-disaccharide at residue 4 of the heptapeptide scaffold [27,29]. All of the Group II sequences hit the UDP-N-acetylglucosamine 2-epimerase (1F6D) [30] as their protein fold template from 3D-PSSM (Table 1).

3.2. Linker prediction of KanF and GtfE

To determine the linker sequences for domain exchanging, the secondary structure of KanF and GtfE was investigated via 3D-PSSM (Table 1). The high homology (E-value of 4.15e–42) between GtfE and GtfD (PDB 1RRV) was strongly indicated that the GtfE was prominent GT-B family. Protein modeling was achieved on KanF to evidence GT-B fold enzyme (Supplementary Data) using 1F6D as a template. The model structure of KanF was composed of two α/β/α domains [31], which form a deep cleft at the domain interface. The C-terminal domain evidences the topology of the Rossmann dinucleotide binding fold which is surrounded by a group of α-helices. This result indicated that the aminoglycoside GT including KanF was also GT-B fold enzyme based on their hit template were all 1F6D. The flexible loop region was presented between two Rossmann fold connecting C-terminal regions of the N-domain and the N-terminal region of the C-domain. The linker sequence of the KanF was ₁₇₆PDVDPDRAEAA₁₈₇ and the GtfE was ₂₂₁ILPDERP₂₂₇ (Fig. 3).

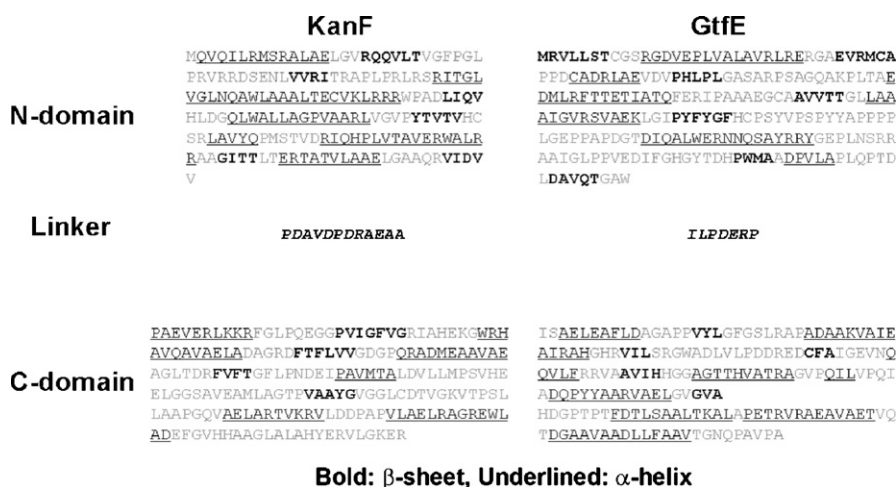


Fig. 3. Secondary structures and linkers of KanF and GtfE predicted by 3D-PSSM. Bold: β -sheet; Underline: α -helix.

3.3. Domain hybridization by linker design

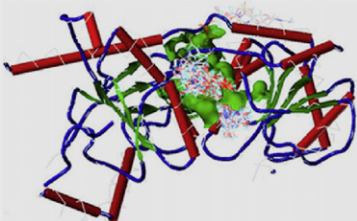
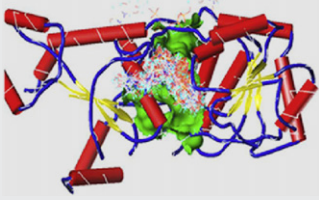
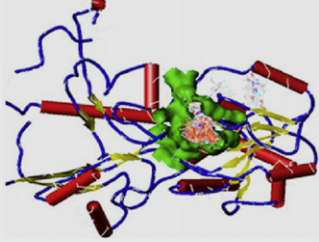
To design optimal linker loop domain for the domain exchanging, three linkers were constructed on the basis of the linker loop sequences of both GtfE and KanF, and the chimeric structures were subsequently assessed through structural modeling (Table 2a). The design with the shortest linker consisted of seven amino acids (ILPDERP) from the linker loop region of GtfE, resulting chimera enzyme with a collapsed active site and an improperly folded N-terminal domain. The design with the longest linker loop (PDAVDPDRAEAAILPDERP) consisted of sequentially fused sequences from both linker loop sequences of KanF and GtfE. The model predictions of this chimera displayed that its N-terminal

domain mainly makes a β -sheet structure, not forming a representative Rossmann-like nucleotide binding conformation. The most plausible linker loop was the hybrid linker loop (PDAVDPDRAERP), which consists of the front nine amino acids (Pro176–Ala184) from the linker loop region of KanF and followed by three amino acids (Glu225–Pro227) from the linker loop region of GtfE.

The hybrid enzymes with the three linker types were over-expressed in *E. coli* and the crude cell extracts were used for the comparison of GT activities for 2-DOS. We applied a pH-sensitive galactosyltransferase assay [32] to do quantitative analysis of the enzyme activity. This method was quite useful to detect glycosylation reaction for non-fluorescent acceptor, such as 2-deoxystreptamine and glycosides, and showed a 50% higher

Table 2

Structure modeling of three hybrid enzymes harboring different representative linker sequences, and the visualization of ligand distribution in the binding site.

Linker origin	Linker sequence	Model	Relative activity
KE chimera	PDAVDPDRAERP		100
KanF+GtfE	PDAVDPDRAEAAILPDERP		87
GtfE	ILPDERP		38

Linker origin: source proteins of linker sequences; relative activity: glucosyl transfer activity of hybrid enzymes toward 2-DOS; KE chimera activity was defined as 100.

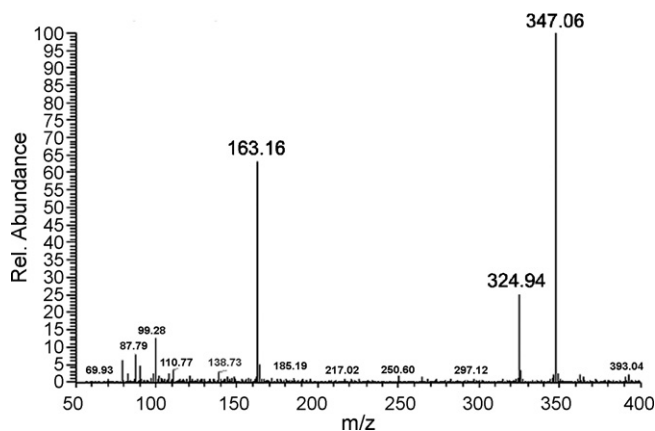


Fig. 4. ESI-MS spectrum for the isolated product. 2-Deoxystreptamine (2-DOS) is found at m/z 163.16 $[M+H]^+$ and mono-glucosylated 2-DOS is found at m/z 324.94 $[M+H]^+$ and at 347.06 $[M+Na]^+$, respectively.

sensitivity than the most common HPLC assay [32]. This technique is based on the detection of UV absorbance changes of pH indicator caused by the O-linked glycosidic bond formation and accompanied proton release. The alteration of absorbance was detected via enzymatic reactions on each of the linker models as described in Section 2. The hybrid GT assay results are somewhat consistent with the modeling results in that the enzyme harboring the hybrid peptide linker loop (PDAVDPDRAERP) resulted the best relative activity (Table 2b), suggesting that the design of the loop region is critical for obtaining good chimeric GT activity.

3.4. Characterization of KE chimera

In order to elucidate the *in vitro* activity and kinetic parameters of the KE chimera (M.W. 40 kDa), it was His₆ tagged and successfully over-expressed in *E. coli*, then purified. To determine the kinetic parameters, the pH-sensitive GT assay method was used as described in Section 2. The HCl calibration curve for the pH-assay and the product formation rate according to the substrate concentration (2-DOS) was measured to determine the kinetic parameters as presented in Supplementary Data. When 2-DOS concentration was varied by fixing the concentration of dTDP-glucose **1** at 10 mM, the k_{cat}/K_m value of KE chimera was estimated as $43 \pm 1 \text{ mM}^{-1} \text{ min}^{-1}$. Mono-glucosylated product was extracted using cation-exchange column and subjected to ESI-MS to verify their molecular weights (Fig. 4). The aglycone substrate 2-DOS was detected at m/z 163.16 $[M+H]^+$ and the mono-glucosylated product was detected at m/z 324.94 $[M+H]^+$ and at 347.06 $[M+Na]^+$. To compare the NDP-sugar specificity of KE chimera with wild type KanF, UDP-N-acetylglucosamine was used as sugar substrate of enzyme reaction. However, any product peak was not detected in ESI-MS (data not shown).

Multiple alignments of KE chimera with vancomycin group GT sequences (GtfA, GtfB, GtfC, GtfD and GtfE) revealed several conserved key residues for the recognition of substrates in the KE chimera (Supplementary Data). These conserved residues were exactly identical to the key residues identified by the ligand docking simulation (Fig. 5). The highly conserved Arg10 in the N-terminal domain of Gtfs and KE chimera was responsible for the interaction with the hexose and the reactive hydroxyl group of the corresponding aglycone. The Arg10 also interacts with Glu251 of the C-terminal domain, and discharges capping the binding pocket, corresponding to the previously described in MurG catalysis [33]. The His₂₆₆ constructed the "HHXXAGT" loop [19], which was highly conserved Gly-rich sequence and a representative binding site of the α -phosphate group of the NDP-sugar donor in GtfA [21].

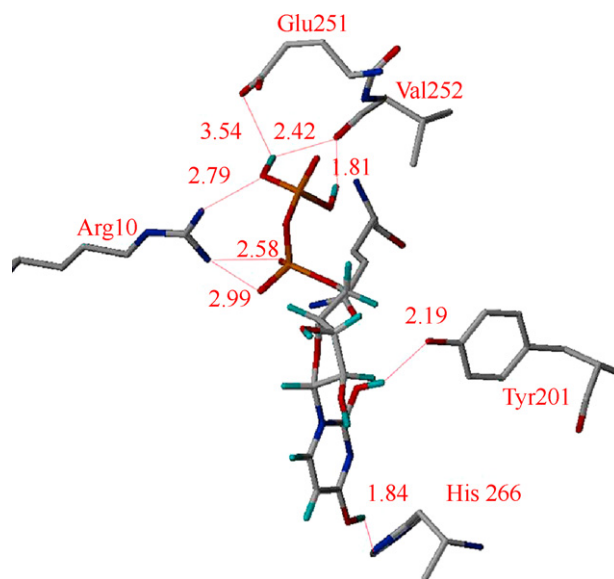


Fig. 5. Active pocket and docking of KE chimera. Binding interactions of NDP with the residues of the N-terminal and C-terminal domains including the key residues.

4. Discussion

There are five published crystal structures of antibiotic glycosyltransferases in the vancomycin series GtfA, GtfB, and GtfD, and the polyketide series, OleI and OleD [20,21,34,35]. From the multiple sequence alignment results of these five GTs, their N-terminal domains were not well conserved compared to the C-terminal domains. Antibiotic GTs showed distinct evolutionary lineages in accordance with their aglycone substrate structures. The evolutionary lineages of the aminoglycoside GTs (Group II) not falling into the GT-1 family [29] were also investigated. According to the results of 3D-PSSM [30] in Table 1, GTs in Group I and Group III are likely to have similar protein structures to the antibiotic GTs in Group III given as template for the comparison, i.e. GtfA (1PNV), GtfB (1IIR) and GtfD (1RRV). Whereas all of the Group II sequences hit UDP-N-acetylglucosamine 2-epimerase (1F6D) as their protein fold template. The function of 1F6D is the supply of an activated sugar donor (UDP-GlcNAc) to bacterial cell, which is critical for the biosynthesis of cell surface polysaccharides [36]. Thus, it is reasonable to imagine that aminoglycoside GT might originate from the non-GT enzyme responsible for primary metabolism. This result is quite agreeable with the recent evolutionary study of phylogenetic analysis of antibiotic GTs [29]. Among the aminoglycoside GTs, the structure of kanamycin GT (KanF) was structurally studied. The first report of the structural analysis of KanF in this study showed that this aminoglycoside GT has a representative GT-B fold, although their ancestral proteins may not be derived from GT.

In order to select model enzymes evidencing different substrate specificities, KanF (aminoglycoside) and GtfE (glycopeptide) were selected for the construction of a domain-swapped hybrid enzyme, referred to as KE chimera. The KE chimera comprising the N-terminal domain of KanF and the C-terminal domain of GtfE was constructed under the expectation that the KE chimera acquires a new ability to transfer glucose from dTDP-glucose to 2-DOS. The expected structure of the products was pseudo-disaccharide often detected in naturally occurring 2-DOS glycosylated products such as neamine, paromamine, lividamine and gentamine [24]. The secondary structures of each enzyme were predicted by 3D-PSSM, and their linker sequences were extracted from the whole amino acid sequences. In order to optimize the domain-swapped hybrid enzyme, three hybrid enzymes were compared via protein model-

ing. Among the three models, only the KE chimera linker sequence in Table 2 proved to be effective for the formation of the Rossmann-like dinucleotide binding GT-B fold hybrid enzyme.

To compare their real activities, the quantitative GT assay was performed using a pH indicator, i.e. cresol red. Cresol red was successfully confirmed to change the color of the colonies on LB agar plate owing to the pH decrease during the O-glycosidic bond formation. This method was especially useful to detect glycosylation of non-fluorescent acceptors of natural products including aminoglycosides and glycosides [32]. By measuring the absorbance change, the glycosylation activities of the three model proteins were determined. Consistent with the modeling predictions, the activities of real enzymes were revealed that the KE chimera sequences showed the highest levels of glucosylation activity toward 2-DOS. From these results, we postulated that one of the key elements of domain exchanging is the length of the linker sequences.

In conclusion, we have successfully demonstrated a domain exchanging strategy between two different groups of antibiotic GTs. The KE chimera harboring the sequence of the N-terminal domain of the KanF-linker-C-terminal domain of GtfE was utilized for the synthesis of glycosylated 2-deoxystreptomycin. The determination of an appropriate and optimal length of the linker peptide was an essential for the successful GT-B fold domain exchanging. This approach may offer a new avenue on the evolutionary study of GT-B fold enzymes including antibiotic GTs, and overcoming major limitations of GT engineering via random mutagenesis to change GT specificities for the combinatorial synthesis of natural products.

Acknowledgments

This work was supported by Ministry of Commerce, Industry and Energy (grant no. 10023194) and by a grant (Code # KRF-2005-005-J16002) from Korea Research Foundation, Republic of Korea.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molcatb.2009.03.006.

References

- [1] D.J. Newman, G.M. Cragg, K.M. Snader, *J. Nat. Prod.* 66 (2003) 1022–1037.
- [2] G. Lancini, R. Lorenzetti, *Biotechnology of Antibiotics and Other Bioactive Microbial Metabolites*, Plenum Press, New York, 1993.
- [3] A.C. Weymouth-Wilson, *Nat. Prod. Rep.* 14 (1997) 99–110.
- [4] P.F. Leadlay, *Curr. Opin. Chem. Biol.* 1 (1997) 162–168.
- [5] H.G. Floss, *J. Ind. Microbiol. Biotechnol.* 27 (2001) 183–194.
- [6] C. Mendez, J.A. Salas, *Trends Biotechnol.* 19 (2001) 449–456.

- [7] X. Fu, C. Albermann, J. Jiang, J. Liao, C. Zhang, J.S. Thorson, *Nat. Biotechnol.* 21 (2003) 1467–1469.
- [8] J. Yang, D. Hoffmeister, L. Liu, X. Fu, J.S. Thorson, *Bioorg. Med. Chem.* 12 (2004) 1577–1584.
- [9] C. Rupprath, T. Schumacher, L. Elling, *Curr. Med. Chem.* 12 (2005) 1637–1675.
- [10] C.E. Melancon III, C.J. Thibodeaux, H.W. Liu, *ACS Chem. Biol.* 1 (2006) 499–504.
- [11] J.A. Salas, C. Mendez, *Trends Microbiol.* 15 (2007) 219–232.
- [12] C. Walsh, C.L. Freil Meyers, H.C. Losey, *J. Med. Chem.* 46 (2003) 3425–3436.
- [13] H.G. Floss, *J. Biotechnol.* 124 (2006) 242–257.
- [14] D. Hoffmeister, K. Ichinose, A. Bechthold, *Chem. Biol.* 8 (2001) 557–567.
- [15] M. Brazier-Hicks, W.A. Offen, M.C. Gershtater, T.J. Revett, E.K. Lim, D.J. Bowles, G.J. Davies, R. Edwards, *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007) 20238–20243.
- [16] G.J. Williams, C. Zhang, J.S. Thorson, *Nat. Chem. Biol.* 3 (2007) 657–662.
- [17] J.A. Campbell, G.J. Davies, V. Bulone, B. Henrissat, *Biochem. J.* 326 (Pt 3) (1997) 929–939.
- [18] P.M. Coutinho, E. Deleury, G.J. Davies, B. Henrissat, *J. Mol. Biol.* 328 (2003) 307–317.
- [19] Y. Hu, S. Walker, *Chem. Biol.* 9 (2002) 1287–1296.
- [20] A.M. Mulichak, H.C. Losey, C.T. Walsh, R.M. Garavito, *Structure* 9 (2001) 547–557.
- [21] A.M. Mulichak, H.C. Losey, W. Lu, Z. Wawrzak, C.T. Walsh, R.M. Garavito, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 9238–9243.
- [22] M.K. Kharel, B. Subba, D.B. Basnet, J.S. Woo, H.C. Lee, K. Liou, J.K. Sohng, *Arch. Biochem. Biophys.* 429 (2004) 204–214.
- [23] H.C. Losey, M.W. Pecuh, Z. Chen, U.S. Eggert, S.D. Dong, I. Pelczar, D. Kahne, C.T. Walsh, *Biochemistry* 40 (2001) 4745–4755.
- [24] I. Hooper, Y. Ito, T. Koeda, S. Kondo, S. Mitsunashi, T. Okuda, N. Tanaka, T. Tsuchiya, K. Umemura, H. Umezawa, S. Umezawa, M. Yokota, *Aminoglycoside Antibiotics*, Springer-Verlag, Berlin, Heidelberg, New York, 1982.
- [25] L.P. Thapa, T.J. Oh, H.C. Lee, K. Liou, J.W. Park, Y.J. Yoon, J.K. Sohng, *Appl. Microb. Biotechnol.* 76 (2007) 1357–1364.
- [26] H.C. Losey, J. Jiang, J.B. Biggins, M. Oberthur, X.Y. Ye, S.D. Dong, D. Kahne, J.S. Thorson, C.T. Walsh, *Chem. Biol.* 9 (2002) 1305–1314.
- [27] R.D. Sussmuth, W. Wohlleben, *Appl. Microbiol. Biotechnol.* 63 (2004) 344–350.
- [28] S. Pelzer, R. Sussmuth, D. Heckmann, J. Recktenwald, P. Huber, G. Jung, W. Wohlleben, *Antimicrob. Agents Chemoth.* 43 (1999) 1565–1573.
- [29] D. Liang, J. Qiao, *J. Mol. Evol.* 64 (2007) 342–353.
- [30] L.A. Kelley, R.M. MacCallum, M.J. Sternberg, *J. Mol. Biol.* 299 (2000) 499–520.
- [31] Z. Zhang, S. Kochhar, M. Grigorov, *Protein Sci.* 12 (2003) 2291–2302.
- [32] C. Deng, R.R. Chen, *Anal. Biochem.* 330 (2004) 219–226.
- [33] Y. Hu, L. Chen, S. Ha, B. Gross, B. Falcone, D. Walker, M. Mokhtarzadeh, S. Walker, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 845–849.
- [34] A.M. Mulichak, W. Lu, H.C. Losey, C.T. Walsh, R.M. Garavito, *Biochemistry* 43 (2004) 5170–5180.
- [35] D.N. Bolam, S. Roberts, M.R. Proctor, J.P. Turkenburg, E.J. Dodson, C. Martinez-Fleites, M. Yang, B.G. Davis, G.J. Davies, H.J. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007) 5336–5341.
- [36] R.E. Campbell, S.C. Mosimann, M.E. Tanner, N.C. Strynadka, *Biochemistry* 39 (2000) 14993–15001.
- [37] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, *Nucleic Acids Res.* 25 (1997) 4876–4882.
- [38] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, *J. Mol. Biol.* 247 (1995) 536–540.
- [39] A.S. Ivanov, V.S. Skvortsov, A.I. Archakov, *Vop. Med. Khim.* 46 (2000) 615–625.
- [40] J.K. Buolamwini, H. Assefa, *J. Med. Chem.* 45 (2002) 841–852.

Further reading

- [41] <ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX>.
- [42] <http://www.sbg.bio.ic.ac.uk/~3dpssm>.