Engineering Microbiology 4 (2024) 100134

Contents lists available at ScienceDirect

Engineering Microbiology

journal homepage: www.elsevier.com/locate/engmic

Original Research Article

Rational design of a highly active *N*-glycosyltransferase mutant using fragment replacement approach

Jiangyu Yang[a,#], Kun Li[a,#], Yongheng Rong[a], Zhaoxi Liu[a], Xiaoyu Liu[a], Yue Yu[a], Wenjing Shi[b], Yun Kong[b], Min Chen[a,*]

[a] State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, China
[b] National Glycoengineering Research Center, Shandong University, Qingdao 266237, China

ARTICLE INFO

Keywords:
N-glycosyltransferase
Computational design
Glycoprotein
Fragment replacement
GT41

ABSTRACT

The modularity of carbohydrate-active enzymes facilitates that enzymes with different functions have similar fragments. However, because of the complex structure of the enzyme active sites and the epistatic effects of various mutations on enzyme activity, it is difficult to design enzymes with multiple mutation sites using conventional methods. In this study, we designed multi-point mutants by fragment replacement in the donor-acceptor binding pocket of *Actinobacillus pleuropneumoniae N*-glycosyltransferase (ApNGT) to obtain novel properties. Candidate fragments were selected from a customized glycosyltransferase database. The stability and substrate-binding energy of the three fragment replacement mutants were calculated in comparison with wild-type ApNGT, and mutants with top-ranking stability and middle-ranking substrate-binding energy were chosen for priority experimental verification. We found that a mutant called F13, which increased the glycosylation efficiency of the natural substrate by 1.44 times, the relative conversion of UDP-galactose by 14.2 times, and the relative conversion of UDP-xylose from almost 0 to 78.6%. Most importantly, F13 mutant acquired an entirely new property, the ability to utilize UDP-glucuronic acid. On one hand, this work shows that replacing similar fragments in the donor-acceptor binding pocket of the enzyme might provide new ideas for designing mutants with new properties; on the other hand, F13 mutant is expected to play an important role in targeted drug delivery.

1. Introduction

Glycosylation reactions are critical for many biological processes. For example, protein glycosylation affects protein degradation. Heavily glycosylated proteins can be challenging for proteases to cleave because of charge alterations and steric hindrance caused by glycans [1,2]. Glycosylation also affects protein function. For example, the *N*-glycosylation status of the Ig-Fc region significantly affects the half-life and effector function of an antibody [3]. Glycosyltransferases are the primary enzymes involved in protein glycosylation.

Carbohydrate-active enzymes catalyze the synthesis, degradation, and modification of carbohydrates and their derivatives, such as bacterial exopolysaccharides, starch, cellulose, and lignin, as well as the glycosylation of proteins and lipids [4]. Glycosyltransferases have 116 families that catalyze many different reactions. Since the substrate adaptability of a particular glycosyltransferase is typically limited, it is necessary to find a new method to expand substrate adaptability and broaden the use of glycosyltransferases.

In the past few years, significant progress has been made in enzyme engineering using sequence- or structure-guided rational approaches [5,6]. For example, by utilizing multiple sequence alignment and structural analysis, researchers can gain valuable insights into the key residues located in the active sites of glycosyltransferases, which can be leveraged to enhance substrate recognition [7,8]. However, the complex structure of the enzyme active sites and the epistatic effect of mutation sites on enzyme activity make it difficult for researchers to design enzymes with multiple mutation sites using classical rational methods [9,10]. The epistatic effect is unavoidable when combining different single-point mutations. When different single-point mutations are combined and exist in a mutant, their effect on protein characteristics is not always additive. Such mutational effects that are not additive suggest that an epistatic effect is at work.

Abbreviations: ApNGT, *Actinobacillus pleuropneumoniae N*-glycosyltransferase; GT41, glycosyltransferase family 41; LB, Luria-Bertani; PDB, Protein Data Bank; RP-HPLC, reversed-phase high-performance liquid chromatography; TLC, thin layer chromatography; UDP-Gal, UDP-galactose; UDP-GlcNAc, UDP-N-acetyl-D-glucosamine; UDP-Glc, UDP-glucose; UDP-Man, UDP-mannose; UDP-Xyl, UDP-xylose; UDP-GlcA, UDP-glucuronic acid.

* Corresponding author.
E-mail address: chenmin@sdu.edu.cn (M. Chen).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.engmic.2023.100134
Received 13 June 2023; Received in revised form 15 November 2023; Accepted 16 November 2023
Available online 30 November 2023

Previously, a multi-point mutation design method was developed based on phylogenetic analysis and Rosetta energy calculations [11–14]. First, multiple sequence alignments based on retained homologs and selected mutation sites with moderate frequency in nature. Second, using the Rosetta atomic model, mutation sites that could severely damage protein stability were identified and removed. Third, all mutants with 3–5 mutation sites were sorted by protein energy, and mutants with lower energy were selected for experimental verification. However, this method also had some limitations. The amino acids near the active center were conserved, which limited the sequence space of the mutants, therefore rendering creation of mutants with new functionality challenging. Compared with other methods, especially *de novo* protein design, our method had an improved success rate. This is beneficial for glycosyltransferases, which are difficult to screen with high throughput. The modularity of carbohydrate-active enzymes facilitates that enzymes with different functions have similar fragments [4]. This study focused on *Actinobacillus pleuropneumoniae N*-glycosyltransferase (ApNGT), an *N*-β-glucosyltransferases (EC:2.4.1.-) that belongs to the glycosyltransferase family 41 (GT41). Using UDP-glucose (UDP-Glc, optimal substrate), UDP-galactose (UDP-Gal), and UDP-xylose (UDP-Xyl) as sugar donors, wild-type ApNGT could glycosylate the acceptor peptide including the N-X-S/T (X≠Pro) sequence [15–18]. Moreover, ApNGT has shown potential for application in glycoengineering [19–21].

We used fragments from other glycosyltransferases to replace sections close to the donor-acceptor binding pocket of ApNGT. Combinatorial mutants with multiple fragment mutations were constructed. To select suitable fragments, we evaluated candidate fragments based on their charge properties, similarity, and number of mutated amino acid residues. Furthermore, we ensured that the candidate fragments did not involve direct changes to active-site residues. The stability of all combined mutants was calculated using the Rosetta atomic model, and the top-ranking mutants were selected for experimental verification.

In conclusion, we constructed a glycosyltransferase mutant (F13), which was more efficient than wild-type ApNGT in producing glycosylated proteins. The F13 mutant displayed enhanced catalytic efficiency, a broader peptide substrate utilization spectrum, and the ability to utilize the sugar donor UDP-GlcA, which was unavailable to wild-type ApNGT.

2. Materials and methods

2.1. Materials

ApNGT (provided in the Supplementary Materials) was synthesized by GenScript Co., Ltd. Competent cells of *Escherichia coli* DH5α and *Escherichia coli* BL21 (DE3) were purchased from Sangon Biotech (Shanghai) Co., Ltd. The multi-point mutation kit was purchased from Vazyme Biotech Co., Ltd. UDP-N-acetyl-D-glucosamine (UDP-GlcNAc), UDP-Glc, and UDP-GlcA were purchased from Sigma-Aldrich Co., Ltd. UDP-Gal was purchased from Macklin Co., Ltd. UDP-mannose (UDP-Man) and UDP-Xyl were purchased from Shanghai Yuanye Biotechnology Co., Ltd. The designed substrate peptide was synthesized by GenScript Co., Ltd.

2.2. Computational design of mutants

Mutants were created using a method called “fragment replacement”. To screen for mutants with superior characteristics, this method replaced fragments in the wild-type protein and calculated characteristics of the mutants. A Python script (provided in the Supplementary Materials) was used to design the mutants. The following paragraph (illustrated in Fig. 1) explains the main logic of the Python script.

First, the structures of several glycosyltransferases were randomly selected from the Protein Data Bank (PDB) database. These structures were split into fragments for replacement during the subsequent processing. For example, to replace a tripeptide fragment, the downloaded glycosyltransferase can be randomly split into tripeptide fragments of the same size. The fragments obtained after splitting constituted the sequence space.

Next, seven loop regions (Figure S1; Table S3) near the donor-acceptor binding pocket in ApNGT (PDB ID: 3Q3E) were selected. Fragment replacement involved replacing some of these loop regions with amino acid fragments of equal length. Subsequently, according to the pre-selected loop regions and several glycosyltransferases, it could be determined which fragments of equal length in each loop region could be replaced. Collection of these fragments can be regarded as an incomplete sequence space. Compared to the space that can be sampled by saturation mutation, the sequence space based on preselected glycosyltransferases can reduce the number of samples while ensuring a reasonable exploration space. However, an incomplete sequence space contains many fragments and direct experimental verification is not feasible. Therefore, five filters were designed for different indicators to reduce the size of the sequence space. Although the filters were independent of each other and did not interfere with each other when performing the filtering functions, the time required by different filters to filter the unit sequences was different. Therefore, for more time-consuming filters, it was preferable to run the filters last.

The first filter was based on the length of each loop region. Two situations were needed to be considered. When the loop region contained five or fewer amino acid residues, it was important to keep at least two amino acid residues unchanged. Alternatively, there must be at least three unchanged amino acid residues in the loop regions longer than five amino acid residues.

The second filter was designed to ensure that important amino acid residues remained unaltered during the replacement process. The criteria used to identify important amino acid residues were as follows: 1) Prior evidence from relevant literature suggesting the importance or impact of the mutation on the expression or catalysis of the enzyme (loss of expression or catalytic ability); 2) Those amino acid residues that were in close proximity to the substrate consistently in the docking experiments. The AutoDock docking software was used for docking experiments.

The third filter pertained to the change in the charge carried by the loop region. The isoelectric point of an amino acid residue was used to assess the charge change in the fragment in the same pH environment.

The fourth filter was based on fragment sequence similarity. The commonly used BLOSUM62 (used by BLAST as the default scoring matrix) did not appropriately characterize the variation of fragments in sequence space because of the conserved number of amino acids set in the first filter; therefore, BLOSUM45 was used to calculate extent of similarity in the fragments before and after mutation. S_w^m implied the similarity of the mutated fragment to the wild-type fragment. S_w^w was the loop region with the highest similarity. Therefore, $\frac{S_w^m}{S_w^w}$ less than 50% of the segments were discarded.

The fifth filter, the most time-consuming step, was to calculate the difference in Gibbs' free energy caused by the mutation ($\Delta\Delta G$, the mutant's energy minus the wild-type's energy). PyRosetta [22] was used to determine the energy change required for the filter. In this filter configuration, we took into consideration that all fragment substitutions with $\Delta\Delta G$ more than 0 were unstable and might have a negative impact on the structure of the enzyme. We only retained the fragment where the $\Delta\Delta G$ was negative. Finally, filtering of the sequence space was completed. However, some loop regions had an empty sequence space after the previous filtering.

The construction of mutants with single-fragment substitutions was not the ultimate goal. A mutant with three replaced fragments was constructed using the following steps and was ultimately employed to measure the relative conversion rate. During this step, mutant filtering resembled that of the fifth filter mentioned earlier, with some differences. In the fifth filter as a compromise between accuracy and speed, the relaxed conformation of the mutant was not considered for calculating the

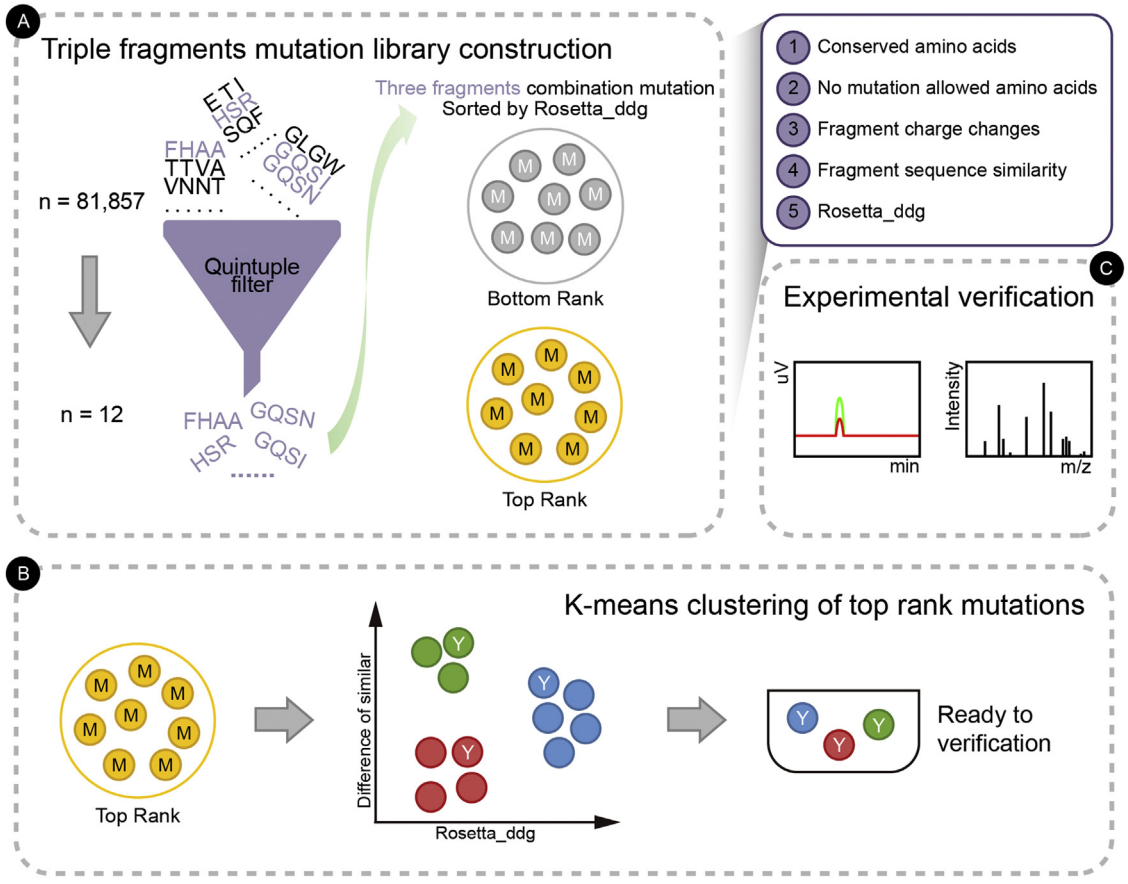

Fig. 1. The main logic of the mutant construction and verification. (A) A mutant library constructed by fragment replacement method. Mutants with lower $\Delta\Delta G$ values were selected for experimental verification. (B) Several different types of mutants selected through K-means clustering, and one of each type is selected for final experimental verification. (C) High performance liquid chromatography and mass spectrometry to detect the reactions catalyzed by the mutants, and to determine their catalytic efficiency.

$\Delta\Delta G$; therefore it was a rough calculation. However, in the current step, as the number of mutants was greatly reduced, the time taken to ensure the accuracy while retaining the speed was acceptable. Therefore, an overall relaxation of the mutant structure was performed before calculating the $\Delta\Delta G$ using PyRosetta.

The verification of every mutant generated by computer design is challenging. Therefore, to select representative mutants for verification, we used a k-means clustering method to divide the existing mutants into several categories according to their similarity difference and $\Delta\Delta G$. We first chose the mutant for each category with the maximum similarity difference. Next, we chose mutants with higher $\Delta\Delta G$ in case the similarity difference was identical. The representative mutants of each category, selected according to the above strategy, were used for verification.

2.3. Construction of mutation vectors

Mutagenic primers were designed according to the instructions of the vazyme multi-point mutation Kit (Order NO. C215), and the online primer design program CE Design (https://crm.vazyme.com/cetool/en-us/simple.html). The formula of Luria-Bertani (LB) liquid medium is 1 L water, 5 g yeast extract purchased from Oxoid Ltd, 10 g tryptone purchased from Oxoid Ltd, 10g NaCl purchased from Sigma-Aldrich LLC. 1.5 g of agar powder purchased from Solarbio is required to make LB solid medium. The DH5α-pET45b-ApNGT strain was cultivated in LB liquid medium containing 0.1 mg/mL ampicillin in an air bath shaker at 170 rpm, 37°C for 12 h, and the plasmid was isolated and used as template. Primers (provided in the Supplementary Materials) were used to amplify fragments carrying mutations and agarose gel electrophoresis was performed to confirm the correct size of the amplified fragments. Dpn I was used to digest methylated template plasmid. Subsequently, the recombination of the mutated fragments was carried out with Exnase as per the manufacturer's protocol, and the recombinant plasmid was used to transform DH5α competent cells. The transformed cells were spread on a solid LB plate containing 0.1 mg/mL ampicillin and incubate at 37°C for 16 h. After selecting the bacterial clones on the plate and sequencing them to confirm the correct mutation, the bacteria were amplified and cultured, and the plasmid was extracted and transferred to BL21 (DE3) cells for expression.

2.4. Purification of NGTs

BL21 (DE3) cells carrying wild-type *ApNGT* or its mutants were inoculated into 1 L LB liquid medium containing 0.1 mg/mL ampicillin. The bacteria were grown at 37°C, 200 rpm until the OD_{600} reached 0.6, and the incubation continued at 16°C for 20 h. The cells were collected by centrifugation and resuspended in 50 mL PBS. The cells were ultrasonically disrupted in ice water to release the protein, followed by centrifugation at 17,420 g for 20 min to separate the precipitate and remove the supernatant to purify wild-type *ApNGT* and its mutant using a Ni column. The Ni column was equilibrated with five volumes of PBS solution (pH 7.2) in advance and eluted successively with 25, 50, 250, and 500 mM imidazole and 250 mM imidazole eluate was collected. The Ni columns were then preserved in 20% ethanol. Ultrafiltration was performed at 3,520 g for 30 min at 4°C with 10 kDa ultrafiltration tubes for 6 times to remove imidazole. The final ultrafiltration step retained 1 mL

of liquid. The protein concentration was determined using a Modified BCA Protein Assay Kit (Order No. C503051; Sangon Biotech [Shanghai] Co. Ltd.).

2.5. NGTs activity identification

The 20 μL reaction system (0.5 μg/μL NGT, 10 mM PBS pH 7.2, 10 mM sugar nucleotide, 0.5 mM substrate peptide) was constructed and incubated at 37°C for 1 h or 12 h. The reaction solution was incubated at 95°C for 5 min after reaction, and then centrifuged at 9,657 g for 10 min. 0.4 μL of the supernatant was used for thin layer chromatography (TLC) detection. The TLC separation solvent used was butanol: acetic acid: water = 2:1:1. After 40–60 min of chromatography, the thin plate was removed, dried with a hair dryer in a fume hood, and placed under an excitation light for observation. The supernatant was desalinated using a ZipTip C18 microchromatography column and diluted 100 times with a 0.1% formic acid solution for mass spectrometry.

2.6. Determination of optimum temperature and pH

Using UDP-Glc as donor substrate in PBS buffer (pH 7.2), the glycosyltransferase activity was measured at different temperatures from 16°C to 60°C for 5 min. The reaction mixtures contained 1 mg/mL wild-type enzyme (or 0.05 mg/mL of the F13 mutant), 10 mM PBS buffer (pH 7.2), 0.1 mM TRAMA-DANYTK, and 0.2 mM UDP-Glc.

Using UDP-Glc as donor substrate at 37°C, the glycosyltransferase activity was measured in buffer systems with pH ranging from 5.0 to 10.0 for 6 min and 30 s. The reaction mixtures contained 0.5 mg/mL of the wild-type enzyme (or 0.025 mg/mL of the F13 mutant), 25 mM buffer, 0.05 mM TRAMA-DANYTK, and 0.1 mM UDP-Glc. MES buffer was used for pH 5.0, 6.0, and 7.0 in the reaction mixtures. Whereas, Tris buffer was used for reaction mixtures of pH 8.0, 9.0, and 10.0.

The analysis was performed as described above, and all reactions were performed in triplicate. The products were identified and quantified using reversed-phase high-performance liquid chromatography (RP-HPLC).

2.7. Determination of kinetic constants

The production of glycosylated peptides was quantified by monitoring the relative conversion of the reaction system by measuring the absorbance of a UV detector (220 nm) of RP-HPLC. The relative conversion of enzymes was measured in 10 μL reaction mixtures containing 1 mg/mL enzyme, 10 mM PBS buffer (pH 7.2), 0.1 mM TRAMA-DANYTK, and different concentrations of UDP-Glc (0.1–5 mM) at 37°C. The relative conversion rate of the wild-type enzyme was obtained by reaction for 15 min, incubation at 95°C for 5 min to inactivate the enzyme, and centrifugation at 9,657 g for 10 min. The supernatant was diluted 20 times, and the relative conversion rate was determined using RP-HPLC. The relative conversion rate of F13 mutant was also measured using a similar procedure, except that F13 mutant reaction time was 5 min and the reaction mixtures contained 0.05 mg/mL enzyme. All enzymes were assessed in three independent replicates. Values and standard errors for the apparent affinity constant (K_m) and enzyme turnover (K_{cat}) were obtained by nonlinear least-squares fitting of the experimental measurements to the Michaelis–Menten model using the GraphPad Prism 8.0.2 software.

3. Results

3.1. Design result of mutants

Seventy-eight glycosyltransferases were randomly selected from the PDB database (Table S1). Glycosyltransferases were selected to construct the fragment replacement library because there may be some shared fragments among the glycosyltransferases, and the replacement of these fragments may result in improved enzyme properties. However, among the 78 glycosyltransferases, a special protein was annotated as a glycosyltransferase in the PDB database, but was annotated as a glycoside hydrolase in the functional annotation of Uniprot. This might be due to the fact that the two seemingly different activities of glycosyl transfer and glycoside hydrolysis could be interconverted in some cases [6].

The initial sequence space was constructed using the previously downloaded 78 glycosyltransferases. To set the filters required for designing the mutants, the following selections were made. The second filter was set for the amino acids that were not allowed to change. However, the second filter may not be necessary in all loop regions. According to the literature [16,23], mutations in Tyr-222, Ala-276, Lys-441, Tyr-501, Asn-521, Asp-525, and the presumed catalytic residue His-277 cause the enzyme to become inactive. In addition, important amino acids (Table S2) that very frequently interacted between the substrate peptide and the enzyme were investigated using AutoDock. Finally, in combination with selected loop regions, Phe-273 was identified as an amino acid that could not be altered. A third filter was used to ensure a benign change in charge near the mutation loop. Binding of the substrate sugar nucleotide may be related to the positive charge of the donor-acceptor binding pocket of ApNGT. Therefore, we removed fragments with a negative overall charge from the sequence space.

The remaining sequence spaces for each loop region after five rounds of filtering are listed in Table S3. Of the seven fragments selected, only five fragments still had mutable fragments after filtering. This may be related to the strict filter conditions that we set.

Fourteen mutants were constructed, and their corresponding mutation sites are listed in Table S4. Despite our mutational strategy for fragment replacement, the majority of replacements resulted in single-point mutations. This result might be due to the fact that we selected short fragments and set strict filtering conditions.

Table 1
Kinetic parameters of wild-type ApNGT and F13 mutant with UDP-Glc as substrate.

Enzyme	K_m (mM)	K_{cat} (min^{-1})	K_{cat}/K_m ($mM^{-1}min^{-1}$)
WT	12.63×10^{-2}	5.008×10^{-2}	0.3965
F13	9.919×10^{-2}	1111×10^{-2}	112.0

Note: WT, wild-type

3.2. Determination of enzymatic properties

Except for the mutants that could not be expressed or purified, all other mutants were identified by TLC. From the TLC results, the relative level between the enzyme activity of the mutant and wild-type enzymes could be roughly determined, and mutants with significantly better enzyme activity were selected for quantitative verification. Through preliminary experiments, we found that the activity of glucose glycosyltransferase was significantly increased in F13 mutant compared to that of the wild-type; therefore, the F13 mutant was the focus of the study for later experiments.

The effect of temperature on the activity of the F13 mutant was observed from 16°C to 60°C (Fig. 2A). In this study, the optimum temperature of F13 mutant was about 40°C. F13 mutant had a similar preferred temperature compared to wild-type ApNGT (37°C) [24,25]. The effect of pH on the activity of F13 mutant was analyzed using different buffer conditions (Fig. 2B). The optimum pH for F13 mutant was 8.0 in the Tris buffer, similar to that of wild-type ApNGT.

We used UDP-Glc as a donor substrate to quantitatively analyze the kinetic properties of wild-type and F13 mutant. RP-HPLC was used to measure reaction velocities at a fixed sugar substrate concentration (Figure S2). The apparent K_m and K_{cat} values are listed in Table 1. The K_m of wild-type ApNGT was 1.27 times that of F13 mutant, and there was no significant difference. In contrast, the K_{cat} of the two differed by

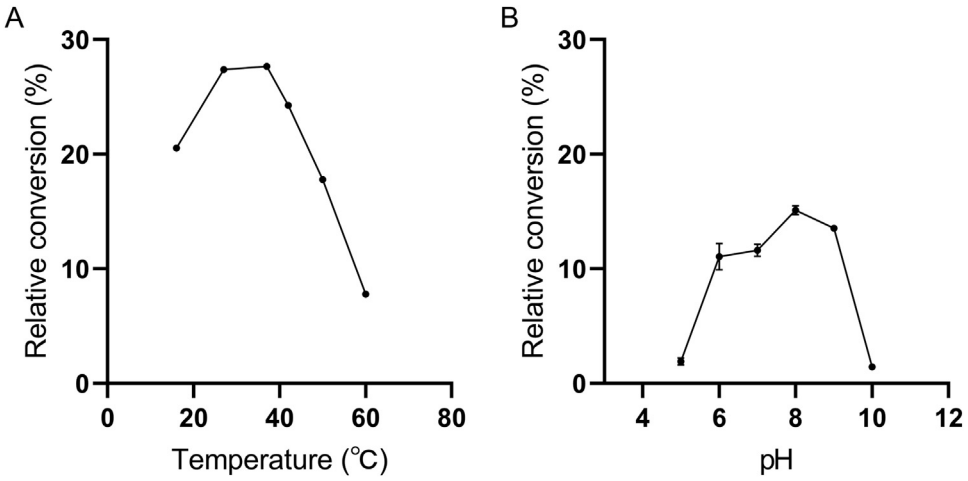

Fig. 2. Biochemical properties of the F13 mutant. The relative conversion is calculated by the peak area of the nude peptide and glycosylated peptide. (A) The reaction at different temperatures (16°C, 27°C, 37°C, 42°C, 50°C, and 60°C). (B) The reaction in buffers with varying pH values (5.0, 6.0, 7.0, 8.0, 9.0, and 10.0).

Table 2
Relative conversions of wild-type ApNGT and F13 mutant (donor: UDP-Glc).

Source	Substrate peptide	Relative conversion (%)	
		ApNGT	F13 mutant
HMW1	TAMRA-DANYTK	69.4	100
Dipeptidyl peptidase 1	CDTPANCTYLDLL	0	20.9
Hemopexin	TAMRA-PAVGNCSSALR	100	100
Hemopexin	TLDDNGTMLFFK	0	17.7

Note: After incubation at 37 °C for 12 h, the supernatant was centrifuged and analyzed by mass spectrometry.

Table 3
Relative conversion of wild-type ApNGT and F13 mutant using various sugar donors.

Sugar donor	Relative conversion (%)	
	ApNGT	F13 mutant
UDP-Glc	69.4	100
UDP-Gal	6.25	88.9
UDP-Xyl	< 1	78.6
UDP-GlcA	0	13.8
UDP- GlcNAc	0	0

222 times. This indicates that mutations in F13 greatly changed the rate of substrate dissociation, but the affinity of substrates for the enzymes changed minimally.

3.3. Recognition of peptide acceptors by the F13 mutant

The relative conversion rates of mutant and wild-type ApNGT were investigated using UDP-Glc as the donor. The reaction system was incubated at 37 °C for 1 h. We tested the relative conversion of the mutants and ApNGT to the natural peptide substrate DANYTK. F13 mutant exhibited the highest conversion activity. The relative conversion of DANYTK by the F13 mutant was higher than that of ApNGT (95.95% ± 4.05% vs. 8.75% ± 1.25%, respectively) (Fig. 3B).

We also determined that the F13 mutant was still an *N*-glycosyltransferase using the higher-energy collisional dissociation (HCD) fragmentation method (Fig. 3A). The b3, b4, and b5 fragments increased the molecular weight of glucose, indicating that the F13 mutant transferred glucose to asparagine.

The reaction system was incubated at 37 °C for 12 h. Four peptides were tested as the substrates. For the Hemopexin fragment PAVGNCSSALR, wild-type ApNGT and F13 mutants showed a relative conversion of 100% (Figure S3 and Table 2). For the other three peptides, the F13 mutant showed significantly improved relative conversion (Table 2). The relative conversion of the natural substrate, DANYTK increased from 69.4% of wild-type ApNGT to 100% (Table 2 and Figure S4). For the dipeptidyl peptidase 1 fragment CDTPANCTYLDLL, F13 showed a relative conversion of 20.9%, and wild-type ApNGT did not show any activity (Table 2 and Figure S5). For the hemopexin fragment TLDDNGTMLFFK, the F13 mutant had a relative conversion of 17.7%. The wild-type ApNGT showed no activity (Table 2 and Figure S6).

3.4. Recognition of sugar donors by F13 mutant

The reaction system was incubated at 37 °C for 12 h. Wild-type ApNGT showed very low utilization of UDP-Gal and UDP-Xyl (Table 3, Figure S7 and S8); whereas, the F13 mutant showed high utilization. Neither the F13 mutant nor wild-type ApNGT used UDP-GlcNAc as a sugar donor. However, the F13 mutant used UDP-GlcA (Fig. 4) as a sugar donor and showed a relative conversion of 13.8% of the DANYTK peptide substrate.

3.5. Enlargement of the acceptor binding channel

The structure of the F13 mutant was predicted using AlphaFold2. Song et al. proposed the hypothesis of an acceptor substrate-binding channel [18]. A comparison of the structural differences between the F13 mutant and wild-type revealed a significant change in one of the four key amino acids in the binding channel (Fig. 5). This may lead to an increased tolerance of the binding channel for substrate peptides.

Combining the data in Table 1, we can see that the substrate dissociation rate of the mutant F13 was greatly increased, which was related to a change in the amino acids in the binding channel. This binding channel was also responsible for the dissociation of glycosylated peptides after the reaction was completed. The dissociation rate changes when the channel is enlarged. Therefore, it could be speculated that the binding affinity of the mutant F13 to the receptor peptide would also change significantly.

4. Discussion

Here, we present a method for constructing mutants by fragment replacement. The traditional rational design scheme is based on key amino acids involved in catalysis or interactions with substrates. Our method focuses on replacing the loops containing key amino acids, thereby providing a better environment for key amino acids to perform their functions. This has allowed researchers to focus on specific amino acid residues in certain regions. However, our method had some limitations. We could not achieve a fully automated mutant design, and manual intervention was needed to set the filtering conditions during

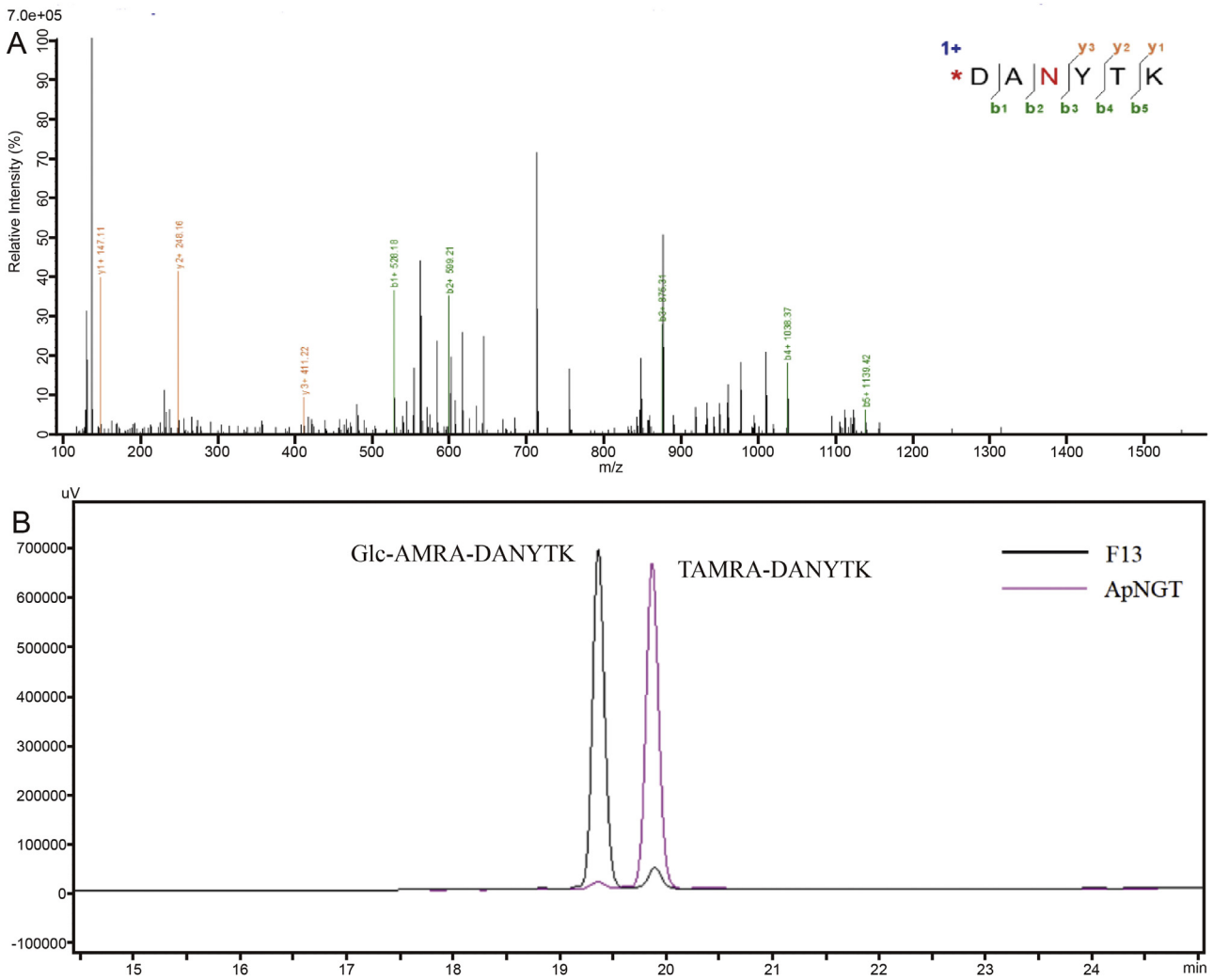

Fig. 3. Mass spectrometric analysis performed by the HCD method (A), and RP-HPLC analysis (B). (A) b3, b4, and b5 fragments increase the molecular weight of glucose, which prove that F13 mutant transfers glucose to asparagine. (B) Because the increased Glc makes the product more polar, the peak of the product appears earlier in the RP-HPLC. After glycosylation, the polarity of the peptide increases and the retention time in RP-HPLC advances.

candidate fragment selection. Particularly, inputs from personnel with a basic understanding of the enzyme donor-acceptor binding pocket (e.g. for determining the charged nature of the catalytic substrate and location of the donor-acceptor binding pocket, and for identifying reports pertaining to effect of the mutation on inactivation of the enzyme) were necessary. The quality of the final mutant design is affected by a set of conditions.

Compared to protein design based on artificial intelligence [26,27], rational design usually has a higher success rate. Rational design is undoubtedly a better choice for glycosyltransferases, which are difficult to determine with high throughput. In this study, we screened 14 mutants and found one with significantly improved catalytic efficiency. However, in *de novo* protein design under the guidance of artificial intelligence, hundreds of proteins must be screened to obtain the target protein. Another type of protein design scheme used was Rosetta [12,13], which uses the same technology as PyRosetta in this study. The difference between our method and these methods is that the sequence was treated differently before using PyRosetta to calculate the energy. These methods use a position-specific scoring matrix to screen sequences; whereas, we combined several carefully designed conditions based on fragment replacement to screen sequences. Our method required only a database of 78 glycosyltransferases. Thus, the time spent on the calculations was greatly reduced.

We applied this method to design ApNGT and successfully obtained a mutant enzyme, F13, which exhibited significant advantages. The F13 mutant not only broadened the original substrate specificity but also exhibited catalytic activity for previously unavailable GlcA. Owing to its negative charge, GlcA is an important sugar molecule that differs from other sugar molecules. Previous studies have shown that GlcA can be used to modify liposomes, such that the GlcA-modified liposomes are highly selective for glucose transporter-1 in pulmonary arterial smooth muscle cells, thereby enabling targeted delivery of drugs [28,29]. This implied that GlcA-modified proteins may have similar targeting effects, providing an important tool for protein-targeted drug delivery. Additionally, as an enzyme derived from bacteria, the F13 mutant exhibits better expression in non-eukaryotes, making it advantageous for heterologous expression in industrial production.

AlphaFold2 is state-of-the-art software developed for protein structure prediction that has demonstrated high prediction accuracy for a broad range of proteins. We used AlphaFold2 to predict the structure of the F13 mutant. Alignment of the structure of the F13 mutant with that of the wild-type, revealed that the F13 mutant had a clear change in the peptide substrate-binding channel.

Although F13 reacted for 12 h, its relative conversion rate was not sufficiently high. However, the acquisition of this new trait is particularly important, and F13 is expected to improve the utilization of GlcA

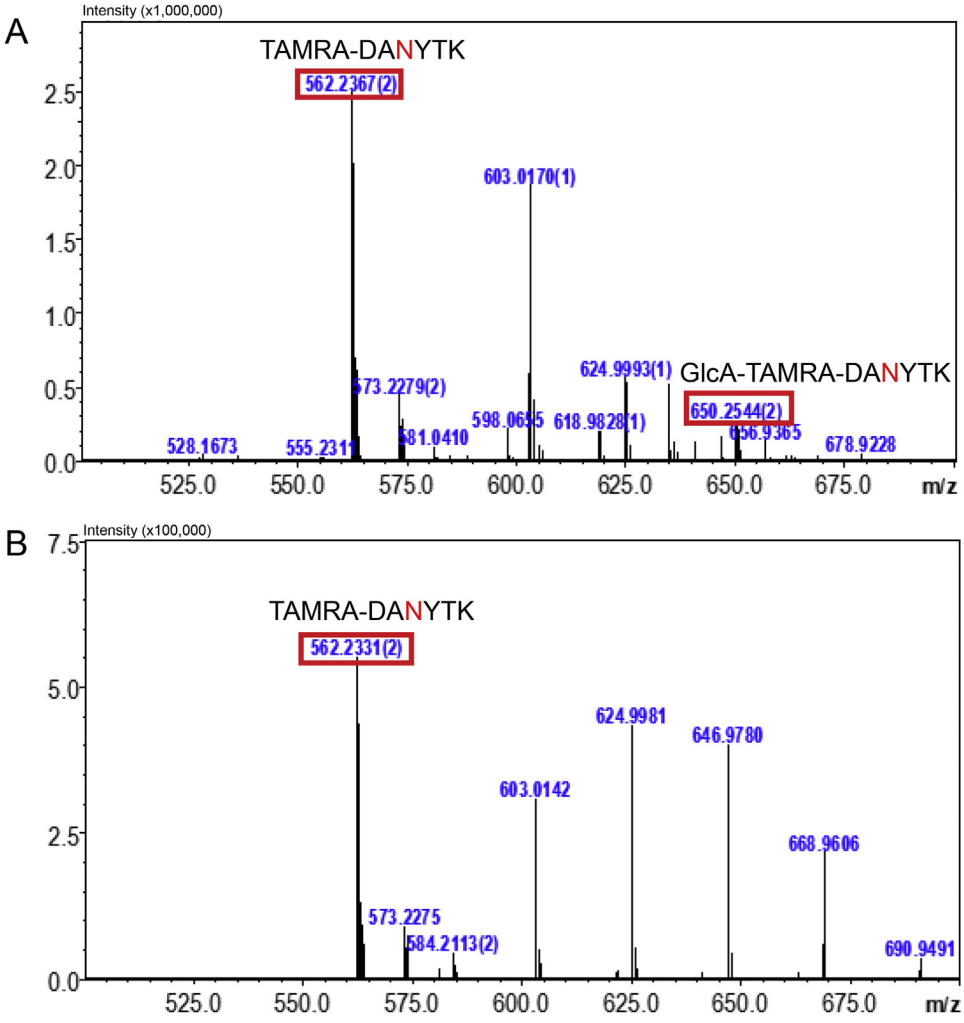

Fig. 4. Mass spectrometry results of the F13 mutant and wild-type ApNGT glycosylated DANYTK fragment using UDP-GlcA as sugar donor. After incubation at 37 °C for 12 h, the supernatant was centrifuged and analyzed by mass spectrometry. (A) F13 mutant and (B) wild-type ApNGT.

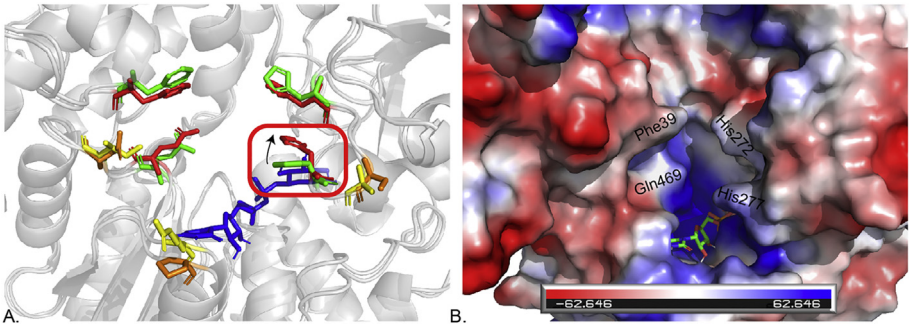

Fig. 5. Changes in acceptor substrate binding channels. (A) Channel changed before and after mutation. Green and red represented the four amino acid residues in the receptor substrate binding channel of wild-type and F13 mutant. Orange and yellow represented the amino acid residues at the mutation site of wild-type and F13 mutant. Blue represents UDP-Glc. (B) The four amino acid residues in the acceptor substrate binding channel might be related to the binding ability of the peptide [18].

through subsequent mutations, thereby achieving a higher relative conversion rate.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Data Availability Statement

All data generated or analyzed during this study are included in this published article (and its Supplementary Materials files).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Jiangyu Yang: Writing – review & editing, Writing – original draft, Software, Conceptualization. **Kun Li:** Writing – review & editing, Writing – original draft, Conceptualization. **Yongheng Rong:** Methodology, Investigation. **Zhaoxi Liu:** Visualization, Resources, Formal anal-

ysis. **Xiaoyu Liu:** Formal analysis, Data curation. **Yue Yu:** Validation, Methodology. **Wenjing Shi:** Visualization, Formal analysis. **Yun Kong:** Software, Resources. **Min Chen:** Writing – review & editing, Writing – original draft, Project administration, Funding acquisition, Conceptualization.

Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFA1304103), the National Natural Science Foundation of China (Grant numbers 32070921), the Key Research and Develop Program of Shandong Province (grant number 2020CXGC010601) and Taishan industry leading talents project (No. tscy20200221).

We would like to thank Jing Zhu, Jingyao Qu, Guannan Lin and Zhifeng Li from State Key laboratory of Microbial Technology of Shandong University for help and guidance in LC-MS. The AlphaFold2 prediction in this paper have been done on the HPC Cloud Platform of Shandong University.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.engmic.2023.100134.

References

[1] E.P. Lillehoj, K. Kato, W. Lu, et al., Cellular and molecular biology of airway mucins, Int Rev Cell Mol Biol 303 (2013) 139–202, doi:10.1016/B978-0-12-407697-6.00004-0.

[2] K.M. Loomes, H.E. Senior, P.M. West, et al., Functional protective role for mucin glycosylated repetitive domains, Eur J Biochem 266 (1999) 105–111, doi:10.1046/j.1432-1327.1999.00824.x.

[3] A.W. Chung, M. Crispin, L. Pritchard, et al., Identification of antibody glycosylation structures that predict monoclonal antibody Fc-effector function, AIDS 28 (2014) 2523–2530, doi:10.1097/QAD.0000000000000444.

[4] S. Benini, Carbohydrate-Active Enzymes: Structure, Activity, and Reaction Products, Int J Mol Sci. 21 (2020) 5, doi:10.3390/ijms21082727.

[5] C.K. Bandi, A. Goncalves, S.V. Pingali, et al., Carbohydrate-binding domains facilitate efficient oligosaccharides synthesis by enhancing mutant catalytic domain transglycosylation activity, Biotechnology and Bioengineering 117 (2020) 2944–2956, doi:10.1002/bit.27473.

[6] C.K. Bandi, A. Agrawal, S.P. Chundawat, Carbohydrate-Active enZyme (CAZyme) enabled glycoengineering for a sweeter future, Curr Opin Biotechnol 66 (2020) 283–291, doi:10.1016/j.copbio.2020.09.006.

[7] D. Chen, S. Fan, R. Chen, et al., Probing and Engineering Key Residues for Bis-C-glycosylation and Promiscuity of a C-Glycosyltransferase, ACS Catalysis 8 (2018) 4917–4927, doi:10.1021/acscatal.8b00376.

[8] Y. Li, L. Wei, Z. Zhu, et al., Rational design to change product specificities and thermostability of cyclodextrin glycosyltransferase from Paenibacillus sp, RSC Advances 7 (2017) 13726–13732, doi:10.1039/c7ra00245a.

[9] E. Dellus-Gur, M. Elias, E. Caselli, et al., Negative Epistasis and Evolvability in TEM-1 beta-Lactamase–The Thin Line between an Enzyme's Conformational Freedom and Disorder, J Mol Biol 427 (2015) 2396–2409, doi:10.1016/j.jmb.2015.05.011.

[10] T.N. Starr, J.W. Thornton, Epistasis in protein evolution, Protein Sci 25 (2016) 1204–1218, doi:10.1002/pro.2897.

[11] G.J. Cadet XF, A van Noord, F Cadet, CG Acevedo-Rocha, Learning Strategies in Protein Directed Evolution, Methods in molecular biology (Clifton, N.J.) 2461 (2022) 225–275, doi:10.1007/978-1-0716-2152-3_15.

[12] A. Goldenzweig, M. Goldsmith, S.E. Hill, et al., Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability, Mol Cell 63 (2016) 337–346, doi:10.1016/j.molcel.2016.06.012.

[13] O. Khersonsky, R. Lipsh, Z. Avizemer, et al., Automated Design of Efficient and Functionally Diverse Enzyme Repertoires, Mol Cell 72 (2018) 178–186 e175, doi:10.1016/j.molcel.2018.08.033.

[14] V. Saez-Jimenez, S. Scrima, M. Lambrughi, et al., Directed Evolution of (R)-2-Hydroxyglutarate Dehydrogenase Improves 2-Oxoadipate Reduction by 2 Orders of Magnitude, ACS Synth Biol 11 (2022) 2779–2790, doi:10.1021/acssynbio.2c00162.

[15] S. Grass, C.F. Lichti, R.R. Townsend, et al., The Haemophilus influenzae HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin, PLoS Pathog 6 (2010) e1000919, doi:10.1371/journal.ppat.1000919.

[16] F. Kawai, S. Grass, Y. Kim, et al., Structural insights into the glycosyltransferase activity of the Actinobacillus pleuropneumoniae HMW1C-like protein, J Biol Chem 286 (2011) 38546–38557, doi:10.1074/jbc.M111.237602.

[17] A. Naegeli, C. Neupert, Y.Y. Fan, et al., Molecular analysis of an alternative N-glycosylation machinery by functional transfer from Actinobacillus pleuropneumoniae to Escherichia coli, J Biol Chem 289 (2014) 2170–2179, doi:10.1074/jbc.M113.524462.

[18] Q. Song, Z. Wu, Y. Fan, et al., Production of homogeneous glycoprotein with multi-site modifications by an engineered N-glycosyltransferase mutant, J Biol Chem 292 (2017) 8856–8863, doi:10.1074/jbc.M117.777383.

[19] J. Cuccui, V.S. Terra, J.T. Bosse, et al., The N-linking glycosylation system from Actinobacillus pleuropneumoniae is required for adhesion and has potential use in glycoengineering, Open Biol 7 (2017) 13, doi:10.1098/rsob.160212.

[20] I.J. Passmore, A. Andrejeva, B.W. Wren, et al., Cytoplasmic glycoengineering of Apx toxin fragments in the development of Actinobacillus pleuropneumoniae glycoconjugate vaccines, BMC Vet Res 15 (2019) 6, doi:10.1186/s12917-018-1751-2.

[21] Y. Xu, Z. Wu, P. Zhang, et al., A novel enzymatic method for synthesis of glycopeptides carrying natural eukaryotic N-glycans, Chem Commun (Camb) 53 (2017) 9075–9077, doi:10.1039/c7cc04362g.

[22] S. Chaudhury, S. Lyskov, J.J. Gray, PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta, Bioinformatics 26 (2010) 689–691, doi:10.1093/bioinformatics/btq007.

[23] C. Martinez-Fleites, M.S. Macauley, Y. He, et al., Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation, Nat Struct Mol Biol 15 (2008) 764–765, doi:10.1038/nsmb.1443.

[24] Z. Liu, K. Li, X. Liu, et al., Production of microhomogeneous glycopeptide by a mutated NGT according FuncLib with unique sugar as substrate, Enzyme and Microbial Technology 154 (2022), doi:10.1016/j.enzmictec.2021.109949.

[25] A. Naegeli, G. Michaud, M. Schubert, et al., Substrate Specificity of Cytoplasmic N-Glycosyltransferase, Journal of Biological Chemistry 289 (2014) 24521–24532, doi:10.1074/jbc.M114.579326.

[26] J.L. Watson, D. Juergens, N.R. Bennett, et al., Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models, bioRxiv (2022) 2022.2012.2009.519842, doi:10.1101/2022.12.09.519842.

[27] S.V. Torres, P.J.Y. Leung, I.D. Lutz, et al., De novo design of high-affinity protein binders to bioactive helical peptides, bioRxiv (2022) 2022.2012.2010.519862, doi:10.1101/2022.12.10.519862.

[28] A. Liu, B. Li, M. Yang, et al., Targeted treprostinil delivery inhibits pulmonary arterial remodeling, Eur J Pharmacol 923 (2022) 174700, doi:10.1016/j.ejphar.2021.174700.

[29] B. Li, W. He, L. Ye, et al., Targeted Delivery of Sildenafil for Inhibiting Pulmonary Vascular Remodeling, Hypertension 73 (2019) 703–711, doi:10.1161/hypertensionaha.118.11932.