DEVELOPMENT OF A CLICK CHEMISTRY BASED SCREENING METHODOLOGY FOR THE DIRECTED EVOLUTION OF GLYCOSYNTHASES

By

AYUSHI AGRAWAL

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Chemical and Biochemical Engineering

Written under the direction of

Shishir P. S. Chundawat

And approved by

_____

_____

_____

New Brunswick, New Jersey

October, 2019

ABSTRACT OF THE THESIS

Development of a click chemistry based screening methodology for the directed evolution of glycosynthases

By AYUSHI AGRAWAL

Thesis Director: Shishir P.S. Chundawat

Glycans are the most abundant biomolecules in nature and play a significant role in biological systems. They are involved in diverse biological processes such as intercellular signaling, intracellular metabolism, and protein folding. Hence, while the importance of glycans in the fields of chemistry, biology, and bioengineering is being increasingly recognized, limited progress has been made due to the lack of well-defined glycan based reagents available for research applications. Development of efficient chemoenzymatic routes for synthesis of the designer carbohydrates (or glycans) has therefore gained momentum in recent years. Glycosynthases are mutant glycosidase enzymes that can catalyze the formation of glycosidic bonds (between monosaccharide groups) instead of breaking/cleaving these bonds. But identifying a suitable glycosidase from a library of mutants that can act as an efficient glycosynthase for targeted donor or acceptor sugars is still very challenging. This study focuses on the development of a high-throughput

screening (HTS) method to rapidly screen and evolve glycosynthase variants generated using traditional directed evolution techniques to enhance catalytic activity or product specificity. A model glycosynthase was used here for validation of this HTS method. Random mutagenesis (via error-prone PCR) was performed on this model glycosynthase to generate a library of mutants that were then screened using this validated HTS method.

Acknowledgements

I would like to express my sincere gratitude to the following individuals who have helped me accomplish my dissertation for the degree. At the forefront, I would like to thank my thesis advisor Dr. Shishir Chundawat whose expertise was invaluable in the designing of the project and methodology in particular. I would also like to thank Dr. Zhang and Dr. Gormley for serving on my committee.

I would like to acknowledge Dr. Falkowski for providing access to his flow cytometer instrument. I would also like to acknowledge my lab colleagues especially Chandra Kanth Bandi for help with experiments, guidance, and immense support during the course of my project. In addition, I would like to thank Rutgers University (SOE) for supporting this research project.

My appreciation also goes out to my parents and brother for believing in me and supporting me throughout the last three years of my course.

Table of Contents

Table of Figures:

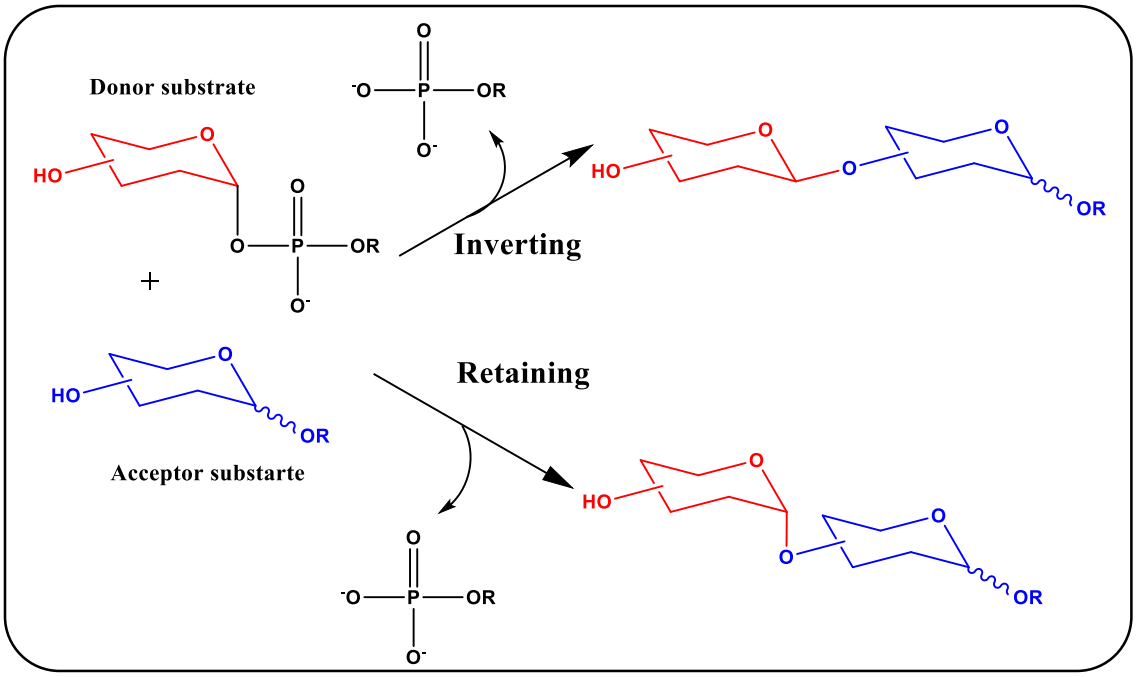
Chapter 1: Introduction

1.1 Background

Glycans consist of a broad class of biomolecules like monosaccharides, oligosaccharides, polysaccharides, and carbohydrate-based glycoconjugates that are attached to other biologically relevant biomolecules like proteins or lipids [1]. Glycans are the building blocks of life that are central to many biological processes such as cell-cell and cell-matrix interactions, protein folding, receptor binding, protein clearance, protein trafficking, serving as point of attachments for viruses and bacteria and other cells, clinical diagnostics such as cancer cells detection, and in recent years as novel ligand targets for drug therapeutics [2], [3], [4]. Since glycans play a crucial role in the life cycle of most living systems, developing catalytic and/or biocatalytic processes that enable their facile synthesis has become critical in recent years.

Over the years, several approaches have been used for glycans synthesis such as chemically catalyzed synthesis and enzymatically/biologically-catalyzed synthesis. Chemical synthesis using either the multistep one-pot solution-phase synthesis approach or the solid phase synthesis approach are traditional paths followed by researchers but these chemical based synthesis routes can be very time consuming, laborious, require multi-step synthetic chemistry methods for functional groups protection/deprotection, lack regioselectivity and stereospecificity, and are not economical [5],[6],[7]. To overcome these drawbacks, enzymatic based methods are used that provide exquisite control over the reaction stereospecificity and regioselectivity, while been conducted in aqueous solutions at close to ambient

temperatures and provide higher reaction efficiency [8],[9], [10]. The enzyme families mostly responsible for biosynthesis of glycans in cellular systems are called Glycosyl Transferases (GTs). GTs are enzymes that catalyze the transfer of a sugar moiety from an activated sugar donor (like UDP-Glucose) onto saccharide or aglycone acceptor group [11]. GTs constitutes a large family of enzymes that are involved in the

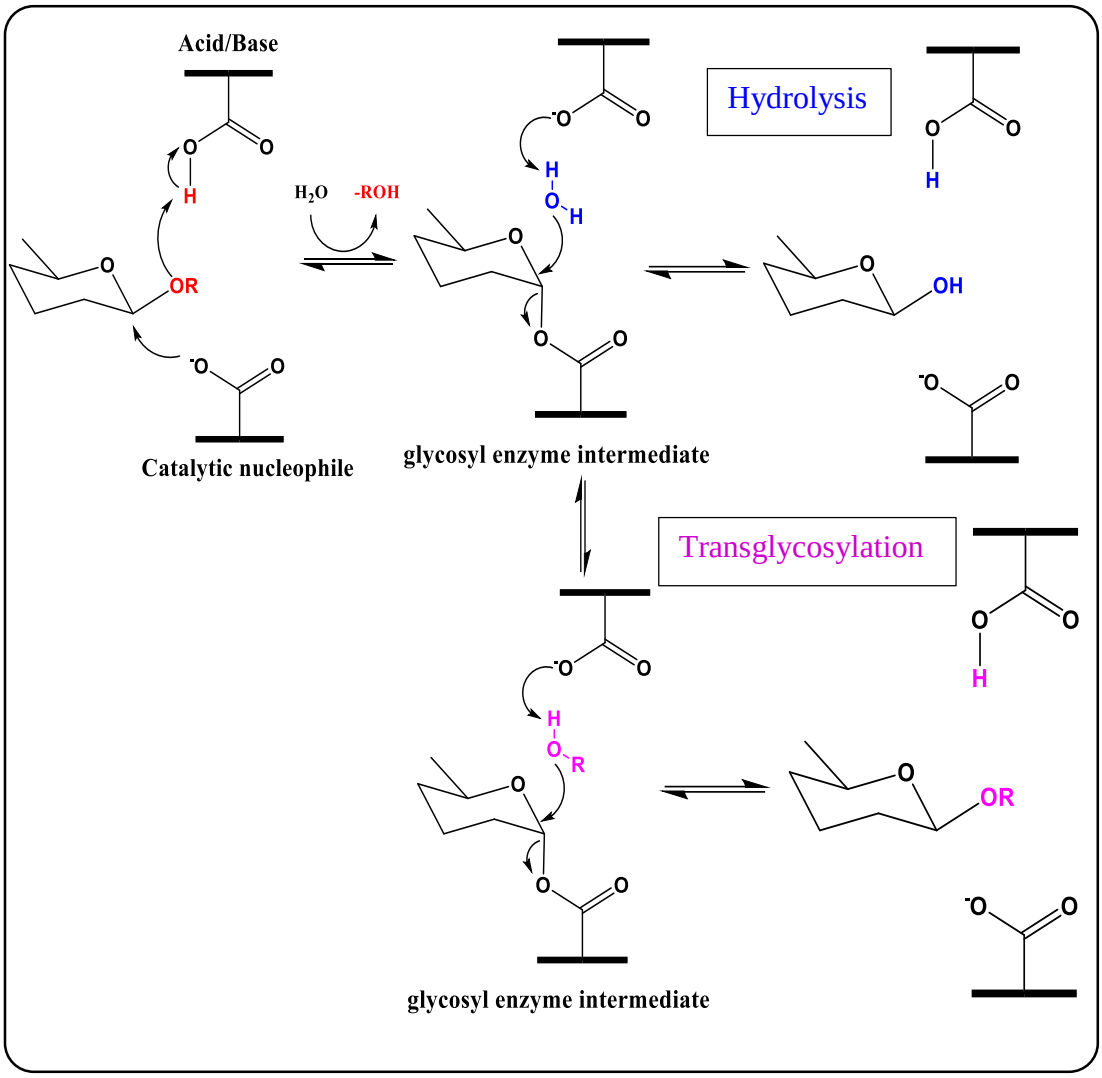


Scheme 1: Glycosyltransferases catalyze glycosyl group transfer with either inversion or retention of the anomeric stereochemistry with respect to the donor sugar [12].

biosynthesis of oligosaccharides, polysaccharides, and glycoconjugates. GTs are prevalent in both prokaryotes and eukaryotes [11]. They are subdivided into two major enzyme classes termed retaining and inverting. Retaining GTs transfer the donor with retention of stereochemistry at the anomeric position, whereas inverting GTs transfer the donor with inversion at the anomeric position as shown in Scheme 1 [12]. However, the scale-up of synthesis of oligosaccharides by GTs is still restricted due to a lack of readily available glycosyltransferases, poor expression capability in *E. coli*, and the

specific but limited substrate specificities of many of these enzymes [13], [14]. An alternative to using glycosyltransferases for glycans synthesis are glycosyl hydrolases (GHs) that typically catalyze the hydrolysis of glycosidic bonds. In nature, this class of enzymes have been shown to be involved in diverse metabolic processes at various spatiotemporal scales such as the enzymatic processing of N-linked or O-linked glycans during protein folding in the ER/golgi; breakdown of complex human milk oligosaccharides (HMOs) in the intestinal tract by gut microbiota into simple carbohydrates like lactose and fucose; biosynthesis and degradation of glycogen in the body to provide energy; and deconstruction of plant biomass into fermentable sugars by cellulolytic microbes to recycle most terrestrial carbon [15], [16], [17]. They are readily used because of their availability, robustness, and low cost and broad range of donor and acceptor group specificities. Along with hydrolysis, some of these enzymes can also synthesize oligosaccharides via the transglycosylation (TG) pathway if an acceptor sugar attacks the glycosyl enzyme intermediate instead of a water molecule as shown in Scheme 2. In general, GHs also follow either an inverting or retaining mechanism. In the retention mechanism, each step goes through an oxocarbenium like transition state. In the first step (glycosylation), the catalytic nucleophile attacks the anomeric center to displace the aglycone and form a glycosyl-enzyme intermediate. The acid/base residue simultaneously acts as an acid and protonates the glycosidic oxygen as the bond cleaves. In the second step (deglycosylation), the reaction can either proceed forward by hydrolysis (based on thermodynamic control) of the glycosyl enzyme intermediate, which is generally the case for GH enzymes, or by

transglycosylation (based on kinetic control) mechanism in the presence of another sugar moiety.



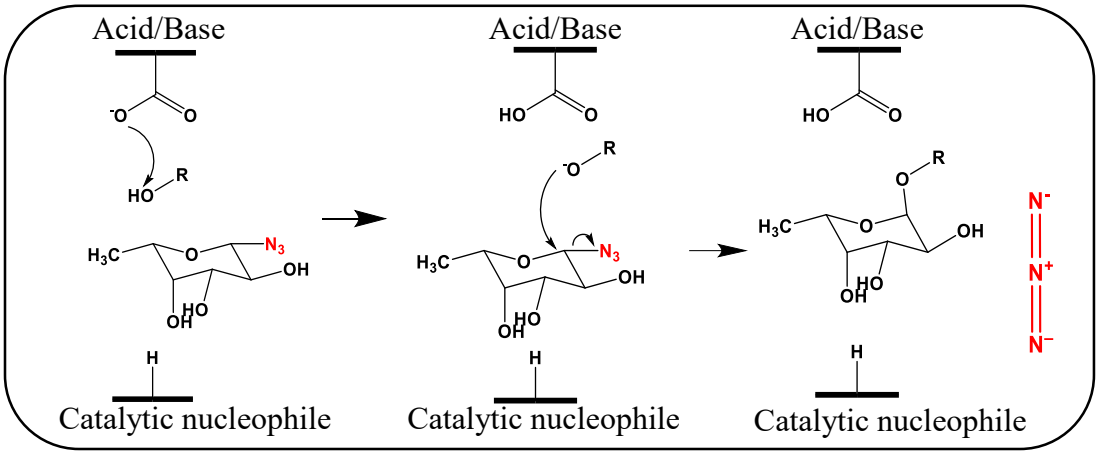
Scheme 2: Mechanism of hydrolysis vs. transglycosylation reactions for GH

The most basic requirement for a transglycosylation reaction to occur is that the acceptor should compete successfully with water in the deglycosylation of the glycosyl enzyme intermediate. The enzyme structural properties and the overall substrate concentration are two critical deciding factors for whether hydrolysis or transglycosylation would ultimately take place. However, the amount of water also

plays a significant role in deciding the glycosidase mechanism since hydrolysis is generally a favored mechanism for GHs, thus aprotic solvents can also facilitate TG reactions [18]. A native GH will only act as a TG under very specific reaction conditions. The major drawback of transglycosylation mechanism is that the products of the reaction are subjected to secondary hydrolysis due to the presence of water in the reaction mixture and most GH give very poor TG product yields (<1-5%). To avoid TG product hydrolysis, hydrolytically inactive GHs could be used that retain their synthetic ability but are engineered to have very low hydrolytic activity.

Glycosynthases (GSs) are a special class of GH enzymes that have been engineered to catalyze the formation of glycosidic linkage. GSs are formed by mutating the amino acid residue at the catalytic nucleophile of a retaining (and some inverting) glycosidase to a small non-nucleophilic amino acid like Alanine, Serine, or Glycine [19]. Glycosynthase reactions take place for α-glycosidases with β-donor sugar with excess β-acceptor sugars or for β-glycosidase with α-donor sugar in the presence of excess α-acceptor sugar. GS reactions do not undergo secondary hydrolysis when compared to TG reactions because their catalytic nucleophile is mutated to be inactive, which is responsible for the hydrolysis step. The mechanism of glycosynthase reaction can be observed in Scheme 3. Engineering GHs into an efficient glycosynthase may also require mutating other neighboring residues in the active site. These additional mutations can be identified either guided rationally using available protein structural information for *in silico* modeling or by random mutagenesis followed by directed evolution and functional screening. Rational engineering can be computationally intensive and requires prior knowledge of the enzyme structure and its reaction

mechanism, which is not trivial for most known enzyme families. Therefore, random mutagenesis of the wild-type enzyme followed by high-throughput (HT) screening is a suitable alternative for selection of enzymes with desired function from a large library of mutants. However, the key limitation of the random mutagenesis approach is the limited availability of HT-screening methods for glycosynthases.



Scheme 3: General mechanism for glycosynthases

A limited number of GS screening techniques have been developed and used by researchers for screening glycosynthase activity in the last two decades. Withers and co-workers from the University of British Columbia (Canada) had first developed a Luria-Bertani (LB) agar plate based colorimetric mutant *E. coli* colonies screening method using a coupled secondary GH enzyme (co-expressed using a dual plasmid system) that will cleave only the product of the mutant glycosynthase reaction to release a chromogenic substrate but will not cleave the acceptor substrate necessary for the glycosynthase activity [20]. Pietruszka and co-workers from the Heinrich-Heine-University (Germany) have designed a glycosynthase activity assay using a silyl ether of p-nitrophenol for detection of the fluoride product released during from the activated

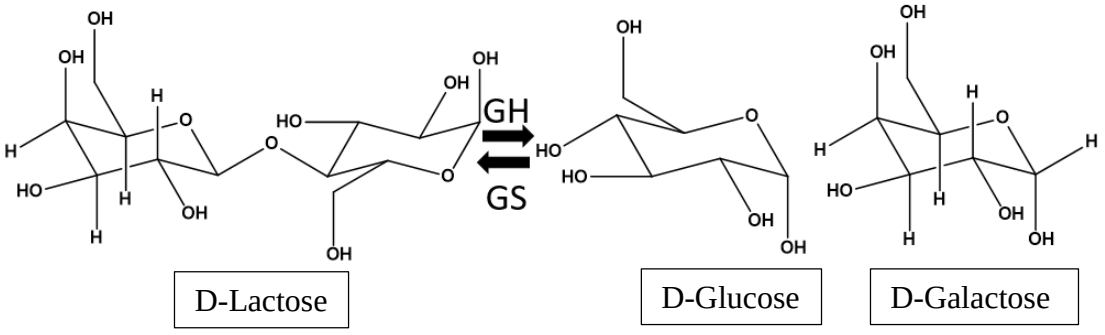
sugar donor for GS reaction at low concentrations [21]. Planas and co-workers from Universitat Ramon Llull (Spain) designed a glycosynthase screening method, specifically for glycosyl fluoride based sugar donors, where a specific chemical sensor is used to transduce fluoride concentration into a fluorescence signal [22]. Soham and co-workers from Technion-Israel Institute of Technology (Israel) developed another LB agar plate based universal screening assay for the directed evolution of glycosynthases for glycosyl fluoride based donor sugars [23]. All these pH or colorimetric LB-plate based screening methods are relatively low-throughput for identifying improved GS mutants, highly specific for a small class of GS families, and/or are mostly specific to only glycosyl fluoride based donor sugars. On the other hand, fluorescence based cell sorting techniques show much higher throughput potential for screening protein/enzyme variants for diverse applications. Jeong and co-workers from KAIST (Korea) have applied high-throughput Fluorescence-activated cell sorting (FACS) methods for *E. coli* producing Poly-(3-hydroxyburate) using fluorescent dyes like BODIPY and Nile red [24]. Similarly Withers and co-workers have also utilized FACS for the directed evolution of glycosyltransferases (GTs) by using fluorescent dyes labeled substrates [25], [26]. However, to-date analogous FACS based screening tools have never been developed for screening mutant glycosynthases as part of a directed evolution approach to increase GS activity and is therefore the focus of this thesis.

To develop a FACS screening method for GS, a model fucosidase from GH 29 was chosen for this study. GHs are currently classified into 156 families based on sequence similarity in the Carbohydrate Active Enzymes or CAZy Database [27]. GH family 29

enzymes are fucosidases that catalyze the hydrolysis of α-(1,2), α-(1,3), or α-(1,4) linked glycosidic bonds between fucose and other sugars. Fucosylated glycoconjugates and other fucosylated biomolecules are involved in a number of necessary biological processes and the fucosidases responsible for their processing therefore play an important role in diverse biological systems. Deficiency of the fucosidase enzyme in the human body is associated with a lysosomal storage disorder called fucosidosis that causes growth retardation, mental retardation, and neurological deterioration [28]. If the FUCA1 gene in the humans, that encodes for an α-L-fucosidase, undergoes any deleterious mutations it can result in incomplete breakdown of glycolipids and glycoproteins that eventually causes cellular malfunction and even cell death. Currently there is no FDA-approved treatment for a rare-disease like fucosidosis except receiving a healthy bone marrow cell transplant to replenish levels of α-L-fucosidase in the patient [29], [30]. Since, α-L-fucosidase plays a crucial role in the life of humans, researchers have been studying mammalian fucosidase enzyme for several decades. However, in recent years there has been also interest to study the role of microbial fucosidase enzymes expressed in the gut microbiome and its role in modulating human health, particularly for pre-term infants [31]. Fucosylated oligosaccharides are a critical carbon source for the establishment of a healthy gut microbiome and therefore there has been tremendous interest to identify enzymes capable of synthesizing fucosylated products.

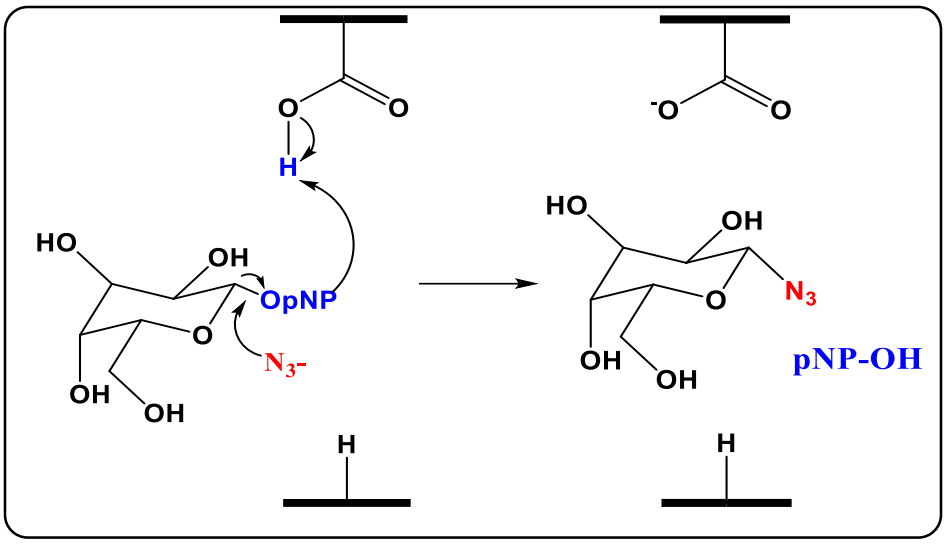
The model GH 29 family alpha-fucosidase chosen for this study was originally isolated from *Thermotoga maritima* and has a crystal structure available that was solved by the Withers group [32]. Several research groups have worked on GH29 enzymes as briefly

summarized below. Cobucci Ponzano and co-workers from the Institute of Protein Biochemistry (Naples, Italy) had identified the catalytic nucleophile site of this α-L-fucosidase enzyme and confirmed the chemical rescue of the catalytically inactive mutant [33]. Carsten Jers and co-workers from the Technical University of Denmark have identified seven novel α-L-fucosidase encoding genes from GH 29 family by functional screening of a soil-derived metagenomic library and then expressed as recombinant 6xHis-tagged proteins for the production of fucosylated human milk oligosaccharides [34]. Yaw-Kuen Li and co-workers from the National Chiao Tung University (Hsinchu, Taiwan) have further identified the essential residues of α-L-fucosidases isolated from the human genome and have tested their catalytic mechanism [35]. Dr. Li Chen and co-workers from Fudan University (Shanghai, China) have identified and characterized a core fucosidase isolated from the bacterium *Elizabethkingia meningoseptica* [36]. Our key aim here was to reverse engineer glycosyl hydrolases to alter their function from "breaking the glycosidic bond between sugar residues" to "catalyze the formation of glycosidic bonds" as exemplified through an example shown in Scheme 4. Here, GH or glycosyl hydrolases break the glycosidic



Scheme 4: GH breaking the bond between the sugar and GS catalyzing the synthesis of bond between sugars

bond between D-Lactose to produce D-Glucose and D-Galactose. Whereas a suitably engineered GS or glycosynthases or transglycosidase could synthesize a glycosidic bond between D-Glucose and D-Galactose to form D-Lactose. The literature suggests that most GH enzymes lose their activity upon mutation of the catalytic nucleophile to another amino acid residues that cannot participate in the reaction. Therefore, external nucleophiles like sodium azide and sodium formate or other anions of similar size and nucleophilicity (i.e., strong electron donors) have been used to drive the reaction forward to reactivate the nucleophilic mutant enzymes [19]. This phenomenon is termed in the literature as chemical rescue and has been reported for various classes of non-GH enzymes as well [37], [38]. A similar analogous approach has been applied in the field of chemical genetics to understand enzyme structure-function relationships [39]. Chemical genetics provides a complementary way to traditional protein engineering methods to probe cellular events using small molecules to modulate protein function. The expected reaction mechanism of chemical rescue of a model mutant GH on a pNP-based glycosyl based substrate is shown in Scheme 5. Anions with higher molecular weights and bulky structures cannot be used as nucleophiles despite of their optimal nucleophilicity because of steric hindrance in most catalytic pockets of enzymes. Since azides are nucleophilic and small in size they have been found to be fairly efficient as chemical rescue agents.



Scheme 5: Suspected mechanism of the chemical rescue reaction

These reagents are added to the donor sugar substrates as an azido group at the anomeric positions for a mutant GH that shows chemical rescue. This also suggests that mutant GH enzymes that are capable of showing chemical rescue are likely also good starting targets to be engineered into glycosynthases where the azide moiety can potentially act as a good leaving group as well if using azido sugars as the donor sugars to synthesize glycans. In the case of an active glycosynthase, if detected the released azide group can be easily detected then it would be possible to screen a large number of mutants to identify the best performing glycosynthase enzyme. Reaction of azides with fluorophore labeled alkyne groups would allow one to use traditional fluorescence based cell sorting techniques (e.g., FACS) for mutant screening. Here, we have developed an azide or azido-group based FACS screening approach for identifying better glycosynthases based on the concept of “click chemistry” where an azide and an alkyne group would react together to form a triazole. The triazole product formed in the case of copper-free click chemistry can also theoretically quench the fluorescence of a fluorophore group attached to the alkyne moiety which could allow one to fine-

tune the sorting parameters as highlighted later in this thesis. Click chemistry is considered a powerful tool in biology that has been used by several researchers in the last decade with broad applicability in the fields of chemical biology, synthetic chemistry, medicinal science, biochemistry, material science, pharmacology, and catalysis. Researchers have published work on various kinds of click reaction as briefly highlighted here. Copper or Cu-catalyzed click chemistry for biological applications in the field of glycosciences was reviewed by Xi Chen's group from the University of California-Davis [40]. Another review by Valery Fokin and Jason Hein from The Scripps Research Institute (California, USA) further highlights the development of Cu-catalyzed click chemistry for different fields of chemical biology [41]. Protocols for Cu-catalyzed click chemistry based bioconjugation studies and relevant applications have been highlighted by the Finn group from The Scripps Research Institute (La Jolla, California) [42], [43]. Zhang and co-workers from the National Institute of Biological Sciences (Beijing, China) have developed a CuAAC (Cu-catalyzed azide-alkyne cycloaddition) bioorthogonal reaction approach and highlighted applications of such bioorthogonal methods for both *in vivo* and *in vitro* systems [44]. Scientists working with CuAAC discovered that copper used in typical Cu-catalyzed click chemistry could be toxic to cells at certain concentrations. Therefore, researchers have avoided this issue by adding sodium ascorbate to eliminate Cu toxicity. However, sodium ascorbate can in turn also generate reactive oxygen species (ROS), thus there arose a need to design click reactions without the use of copper. Carolyn Bertozzi and co-workers from the University of California (Berkeley, USA), who is now at Stanford University, first developed bioorthogonal Cu-free click cycloaddition reactions and discussed its

applications in biological systems by using a strained alkyne like DIFO to drive the Click reaction forward due to the ring strain and electron donating property of fluorine groups attached to the ring in the absence of copper [45]. Bertozzi then later utilized this copper-free click chemistry to study the dynamics of glycan trafficking inside mammalian cells using azido sugars as the carbon source to identify the relative population of sialoglycoconjugates inside a cell and the sugar internalization kinetics using high resolution fluorescence imaging [46]. Since that seminal work, researchers have been routinely using Cu-free click chemistry for various glycobiology applications. This thesis demonstrates the use of copper free click chemistry to introduce a fluorophore label to be attached to released azides (or azido sugars) inside *E. coli* cells expressing mutant glycosynthases. We report proof-of-concept results that confirm the workflow needed to identify positive mutants based on different fluorescence intensities using a Fluorescence-activated cell sorting approach. Like the Withers group that utilized FACS to evolve sialyltransferases using a fluorophore based substrate probe [25], Uttamchandani and co-workers have recently also utilized a Fluorescence activated cell sorting method for the directed evolution of α-N-acetylgalactosaminidases using a quenched activity-based probe (qABP) [47]. However, to-date no one has yet developed a generic FACS based approach, specifically utilizing Click chemistry, for the directed evolution of glycosynthases, glycosyl hydrolases, or Transglycosidases.

1.2 Aims/Objectives of thesis:

1. To identify a potential glycosynthase mutant among the various catalytic nucleophilic mutants (i.e., Alanine, Serine and Glycine) of a model GH 29 fucosidase enzyme based on the hydrolytic activity of the mutant on α-p-nitrophenyl-fucopyranoside and chemical rescue using exogenous nucleophiles like sodium azide and sodium formate.
2. Optimization of *in-vitro* and *in-vivo* click chemistry conditions based on the reaction of strained fluorescent alkyne DBCO-PEG4-Fluor 545 with both organic and inorganic azides along with confirmation of permeation of fluorescent alkyne probe into *E. coli* cells using confocal microscopy.
3. Finally, generation of random GH 29 fucosidase mutant library using error-prone PCR and mutant library diversity confirmation to identify a potentially improved glycosynthase mutant with increased catalytic activity and stereospecificity upon sorting the library of random mutants using Fluorescence activated cell sorting (FACS).

Chapter 2: Site-directed mutagenesis, expression, purification, chemical rescue, GS reaction for model GH 29 family fucosidase

2.1 Materials and methods

2.1.1 Generation of designs

A model GH 29 family fucosidase enzyme isolated from a hyperthermophile *Thermotoga maritima* was chosen for this study, also known as Tm-alpha-fucosidase (TmAfc). The native (or wild type) gene Tm0306 that encodes TmAfc was codon optimized for *E. coli* expression and custom synthesized with AsiSI and BamH1 restriction sites specific flanking residues in pUC57 by Genscript Biotech Corporation (Piscataway, NJ). The Tm0306 gene was sub-cloned from pUC57 into our customized pEC vector using restriction cloning [48]. The catalytic nucleophile of TmAfc was identified as aspartic acid at the 224^{th} position (D224) [49] and mutated in to either alanine, serine, or glycine using standard site-directed mutagenesis protocols. Briefly, 0.5 µM of forward and reverse primers for mutagenesis (indicated in Table 1) was mixed with 20 ng of plasmid DNA in a 10 µl reaction volume. The reaction was performed using 1X Master Mix (Phusion DNA polymerase, 200 µM dNTPs, 1X Phusion HF buffer, 1.5 mM $MgCl_2$) with 5% DMSO and the reaction volume was made up by adding nuclease free PCR water. The PCR conditions used for mutagenesis are indicated in Table 2 below.

Primer name	Primer sequence	Melting temperature
Tm0306_D224A_Forward	GATGTTCTGTGGAACGCCATGGGTTGGCCGGAG	69 °C
Tm0306_D224A_Reverse	CTCCGGCCAACCCATGGCGTTCCACAGAACATC	69 °C
Tm0306_D224S_Forward	GATGTTCTGTGGAACTCCATGGGTTGGCCGGAG	67.3 °C
Tm0306_D224S_Reverse	CTCCGGCCAACCCATGGAGTTCCACAGAACATC	67.3 °C
Tm0306_D224G_Forward	GATGTTCTGTGGAACGGCATGGGTTGGCCGGAG	69 °C
Tm0306_D224G_Reverse	CTCCGGCCAACCCATGCCGTTCCACAGAACATC	69 °C

Table 1: Site-directed mutagenic primer sequences for D224A, D224S, D224G mutants

Process	Temperature	Time
Initial Denaturation	98 °C	30 s
Denaturation	98 °C	10 s
Annealing	69 °C and 72 °C	30 s
Extension	72 °C	3 min 10 s
Final Extension	72 °C	5 min
Hold	10 °C	∞
No. of cycles	20	

Table 2: PCR conditions used for site-directed mutagenesis

The PCR reaction samples were prepared to be run on a DNA gel (0.7% agarose gel with 0.5 µg/ml ethidium bromide) for electrophoresis as follows: 5 µl of the PCR sample was mixed with 1 µl of the purple loading dye and loaded directly on to the gel alongside 5 µl of the DNA ladder. The gel was run at 120 V for 40 minutes. The reaction mixtures were then digested with 10 U of Dpn1 enzyme (New England Biolabs) at 37°C for 1 hour and transformed into E. Cloni 10 g competent cells (Lucigen, WI) using the Zymo

transformation kit and plated on LB agar plates. Several random colonies were selected for inoculation as starter cultures with kanamycin as antibiotic (50 µg/ml) for 12-15 hours growth at 37°C in an incubator-shaker. Plasmid DNA was next extracted from the overnight grown starter cultures using IBI Scientific mini-prep kit according to the manufacturer's instructions and verified by DNA sequencing (Genscript Biotech Corporation).

2.1.2 Protein expression and purification

The recombinant wild type (Tm0306_WT) and nucleophile mutants (Tm0306_D224A/S/G) DNA plasmids were then transformed into *E. coli* BL21(DE3) competent cells and plated on LB agar supplemented with 50 µg/ml kanamycin. A single colony was picked from each LB agar plate and used for inoculation of a 50 ml starter culture comprising of LB media supplemented with kanamycin antibiotic (50 µg/ml) and incubated at 37°C with shaking at 200 rpm for 12-16 hours. Overnight grown 50 ml starter cultures were next transferred in to 1000 ml LB media containing 50 µg/ml kanamycin and grown at 37°C until the culture density reached an OD_{600} of 0.4-0.8 (exponential phase for *E. coli* growth). The protein expression was then induced using 0.5 mM IPTG (Isopropyl-β-D-thio-galactoside, Catalog number: BP175510, Fisher Scientific) and cultures were incubated using the following conditions: a) 37°C for 4 hours, b) 25°C for 20 hours and c) 16°C for 24 hours, to identify the best expression conditions. The cell pellets were recovered by centrifugation of the cell cultures at 10,000 rpm for 15 minutes at 4°C.

The expressed his-tagged proteins were purified using Ni-immobilized metal affinity chromatography (IMAC). The cell pellets were suspended in cell lysis buffer (20 mM sodium phosphate, 500 mM NaCl and 20% glycerol, pH: 7.4), along with protease inhibitor

cocktail (1 µM E-64, 0.5 mM benzamidine and 1 mM EDTA) and lysozyme (10 µg/ml) and then lysed by sonication. After sonication, the cultures were centrifuged at 11,000 rpm for 60 minutes at 4 °C and the supernatant cell lysate (containing the desired soluble protein) were recovered. The his-tagged protein of interest was separated from the other undesired *E. coli* proteins using an IMAC column on NGC-FPLC (Bio Rad, Hercules, CA). Briefly, the Ni-column was equilibrated with the IMAC binding buffer (100 mM MOPS, 10 mM imidazole, 500 mM NaCl, pH: 7.4). The cell lysate was then loaded onto the column and the IMAC binding buffer was run to remove any non-specifically bound proteins from the IMAC column. The protein of interest was next eluted with the IMAC elution buffer (100 mM MOPS, 500 mM imidazole, 500 mM NaCl, pH 7.4). The protein was buffer exchanged using desalting columns (GE Healthcare, Catalog number: 17-0851-01) into long-term storage buffer (10 mM of 2-morpholin-4-ylethanesulfonic acid or MES, pH 6). For buffer exchange, the gravity columns were equilibrated with 6 column volumes of 10 mM MES pH 6.0 (1 column volume=5 ml). Next, 2.5ml of the eluted protein from NGC was run through the gravity column and buffer exchanged protein was then eluted using 3.5 ml of storage buffer (10 mM MES pH=6.0).

2.1.3 Purified protein characterization

Protein concentration estimation: The purified protein concentration was estimated using the Spectradrop UV spectrophotometer (SpectraMax M5e) based on 280 nm absorbance. The samples were spotted in duplicates of 5 µl each for Tm0306_WT and Tm0306_D224A/S/G and 5 µl of 10 mM MES pH=6.0 was taken as the blank on a SpectraDrop micro-volume sample slide. The absorbance at wavelength λ=280 nm was measured and protein concentration was determined based on the Beer-Lambert's formula.

The protein sequence was determined using the Geneious software and copy pasted into the ExPASy- ProtParam online web-tool to determine standard parameters like molecular weight and extinction coefficient (for all disulfide bonds reduced). Beer-Lambert's law was then used to determine the protein concentration:

$A = \varepsilon CL$

Here, A stands for absorbance at λ=280 nm, ε is the extinction coefficient and L is the length of the path between the micro volume microplate and the cover slide.

Protein gel electrophoresis: Purity of all enzymes was confirmed by SDS-PAGE based densitometric analysis using pre-cast stain-free (Bio-Rad) protein electrophoresis gels. Samples were prepared using 1:1 protein:SDS-PAGE buffer (i.e., 95% laemmli buffer+ 5% β-mercaptoethanol) and denatured at 95°C for 5 minutes. After denaturation, 10 µl of the sample was loaded alongside 5 µl of the protein molecular weight ladder standard and the gel was run at 200 V for 30-40 minutes. The gel was stained in Coomassie blue staining solution (50% ethanol, 40% water, 10% acetic acid, 0.1% Coomassie brilliant blue R 250, Catalog number: C.I. 42660) by microwaving for 45 seconds and incubating at room temperature for 10 minutes. The gel destaining was performed in destaining buffer (i.e., 50% water, 40% methanol and 10% acetic acid) by microwaving for 45 seconds and incubating at room temperature for 10 minutes, followed by water destaining overnight. The destained gel was imaged using Gel Doc EZ Imager (Bio-Rad Laboratories) and purity was estimated using Image Lab software based standard protein bands densitometric analysis.

2.1.4 Activity assays

The activity of the purified enzymes Tm0306_WT, Tm0306_D224A/S/G was evaluated using pNP-F (4-nitrophenol α-fucopyranoside) and 2Cl-4NPF (2-Chloro,4-nitrophenol α-fucopyranoside) as substrates that were purchased from Carbosynth Limited. For determining the hydrolytic activity of the enzymes, 1 µg of protein was added to 2 mM pNP-F and 2Cl-4NPF in a reaction buffer containing 50 mM MES pH 6 and incubated at 60°C for 1.5 hours. Blank wells with pNP-F and 2Cl-4NPF were also taken in buffer alone without the protein as controls. After 1.5 hours of the reaction, 100 µl of the reaction mixture was transferred to a new 96-well transparent well microplate with 100 µl of 1 M NaOH and the absorbance was measured at 410 nm in the UV spectrophotometer (SpectraMax M5e) to determine pNP or 2Cl-4NP absorbance upon substrate hydrolysis. pNP or 2Cl-4NP standards at different concentrations of 0.025 mM, 0.0625 mM, 0.125 mM, 0.25 mM, 0.5 mM, 1 mM, 2 mM were used to define the pNP calibration curve and find the relationship between the measured absorbance and estimated concentration.

2.1.5 Chemical rescue assays

In order to recover the hydrolytic activity of the nucleophile mutants, high concentrations of external nucleophiles like sodium azide and sodium formate (2 M) were added to reaction mixtures containing 1 µg of the protein with 2 mM substrate (either pNPF or 2Cl-4NPF) in 50 mM MES pH 6 and incubated at 60°C for 2 hours at 400 rpm. Blank wells with pNP-F were also taken in buffer alone without the protein as controls. After 2 hours of the reaction at 60 °C, 30 µl of the reaction mixture was transferred to a new 96-well transparent microplate and mixed with 70 µl of DI water and 100 µl of 0.1 M NaOH. The absorbance was measured at 410 nm to check for pNP release using the UV

spectrophotometer (SpectraMax M5e). pNP standards at different concentrations 0.025 mM, 0.0625 mM, 0.125 mM, 0.25 mM, 0.5 mM, 1 mM, 2 mM were taken to determine the pNP calibration curve to find the relationship between absorbance and concentration.

2.1.6 DNS reducing sugar assays

To estimate the amount of total reducing sugars in the reaction mixtures released after the chemical rescue experiment, a standard DNS (3,5 dinitrosalicylic acid) based colorimetric assay was performed. Briefly, 30 µl of the reaction mixture was added to PCR tubes or microplates and mixed with 60 µl DNS stock reagent as highlighted previously by Miller and others [50], [51]. 30 µl of glucose standards at the following concentrations: 0.1125 g/L, 0.45 g/L, 1.125 g/L, 2.25 g/L, 4.5 g/L were used to prepare the calibration curve. The sealed PCR tubes or plates were incubated at 95°C for 5 minutes and cooled at 10°C for 10 min. After incubation, 36 µl of the reaction mixture from the PCR tube or plate was transferred to a clear bottom microplate and mixed with 160 µl of DI water. The contents of the reaction mixture were mixed well and read for total absorbance at 540 nm using a UV Spectrophotometer (SpectraMax M5e). The calibration curve was built using DNS standards to find the relation between the absorbance and concentration.

2.1.7 Glycosynthase *in-vitro* reaction assay

For evaluating the glycosynthetic activity of Tm0306_WT and Tm0306_D224G, 40 µg of the protein was added to a mixture of 10 mM β-L-fucopyranosyl azide (Catalog number: 66347-26-0, Chemily Glycosciences) and 50 mM pNP-β-D-Xylose (Carbosynth Limited) and incubated at 60°C for 24 hours in 50 mM MES buffer pH 6.0. The reaction mixture was analyzed using Thin Layer Chromatography (TLC) using Silica Gel 60 F254 TLC plates from Merck & Co.. The mobile phase used for TLC was ethyl acetate: methanol:

water (at 70:20:10 v/v ratios). Standards were also run on the TLC plate to determine the unknown detected spots in reaction sample based on retention factor (R_f) value. The plate was epi-illuminated and directly imaged under UV light at wavelength λ=305 nm. The plates were then sprayed with visualization solution containing 0.1% orcinol dye in 10% H_2SO_4, then dried and heated at 100 °C for 15 min. The charred TLC plates were then imaged and analyzed using Gel Doc EZ Imager from Bio-Rad. The intensity of spots from the UV image was evaluated using Image J software to determine the percentage conversion of the reactants into the fucosynthase or hydrolysis products.

2.2 Results and discussion

2.2.1 Cloning, protein expression, purification and characterization

Site-directed mutagenesis was performed on pEC_Tm0306_WT to create the nucleophilic mutants pEC_Tm0306_D224A/S/G. Figure 1 shows the agarose DNA gel image that was run with the PCR products at different annealing temperatures after site-directed mutagenesis of Tm0306_WT into the three desired nucleophile site mutants Tm0306_D224A/S/G. The wild type plasmid contains 6,304 base pairs of nucleotides and the bands in the gel shown in Figure 1 also correspond to the expected size of the plasmid. Furthermore, Dpn1 digestion was performed on the remaining PCR mixtures and transformed into E.cloni 10 g cells and plated on LB agar plate. Random colonies

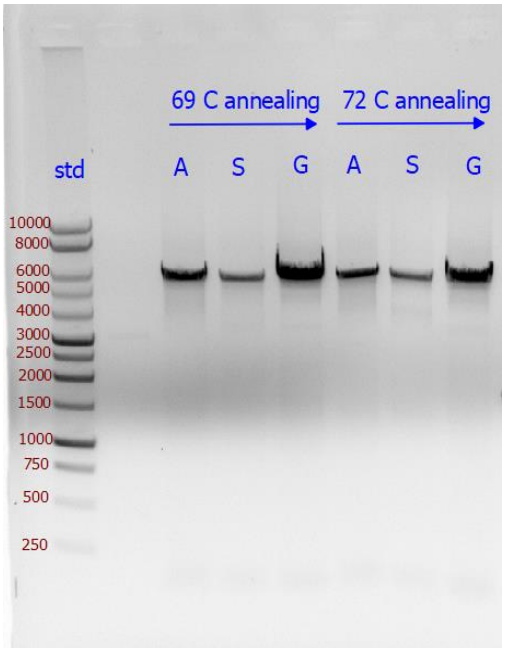


Figure 1: DNA gel for site-directed mutagenesis of Tm0306_WT and Tm0306_D224A/S/G (Image obtained by Chandra)

growing on the plate were selected and grown overnight in LB media to extract the DNA plasmid via miniprep. The samples were then sent for DNA sequencing and the mutations at the catalytic nucleophile site were confirmed using the Geneious software.

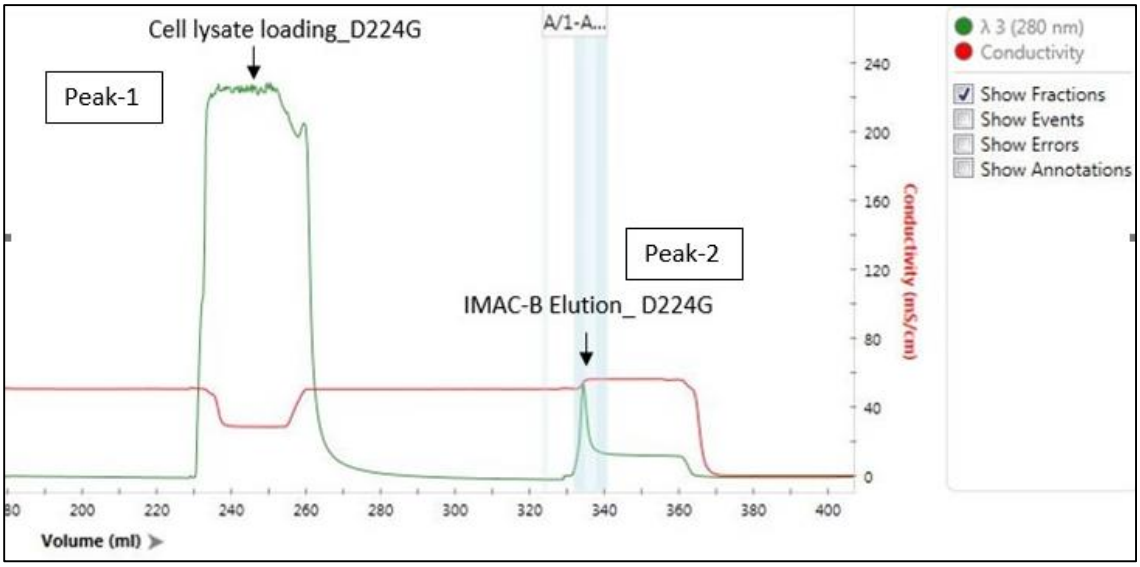


Figure 2: NGC chromatogram for the purification of Tm0306_D224G protein

The positively identified plasmid DNA constructs were transformed into expression strain *E. coli* BL21 (DE3) and protein expression was carried out as reported below. The proteins were purified using Ni-affinity IMAC on the NGC-FPLC. Figure 2 shows a representative chromatogram obtained from NGC-FPLC instrument during the protein purification run for the Tm0306_D224G construct. The green line in the chromatogram stands for λ=280 nm absorbance reading used to monitor protein elution from the column while the red line stands for the conductivity of the eluting mobile phase. Initially, the NGC-FPLC flow cell/tube lines and IMAC column were equilibrated using the IMAC binding buffer until λ280 nm is at baseline and constant. As observed in the chromatogram, the cell lysate (which includes the expressed protein of interest) was loaded after column equilibration that gave an elution peak (peak-1) likely comprised predominantly of undesired *E. coli*

background or non-tagged proteins present in the cell lysate. The column was washed with the IMAC binding buffer to remove the undesired non-tagged proteins from the column. The IMAC elution buffer was then run through the column to elute his-tagged protein shown as peak-2, which was collected for further analysis. The eluted protein was buffer exchanged from the IMAC elution buffer in to the 10 mM MES pH 6 storage buffer using size-exclusion gravity columns. The protein was transferred as 50 µl aliquots in to PCR tubes and flash-frozen using liquid nitrogen and stored at -80°C for future experiments. The protein concentration was measured using 280 nm absorbance readings and approximately 1 mg of the final purified protein was obtained from 1 L starting cell culture. Figure 3 shows the SDS-PAGE gel that confirms the purity and expected molecular weight (~53 kDa) of all the proteins reported in this work.

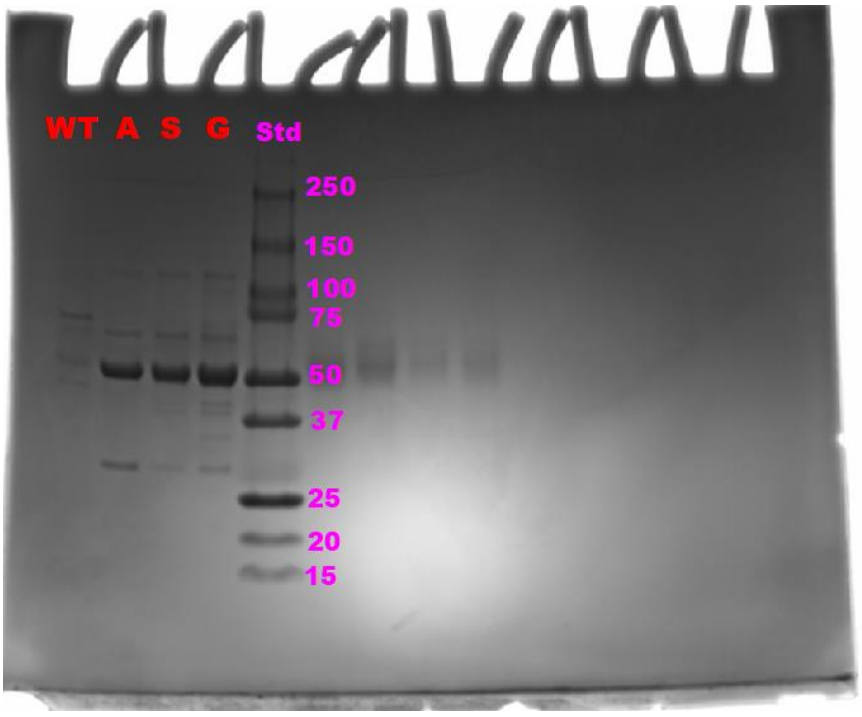
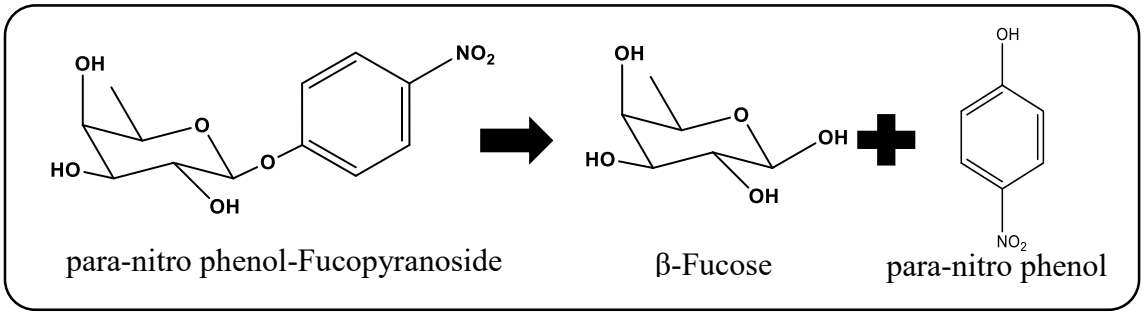


Figure 3:SDS-PAGE Gel for Tm0306_WT and Tm0306_D224A/S/G

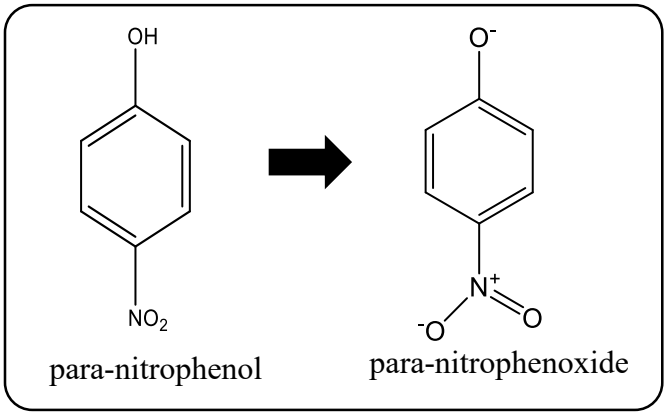
2.2.2 Activity assays of Tm0306_WT and Tm0306_D224A/S/G

The hydrolytic activity of Tm0306_WT and Tm0306_D224A/S/G was evaluated using the substrate para-nitrophenyl-fucopyranoside (pNPF) at 60°C using a 10 mM MES pH 6.0 reaction buffer. The wild-type fucosidase is expected to hydrolyze pNPF in to para-nitrophenol and fucose as shown in Scheme 6.



Scheme 6: Release of pNP observed by reaction of enzyme and substrate pNP-fucose

Para-nitrophenol exists as para-nitrophenoxide anion in solution that is responsible for the yellow color observed during pNP release as shown in Scheme 7. Para-nitrophenoxide has absorbance maxima at 405 nm (ε = 18.3 to 18.4 mM^{-1} cm^{-1}).



Scheme 7: Conversion of para-nitrophenol to para-nitrophenoxide in solution

The absorbance values of pNP standards were used to draw the pNP calibration curve as shown in Figure 4. A linear curve was fitted through the data points using Origin Pro

software and the R-square value and the slope of the graph were determined from the linear regression fit. The R-square value for the fit was 0.99 and the slope of the fitted line was calculated to be 4.0221.

Using the formula, Absorbance= Slope of the graph*Concentration of pNP released, the unknown concentration of pNP released in each reaction well was determined. The absolute concentration of pNP released in wells was calculated by deducting the average blank from the reaction wells. The percentage conversion was calculated by using the following formula:

$$\%\ conversion = \frac{Initial\ substrate\ concentration - Final\ substrate\ concentration}{Initial\ substrate\ concentration} \times 100\%$$

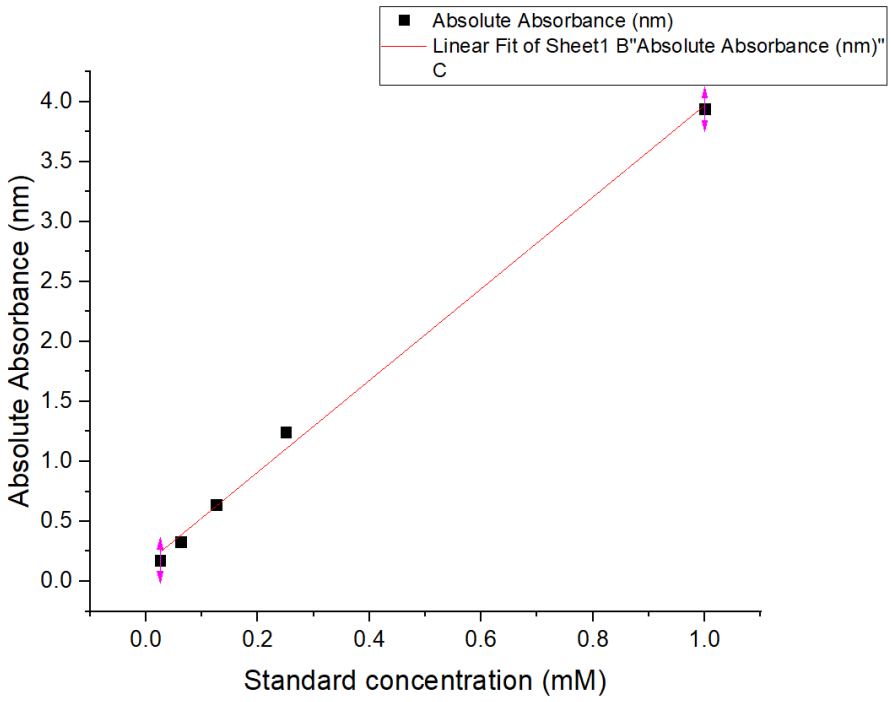


Figure 4: Plot of standard curve between absorbance at 405 nm and pNP concentration (mM)

The percentage conversion values for triplicates samples were averaged and Figure 5 shown depicts the relative % conversion activity of Tm0306_WT versus Tm0306_D224A/S/G.

It can be observed from Figure 5 that Tm0306_WT gave close to 94% conversion of pNP-fucose in to pNP while the D224A and D224S mutant enzymes showed no measurable activity. This was expected since D224 is a true catalytic nucleophile that plays a major role in the hydrolysis mechanism by attacking anomeric carbon of the substrate to form an intermediate enzyme-substrate or Michaelis-Menten complex. Interestingly, the D224G mutant showed marginal activity of ~11% hydrolysis of pNPF suggesting partial recovery of hydrolytic activity. These results are in agreement with results reported by Cobucci and co-workers [19].

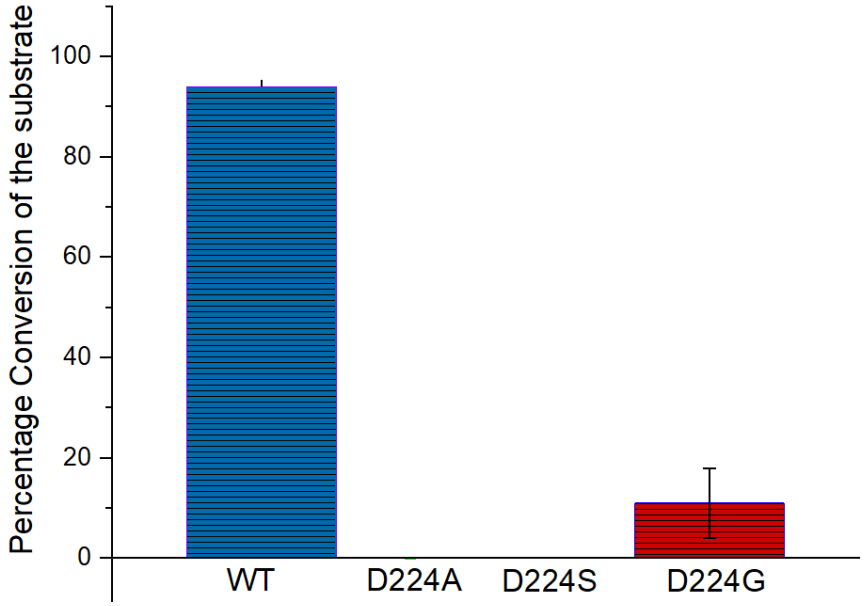


Figure 5: Activity assay of Tm0306_WT and Tm0306_D224A/S/G with pNP-fucopyranoside

2.2.3 Chemical rescue or reactivation of Tm0306 nucleophile mutants

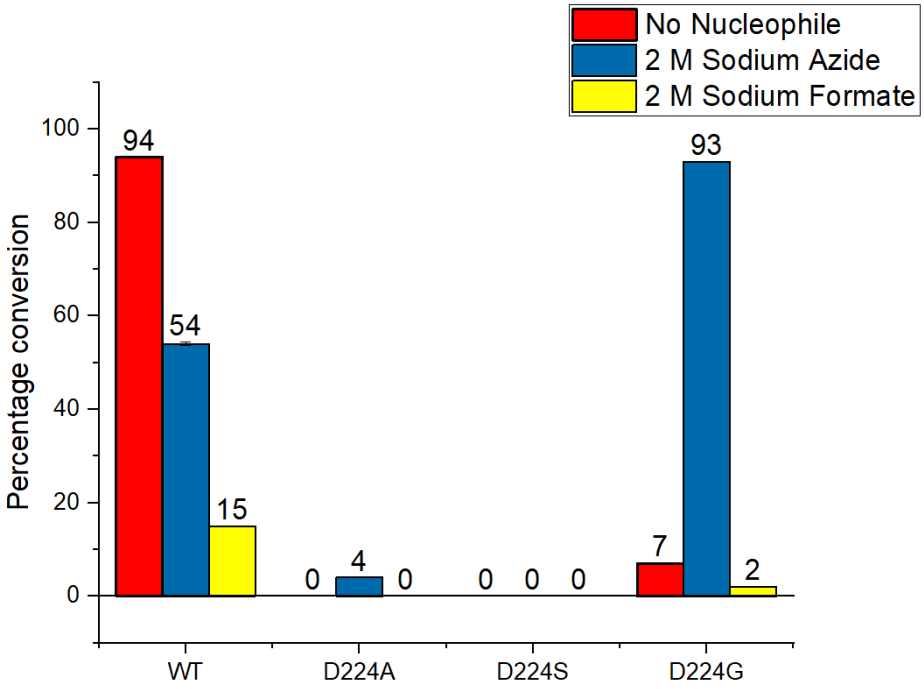


Figure 6: Chemical rescue of Tm0306-D224A/S/G mutants with 2M sodium azide and sodium formate nucleophiles at 60 °C in 50 mM MES buffer, pH=6.0.

As highlighted in the previous section, the catalytic nucleophile mutants Tm0306_D224A/S/G showed either no activity or marginal hydrolytic activity. Exogenous nucleophiles such as sodium azide (NaN_3) and sodium formate (HCOONa) have been used in the past to identify the true nucleophile residue by rescuing the hydrolytic activity of GH2 family mutant enzymes [52]. The chemical rescue experiments were carried out using pNPF as the substrate with enzymes Tm0306_D224A/S/G as well as Tm0306_WT in the presence of 2 M sodium azide and sodium formate as the exogenously added external nucleophiles in 50 mM MES buffer pH 6.0 at 60°C.

We observed that the activity of D224A and D224S mutants was not significantly rescued with either sodium azide or sodium formate while D224G gave fully restored activity in

the presence of sodium azide alone. Comparing the side chain moieties of Alanine, Serine and Glycine, it should be noted that the glycine side chain (–H) is much smaller as opposed to -CH3 and –CH2-OH bulkier side chains in alanine and serine, respectively (Figure 7). The smaller side chain of glycine most likely enables the azide to readily occupy the cavity created upon mutation of aspartic acid residue, which might not be the case with alanine and serine due to steric hindrance. Since azide acts as a pseudo-nucleophile attacking the docked pNP-fucose substrate to facilitate the release of pNP, it is critical for the pseudo-nucleophile to orient correctly in the active-site pocket created upon mutation to glycine. Analogously, the formate anion is larger in size than azide that could hinder its proper docking and orienting in the active site thus making it a poor chemical rescue agent. Hence, D224A and D224S mutants were not analyzed further and Tm0306_D224G was identified as the potential glycosynthase mutant because of its optimal chemical rescue behavior with azide for future work.

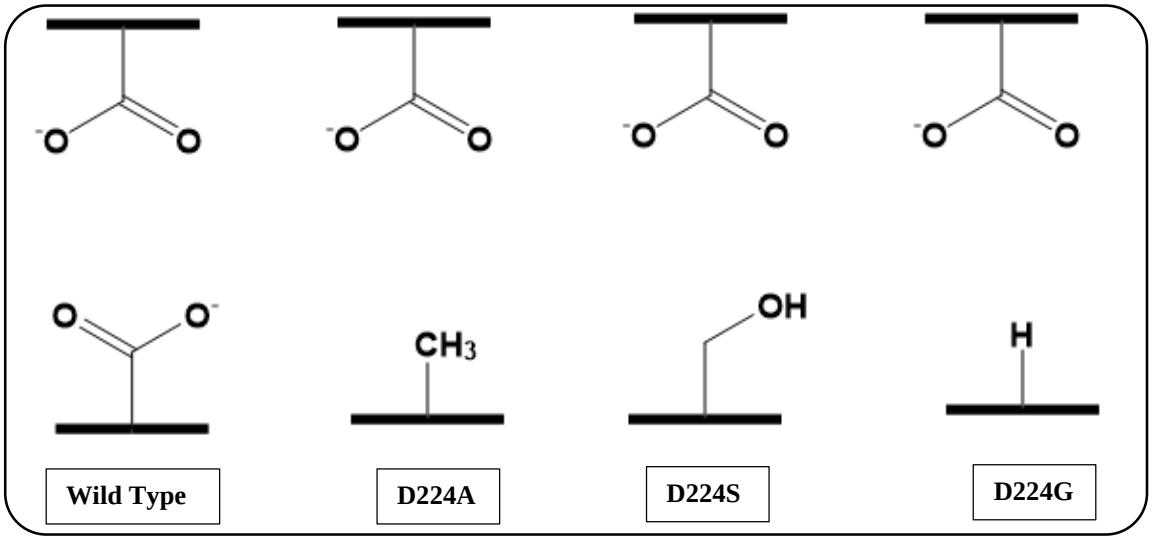


Figure 7: Active site of the fucosidase enzyme: Tm0306_WT and mutants Tm0306_D224A/S/G

We also observed that Tm0306_WT lost significant activity upon addition of external nucleophiles like sodium azide. Sodium azide concentration in the reaction wells (2M) was very high which might be influencing the hydrolytic activity of the WT enzyme. Thus, we also tested lower sodium azide and sodium formate concentrations (0.1M, 1M, 2M) for Tm0306_WT and Tm0306_D224G. Figure 8 and Figure 9 show the effect of varying sodium azide and sodium formate concentrations on Tm0306_WT and Tm0306_D224G enzymes in 50 mM MES buffer pH 6.0. As the concentration of the sodium formate increases in the reaction mixture, the activity of Tm0306_WT reduces as well. Sodium formate is clearly shown to also have an inhibitory effect on the activity of wild type enzyme. Also, lower concentrations of sodium formate also did not rescue the activity of Tm0306_D224G mutant suggesting that the structural steric hindrance in the active site was likely the major issue. On the other hand, increasing sodium azide concentration did not have significant impact on lowering enzyme activities and chemical rescue was seen at all tested concentrations.

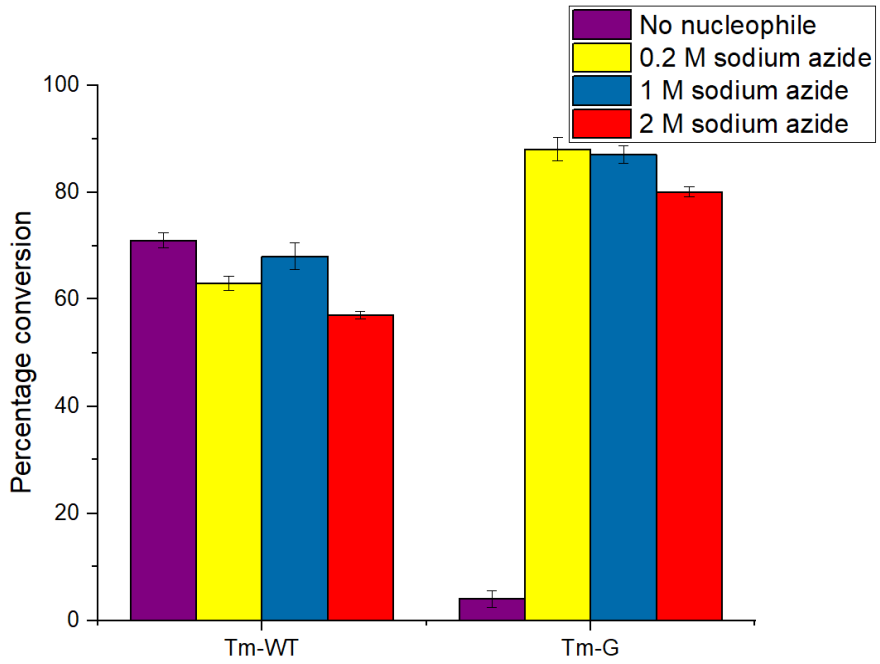


Figure 8: Effect of varying concentrations of sodium azide on Tm0306_WT and Tm0306_D224G enzymes

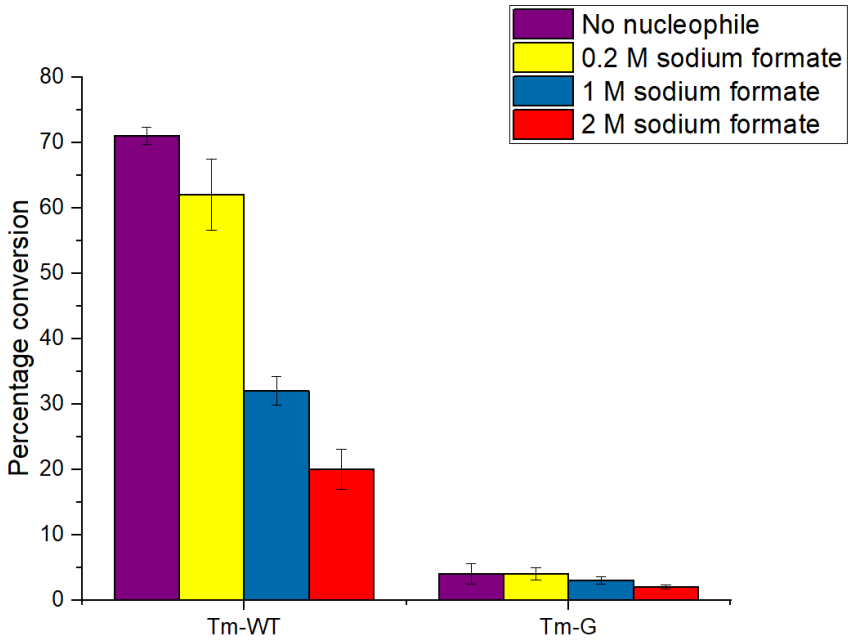
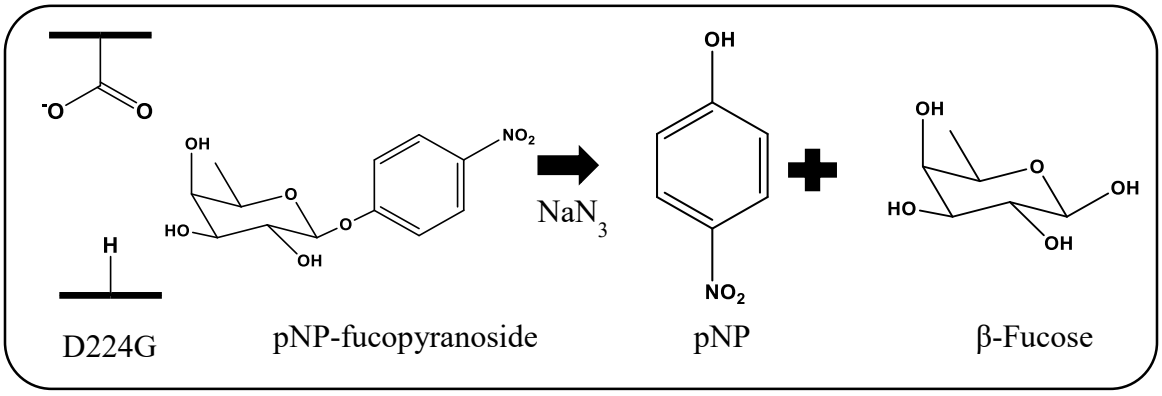


Figure 9: Effect of varying concentrations of sodium formate on Tm0306_WT and Tm0306_D224G enzymes

2.2.4 DNS assay to determine reducing sugars released during chemical rescue

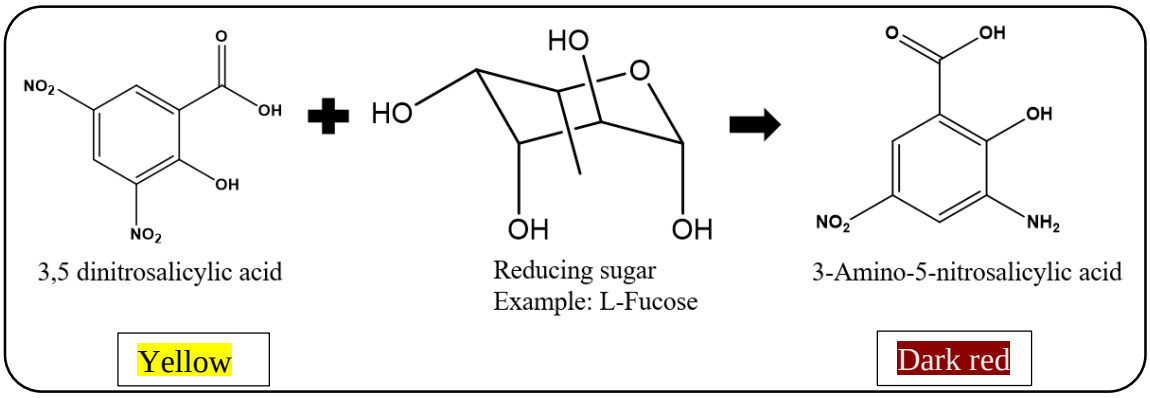
From the chemical rescue experiments, we identified Tm0306_D224G as a potential glycosynthase mutant. Ideally during chemical rescue the mutant Tm0306_D224G should have cleaved pNP-fucose to form pNP and β-Fucose as shown in Scheme 8.



Scheme 8: Hypothesized reaction of Tm0306_D224G in the chemical rescue experiment

If Scheme 8 holds true, then pNP and fucose should have both been released in the reaction mixture. The release of pNP was indeed observed by increased absorbance at λ=405 nm under alkaline conditions as shown in Figure 8 and Figure 9. However, to also check the fucose concentration in the reaction mixture, a reducing sugar or DNS assay was performed. DNS or dinitrosalicylic acid is an aromatic compound that reacts with reducing sugars in the reaction mixture (like fucose) and forms 3-amino-5 nitrosalicylic acid which shows increased absorbance at its characteristic wavelength of 540 nm as shown in Scheme 9 and the color of the reaction mixture changes from yellow to dark red upon formation of 3-amino-5 nitrosalicylic acid. The absorbance values of the reaction mixtures were measured at 540 nm, which were used to then build the calibration curve for DNS standards (i.e., glucose standards of known concentrations) and to determine the percentage

conversion of pNPF into fucose in the reaction mixture and provide an alternative approach to close the mass balance for the chemical rescue assays.



Scheme 9: Reaction of DNS (dinitrosalicylic acid) with a reducing sugar like fucose

Calculations for slope, concentration and percentage conversion were carried out as described in Section 2.2.1. From the DNS calibration curve (Figure 10), the R-square value was found to be 0.9995 and the slope calculated was 0.3279. The estimated concentration of reduced sugars in each reaction mixture is plotted in Figure 11.

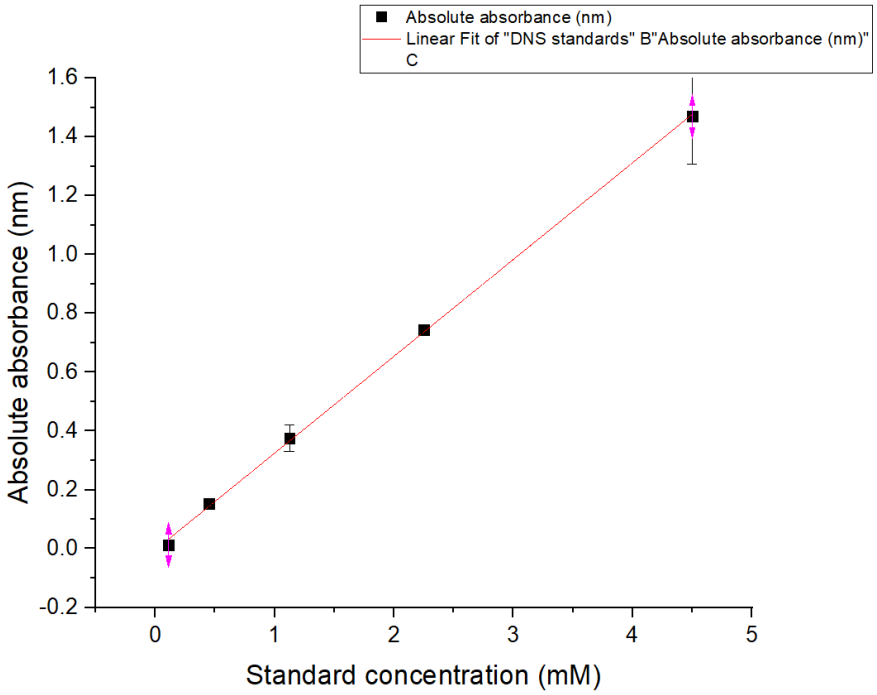


Figure 10: Plot of glucose concentration (mM) and absorbance measured at 540 nm

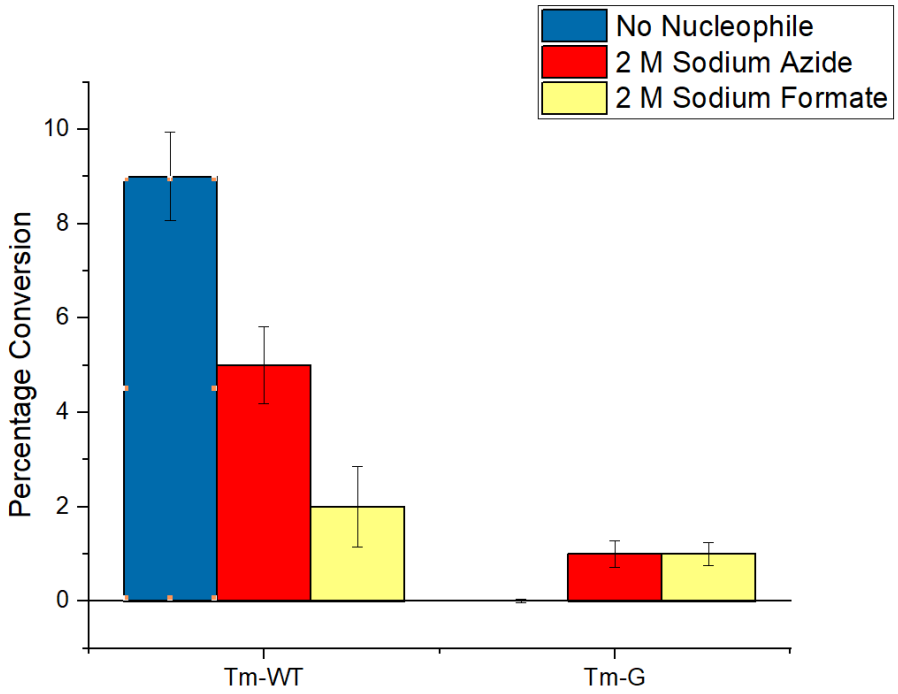
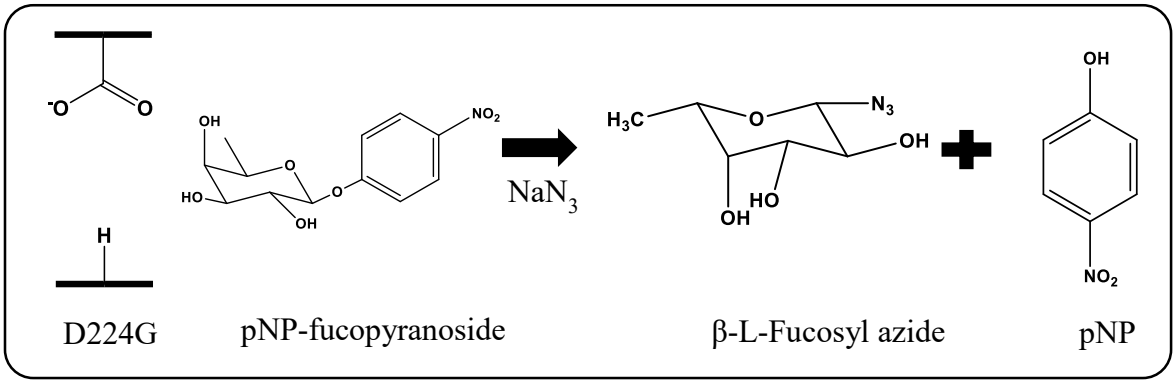


Figure 11: DNS assay results for Tm0306_WT and Tm0306_D224G

On comparing Figure 11 with Figure 8 and Figure 9, we observed that Tm0306_D224G showed approximately 86% activity rescue for reaction with pNP-F and 2M sodium azide. But, the DNS assay results suggest that the amount of fucose released is very low significant quantities suggesting that there might be an alternative reaction scheme also taking place in addition to Scheme 3. During the chemical rescue reaction pNP was released but fucose was not formed, which suggested an intermediate compound β-L-fucopyranosyl azide (data not shown) is likely being formed as a by-product in the reaction mixture as shown in Scheme 10 as suggested previously by Cobucci and co-workers [19].

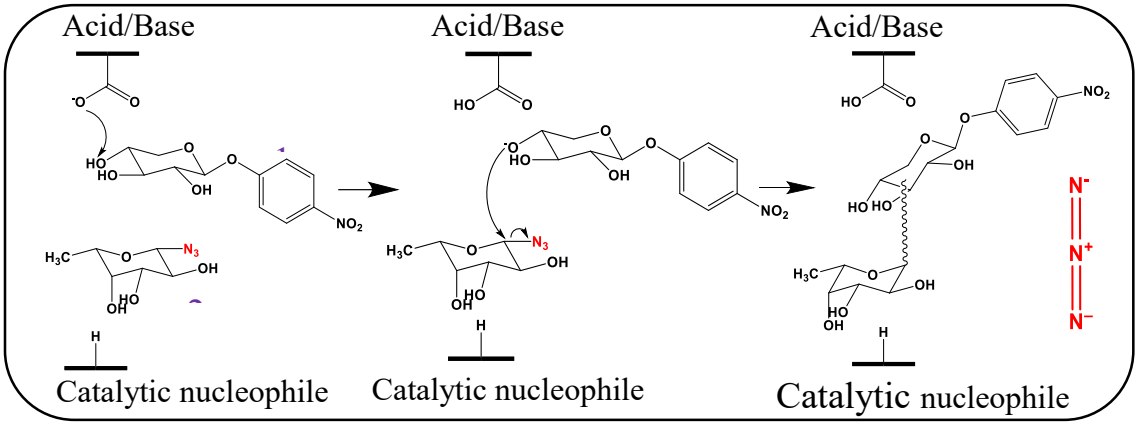


Scheme 10: Hypothesized reaction of Tm0306_D224G with pNP-fucopyranoside in the presence of sodium azide.

2.2.5 Glycosynthase reaction *in-vitro* assays

Glycosynthases (GS) are catalytic mutants of retaining glycosidases that catalyze the synthesis of glycosidic linkages between corresponding donors and suitable acceptor sugars to drive the synthesis of oligosaccharides. In our experimental setup, the donor sugar is β-L-fucopyranosyl azide and the acceptor sugar is pNP-β-D-xylopyranoside as reported previously by Cobucci and co-workers [19]. In Scheme 11, we have outlined the expected

glycosynthase reaction mechanism for our mutant enzyme based on the classical SN2 type mechanism. First, the acid/base residue of D224G deprotonates the –OH group at the C4 position of pNP-β-D-xylopyranoside (1). Next, the deprotonated pNP-β-D-xylopyranoside anion would attack the anomeric carbon of β-L-fucopyranosyl azide (2) present near/inside



Scheme 11: Glycosynthase reaction mechanism of Tm0306_D224G with pNP-β-D-xylopyranoside and β-L-fucopyranosyl azide

the catalytic nucleophile D224G and azide ion would be released. Finally, pNP-β-D-xylopyranoside forms a bond with β-L-fucose to produce β-L-fucopyranoside- β-D xylopyranoside- pNP (3), which is the ultimate product of this glycosynthase reaction. The *in-vitro* GS reaction was carried out using Tm0306_D224G added along with β-L-fucopyranosyl azide and pNP-β-D-xylopyranoside (1:5 molar ratio of donor to acceptor sugar) and incubated at 60 °C or 37 °C for 24 hours. Reaction products detection was performed using Thin-layer chromatography (TLC) and Image-J software to perform densitometric analysis to quantify final GS reaction product yields.

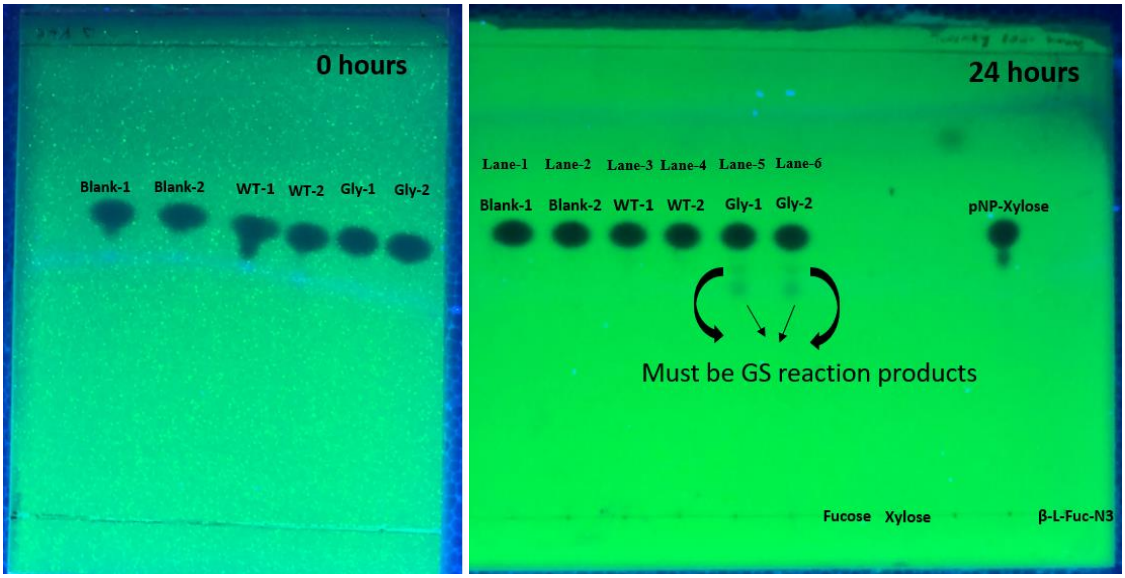


Figure 12: pNP image of the TLC analysis of glycosynthase reaction of D224G with β-L-fucopyranosyl azide and pNP-β-D-xylopyranoside in 50 mM MES pH 6.0 at 60 C at 0 and 24 hours.

From Figure 12, we can observe that at the zeroth hour, only pNP-β-D-xylopyranoside can be visualized on the TLC plate under UV light. Under UV light, only products with a fluorescence or with a fluorophore moiety would be readily visible. After 24 hours, we observe that there is formation of two products that are likely the previously reported glycosynthase reaction products formed by the reaction of Tm0306_D224G with β-L-fucopyranosyl azide and pNP-β-D-xylopyranoside. We cannot see fucose, xylose or β-L-fucopyranosyl azide under the UV image, as these reagents are not fluorescent. These results also corroborate with the Cobucci paper [19]. The two products reported in the previous study were determined to be α-L-Fuc-(1,4)-β-D-Xyl-pNP and α-L-Fuc-(1,3)-β-D-Xyl-pNP. These products were obtained in a yield of 55% and 45%, respectively.

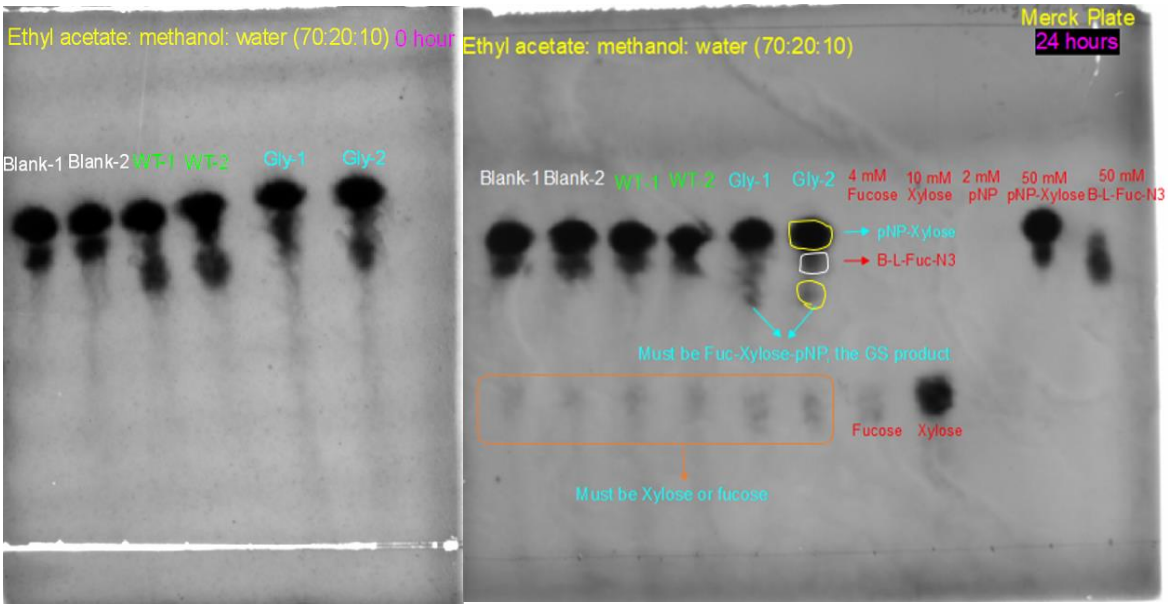


Figure 13: Gel Doc image of the TLC analysis of glycosynthase reaction of D224G with β-L-fucopyranosyl azide and pNP-β-D-xylopyranoside in 50 mM MES pH=6.0 at 60 C at 0 and 24 hours.

The orcinol visualization reagent was then sprayed on the TLC plate and then incubated at 100 °C for 15 minutes. The plate was imaged in the Gel Doc EZ Imager under epi-illumination using white light to perform densitometric analysis on the charred spots observed using the visualization agent Figure 13. From Figure 13, we can observe that one of the potential glycosynthase reaction products cannot be clearly observed after Orcinol staining. This is because residual β-L-fucopyranosyl azide and the glycosynthase reaction product/s have similar retention factors (Note that Retention factor is a ratio of analyte distance to the solvent distance in a TLC plate with respect to point of sample origin/application). Thus after Orcinol staining, the glycosynthase reaction product is hidden behind the dominating spot of residual unreacted β-L-fucopyranosyl azide. We also observe some amount of minor hydrolysis taking place for the donor sugar (β-L-fucopyranosyl azide) as evident from release of fucose that is also seen on the plate. We also performed Image J spot intensity analysis using the UV image to quantify the total GS

product yield. For Image J intensity analysis, Image type 32 bit was chosen to convert the picture in to a black and white image before densitometric analysis. Then, the lane containing the desired spot was selected and plotted using the plot lanes option. The software gives peaks equal to the number of spots in the particular lane that was chosen.

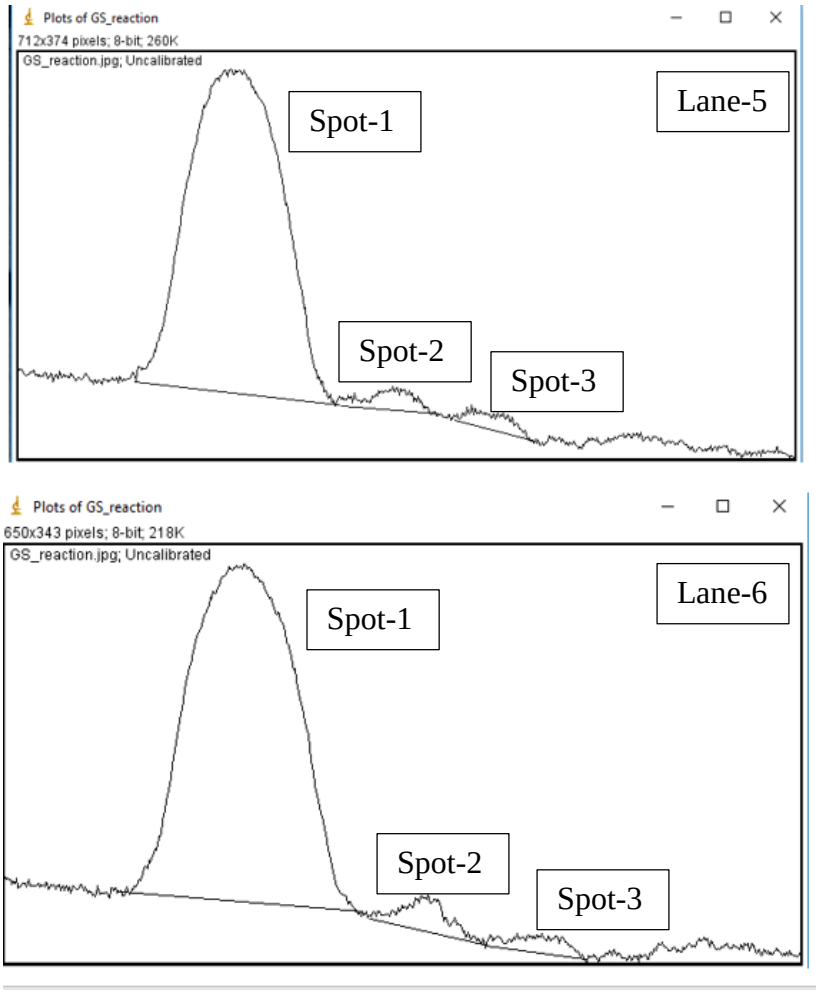


Figure 14: Image J analysis of Lane 5 and Lane 6 of Figure 12. Peak depicting the intensity of the spot.

We found one plot each for the duplicates of Tm0306_D224G. We observed that one peak is obtained corresponding to the number of spots as shown in Figure 14. The intensity of the spots was determined and thus the percentage conversion of the non-limiting reagent

(i.e., acceptor sugar pNP-X) into the final GS product. The values were around 5-6% for each spot, which is an extremely low yield of the product as shown in Table 3. This suggests that D224G is not a very efficient glycosynthase mutant but is still suitable for further development of the FACS based screening platform.

	Spots	D224G	D224G	Blank		% Conversion with blank as denominator
pNP-Xylose	Spot 1	33735.3	29747.96	34650.18	pNP-Xylose	
Product 1	Spot 2	1196.882	1039.933	35845.29	Sum of D224G spots	6%
Product 2	Spot 3	913.104	684.569	31472.46	Sum of D224G spots	5%

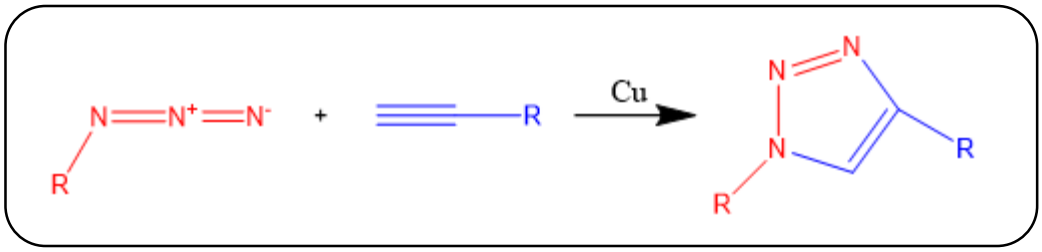
Table 3: Spot intensity analysis of the GS reaction in Figure 12.

Chapter-3: *In-vitro* and *in-vivo* click chemistry optimization for azide detection

3.1 Introduction

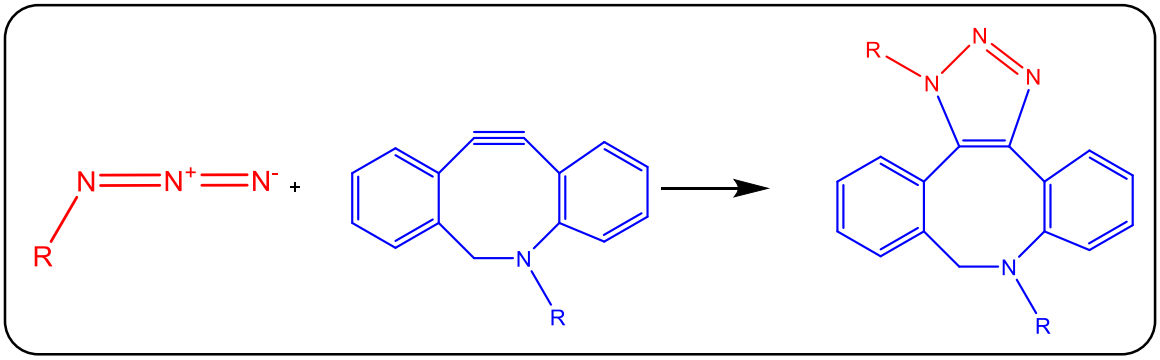
Dr. Karl Barry Sharpless of the Scripps Research Institute proposed the concept of click chemistry in 2001. Click chemistry refers to a reaction where a new molecule can be formed "quickly" based on the chemical ligation/reaction of two suitable chemical moieties. The most common click chemistry reaction is between an alkyne and an azide/azido group to yield cyclic triazole. A click reaction also has the following properties: it uses benign solvents like water, is highly stereospecific, does not need extreme conditions of temperature and pressure, has very high product yields, generates non-toxic byproducts that can be removed without chromatography, and the products are physiologically stable. There are various kinds of click reaction between azide-alkynes: (i) Copper-catalyzed Azide-Alkyne cycloaddition (CuAAC), (ii) copper free click reactions like Strain promoted azide alkyne cycloaddition (SPAAC), (iii) Strain promoted Alkyne-nitrone cycloaddition (SPANC), and (iv) Tetrazine-Alkene ligation [53].

Copper-catalyzed click chemistry: Reaction between a terminal alkyne and an aliphatic azide group in the presence of copper to yield a triazole is classified as Cu-catalyzed click reaction. Scheme 12 gives a pictorial representation of the Cu-catalyzed click reaction [54]. The presence of copper accelerates the reaction rate and provides higher yields than the uncatalyzed reaction. Extreme conditions of temperature are not required for this reaction; it can proceed even below room temperature.



Scheme 12: Copper-catalyzed click chemistry between alkyl azide and terminal alkyne

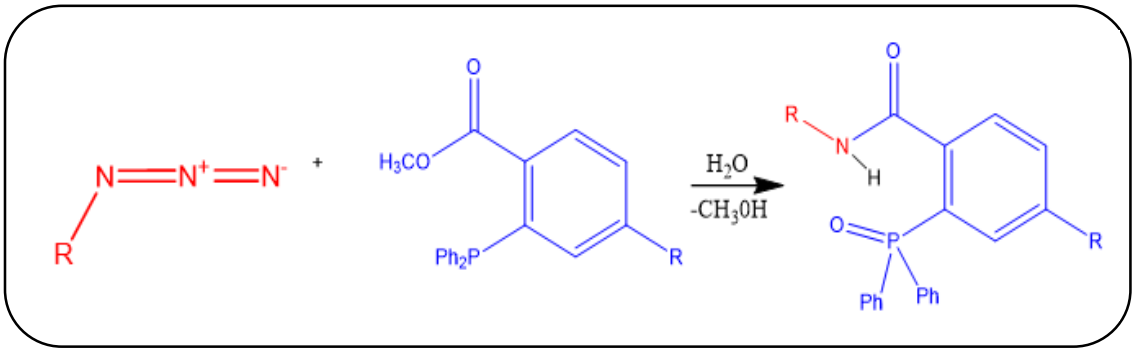
Strain-Promoted alkyne-azide cycloaddition: This reaction is an extension to the classical CuAAC reaction. Since the cytotoxicity of Cu affects the reaction, sodium ascorbate is added to nullify copper toxicity. But this leads to the formation of reactive oxygen species instead. So, strain-promoted alkyne-azide cycloaddition was designed to carry out the click reaction without the use of copper. Reaction between a strained cyclooctyne and phenyl or aliphatic azide to yield a triazole is termed as Cu-free click reaction as shown in Scheme 13. The ring strain of the cyclooctyne makes the reaction proceed forward at a faster rate without the use of cytotoxic transition metals like Cu.



Scheme 13: Strain-promoted alkyne-azide cycloaddition between alkyl azide and aromatic alkyne

Azide-Staudinger Ligation: Reaction between triphenyl phosphine and alkyl azide in the absence of catalyst to yield a triazole is called Azide-Staudinger ligation as shown in

Scheme 14. This reaction achieves high product yields and does not require any catalyst. In this reaction, an amide bond is formed between one nitrogen atom of the azide and triarylphosphine. This reaction produces no toxic by-products and proceeds readily in physiologically relevant aqueous conditions.



Scheme 14: Azide-Staudinger Ligation between alkyl azide and triarylphosphine

3.2 Materials and methods

3.2.1 Click chemistry optimization for *in-vitro* conditions

Dibenzocyclooctyne-Polyethylene Glycol4-Fluor 545 (DBCO-PEG4-FLUOR 545, Catalog number: 760773) was purchased from Sigma Aldrich. Next, 1 mM DBCO-PEG4-FLUOR 545 stock (Molecular weight: 936.06 g/mol) was prepared by dissolving 5 mg of DBCO-PEG4-FLUOR 545 in 5340 µl DI water by vortexing. Then, 100 mM stock of β-L-fucopyranosyl azide (Molecular weight: 189.17 g/mol) was prepared by dissolving 18.917 mg β-L-fucopyranosyl azide in 1 ml of DI water by gentle vortexing. 10 mM β-D-glucopyransoyl azide (Molecular weight: 205.17 g/mol) was prepared by dissolving 11 mg β-D- glucopyransoyl azide in 5 ml of DI water by gentle vortexing.

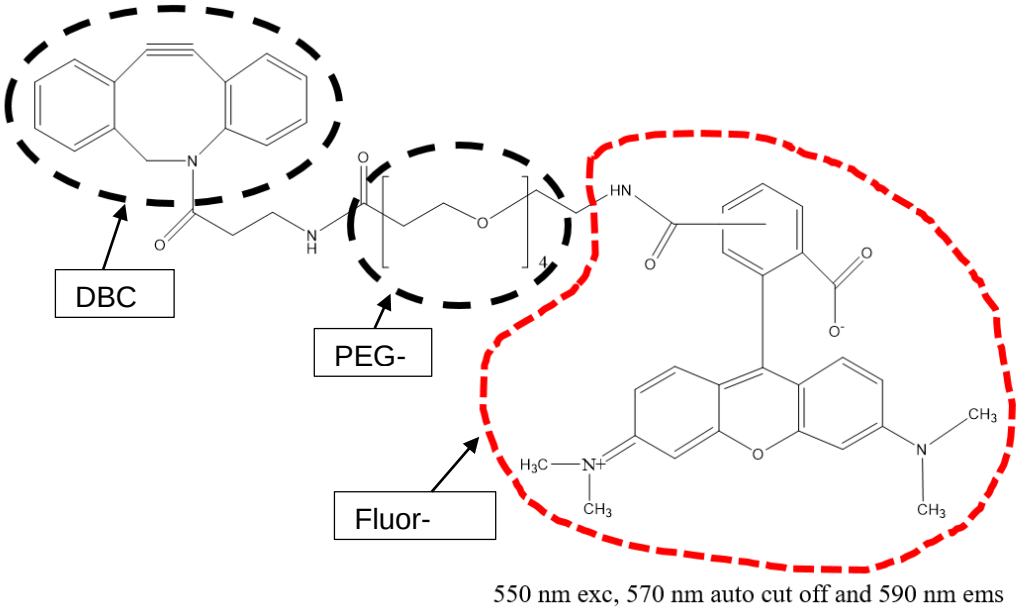


Figure 15: Molecular structure of DBCO-PEG4-Fluor 545

***In-vitro* reaction optimization for fluorescence analysis:** For copper-free click reaction, 200 µM DBCO-PEG4-FLUOR 545 and 400 µM of respective azides (sodium azide or β-D-glucopyransoyl azide or β-L-fucopyranosyl azide) were mixed in 1X PBS reaction buffer at pH 7.4. Next, 200 µM of DBCO-PEG4-FLUOR 545 with 1X PBS buffer pH 7.4 without azides was taken as the DBCO-PEG4-FLUOR 545-control. Also, 400 µM azides with 1X PBS buffer without DBCO-PEG4-FLUOR 545 were taken as the azide controls for the reaction. The reaction was incubated for 5 hours in 384 wells MatriCal plates (GE Healthcare, 28-9324-02) at the following temperatures for optimization: 37 °C, 25 °C, 10 °C at 400 rpm and fluorescence was recorded every 30 minutes at 550 nm excitation, 570 nm auto cut off and 590 nm emission using a spectrophotometer (SpectraMax M5e).

3.2.2 Click chemistry reaction mixture absorbance analysis

The progress of the click chemistry reaction was monitored by the disappearance of the characteristic UV at 309 nm to follow the formation of the triazole moiety [55]. Briefly,

200 µM of DBCO-PEG4-FLUOR 545/ DBCO-NHS was again mixed with 400 µM azides (sodium azide and β-D- glucopyransoyl azide) in 1X PBS buffer at pH 7.4. Next, 200 µM of DBCO-PEG4-FLUOR 545/ DBCO-NHS was taken with 1X PBS buffer pH=7.4 without azides as the DBCO-PEG4-FLUOR 545/ DBCO-NHS controls. Azides were taken with 1X PBS buffer pH=7.4 without DBCO-PEG4-FLUOR 545/ DBCO-NHS as azide controls. The reaction was incubated at 37°C for 3 hours at 400 rpm with regular monitoring of absorbances at 309 nm.

3.2.3 Fluorescence reduction test:

Figure 16 shows that Fluor-545 (Tetramethyl rhodamine) and Rhodamine B (Tetraethyl rhodamine) dyes are structurally similar with a minor structural difference with the methyl groups in Fluor-545 replaced by ethyl groups in Rhodamine-B. Our aim was to study the effect on the fluorescence of the dye alone and in the presence of exogenously added azides or triazole products. This experiment was conducted to explain why the dye fluorescence differentially reduces upon completion of the click reaction for organic versus inorganic azide. Fluor-545 moiety alone is not readily available from commercial sources, it was either tagged with esters, azides, or other functional groups that might interfere with the reaction and prevent us from testing our hypothesis. So, we studied the effect of free azide or triazole products on the fluorescence of Rhodamine-B instead.

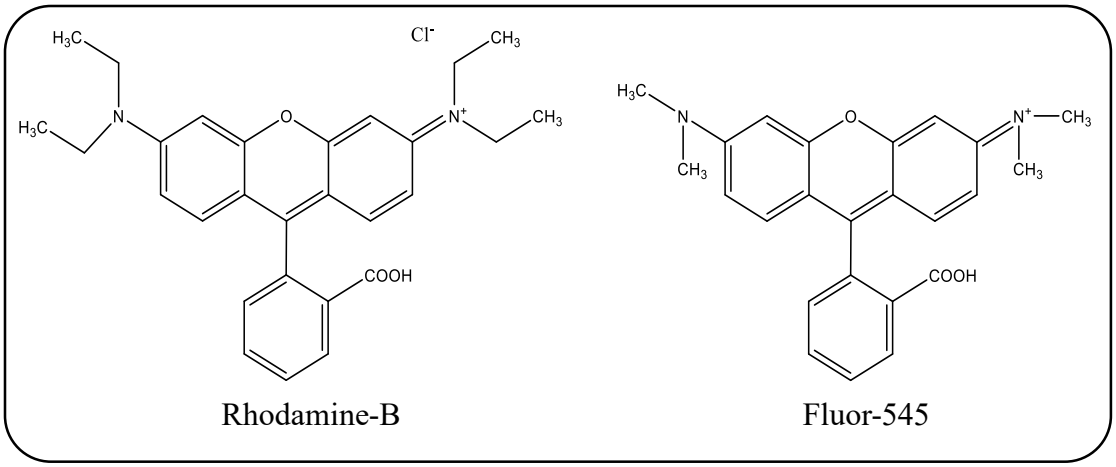


Figure 16: Molecular structures of Rhodamine-B and Fluor-545 group

Effect of free azides on the fluorophore group

To test the effect of exogenously added free azide on the fluorophore group, 200 µM Rhodamine-B was mixed with 400 µM of azides (sodium azide and β-D- glucopyransoyl azide) in 1X PBS buffer pH=7.4. Also, 200 µM Rhodamine-B in 1X PBS buffer pH=7.4 without azides was taken as the Rhodamine-B only control. Finally, 400 µM azides in 1X PBS buffer pH=7.4 without Rhodamine-B were taken as the azide only controls. The reaction was incubated at 37°C for 3 hours and the fluorescence spectra for the solution was recorded every 30 minutes at 550 nm excitation, 570 nm auto cutoff and 590 nm emission in a UV spectrophotometer SpectraMax M5e.

Effect of free triazole products on the fluorophore group

To test the effect of triazole moiety on the fluorophore group, 200 µM of DBCO-NHS was first mixed with 400 µM azides (sodium azide and β-D- glucopyransoyl azide) in 1X PBS buffer pH=7.4 to allow the click reaction to take place. Here, 200 µM of DBCO-NHS with 1X PBS buffer pH=7.4 without azides was taken as the DBCO-NHS control. Azides were taken with 1X PBS buffer pH=7.4 without DBCO-NHS as azide controls. Only 1X PBS

buffer pH=7.4 was taken as the blank for the reaction. The reaction was incubated at 37 °C for 3 hours at 400 rpm. The progress of the click reaction was monitored by the disappearance of the characteristic UV at 309 nm to confirm the reaction had gone to completion [55]. Finally, to study the effect of click reaction formed triazole products on the fluorophore group, 200 μM of Rhodamine-B was added to all the wells, mixed well, and further incubated at 37 °C for 3 hours. The fluorescence of the reaction mixture was recorded every 30 minutes at 550 nm excitation, 570 nm auto cutoff and 590 nm emission.

3.2.4 Click chemistry reactions with mixed organic and inorganic azides

Here, we wanted to study the linear response range of the drop in fluorescence of the Fluor545 moiety upon completion of the click reaction using inorganic, organic, and mixed inorganic-organic azides. Briefly, 200 μM DBCO-PEG4-FLUOR 545 was reacted with 100% sodium azide (400 μM), 100% β-D- glucopyransoyl azide (400 μM), and with a mixture combination of 50% sodium azide (200 μM) and 50% β-D-glucopyransoyl azide (200 μM) in 1X PBS reaction buffer pH=7.4. Here, 200 μM of DBCO-PEG4-FLUOR 545 with 1X PBS buffer pH=7.4 without azides was taken as the DBCO-PEG4-FLUOR 545-control. The reaction was incubated at 25°C for 3 hours in 384 wells MatriCal plates and the plate was read for fluorescence at 550 nm excitation, 570 nm auto cutoff and 590 nm emission.

3.2.5 Confocal Microscopy

Starter culture was inoculated with *E. coli* BL-21 (DE3) glycerol stock for pEC_Tm0306_WT plasmid in 10 ml LB media with 50 μg/ml kanamycin. Here, 5 ml LB media with 50 μg/ml kanamycin alone was taken as a control. The starter culture and the control were incubated at 37 C for 16 hours. Next, 2.25 ml of the starter culture was

transferred to 45 ml minimal media with 45 µl kanamycin and 5 ml minimal media with 5 µl kanamycin was taken in a separate tube as control and incubated at 37 C for 16 hours until OD600 of Tm0306_WT reached about 2. The cell culture was centrifuged at 8,000 rpm for 15 minutes and the supernatant was discarded. The culture was washed thrice with equal amount of 1X PBS buffer pH 7.4 and centrifuged at the same conditions as described above. The washed culture was now re-suspended in same amount of 1X PBS buffer pH 7.4. OD600 was measured again and it was found to be 2 again which remains in consistency with the amount of cells in the culture before the washing step. Now, the cell culture was aliquoted into various sterile micro-centrifuge tubes and divided into two categories as described below:

Type-1: Unwashed samples preparation

Here, 2 tubes (labeled as C1 and C3) were prepared as control for the experiment with 200 µl cells and 200 µl DI water. Another 2 tubes (labeled as C5 and S1) were prepared with 200 µl cells and 66 µl of 0.5 mM of DBCO-PEG4-FLUOR 545. All tubes (C1, C3, C5 and S1) were incubated at 37 C for 30 minutes. The samples were centrifuged at 10,000 rpm for 3 minutes and the supernatants were discarded. Now, 50 µl of freshly prepared 4% paraformaldehyde was added to all the samples, mixed well, and incubated at 37 C for 10 minutes. The samples were centrifuged at the same conditions as described above and the supernatants were discarded. The cell pellets obtained were washed twice with 1X PBS buffer pH=7.4 followed by re-suspending in 266 µl of 1X PBS buffer pH=7.4, mixing well, and centrifuging at 10,000 rpm for 3 minutes and finally discarding the supernatant. Next, 50 µl of 0.1 µg/ml Hoechst 33342 was added to S1, mixed well and incubated at 37 C for 10 minutes. S1 was centrifuged at 10,000 rpm for 3 minutes and the supernatant was

discarded. S1 was washed twice with 1X PBS buffer pH=7.4 and the supernatants were discarded. C1, C3, C5 and S1 were re-suspended in 100 µl of 1X PBS buffer pH= 7.4 and mixed well. Next, 2 µl of the samples were mixed with 50 µl mounting media (Prolong diamond antifade mounting agent, Catalog number: P36965, Thermo Fisher Scientific) in PCR tubes and centrifuged to remove bubbles. Finally, 10 µl of the samples were placed on a glass slide covered with transparent glass cover slip and incubated at 25 C for 24 hours in dark and visualized under a confocal microscope.

Type-2: Washed samples preparation

Here, 2 tubes (labeled as C2 and C4) were prepared as control for the experiment with 200 µl cells and 200 µl DI water. Another 2 tubes (labeled as C6 and S2) were prepared with 200 µl cells and 66 µl of 0.5 mM of DBCO-PEG4-FLUOR 545. All tubes (C2, C4, C6 and S2) were incubated at 37 C for 30 minutes. The samples were then centrifuged at 10,000 rpm for 3 minutes and the supernatants were discarded. The samples were washed thrice in an equal amount of 1X PBS buffer pH= 7.4 and the supernatants were discarded. Now, 50 µl of freshly prepared 4% paraformaldehyde was added to all the samples, mixed well, and incubated at 37 C for 10 minutes. The samples were centrifuged at the same conditions as described above and the supernatants were discarded. The cell pellets obtained were washed twice with 1X PBS buffer pH=7.4 by re-suspending in 266 µl of 1X PBS buffer pH=7.4, mixing well, and centrifuging at 10,000 rpm for 3 minutes and discarding the supernatant. Next, 50 µl of 0.1 µg/ml Hoechst 33342 was added to S1, mixed well and incubated at 37 C for 10 minutes. S1 was then centrifuged at 10,000 rpm for 3 minutes and the supernatant was discarded. S1 was washed twice with 1X PBS buffer pH=7.4 and the supernatants were discarded. C1, C3, C5 and S1 were re-suspended in 100 µl of 1X PBS

buffer pH= 7.4 and mixed well. Next, 2 µl of the samples were mixed with 50 µl mounting media (Prolong diamond antifade mounting agent, Catalog number: P36965, Thermo Fisher Scientific) in PCR tubes and centrifuged to remove bubbles. Finally, 10 µl of the samples were placed on a glass slide covered with transparent glass cover slip and incubated at 25 C for 24 hours in dark and visualized under a confocal microscope.

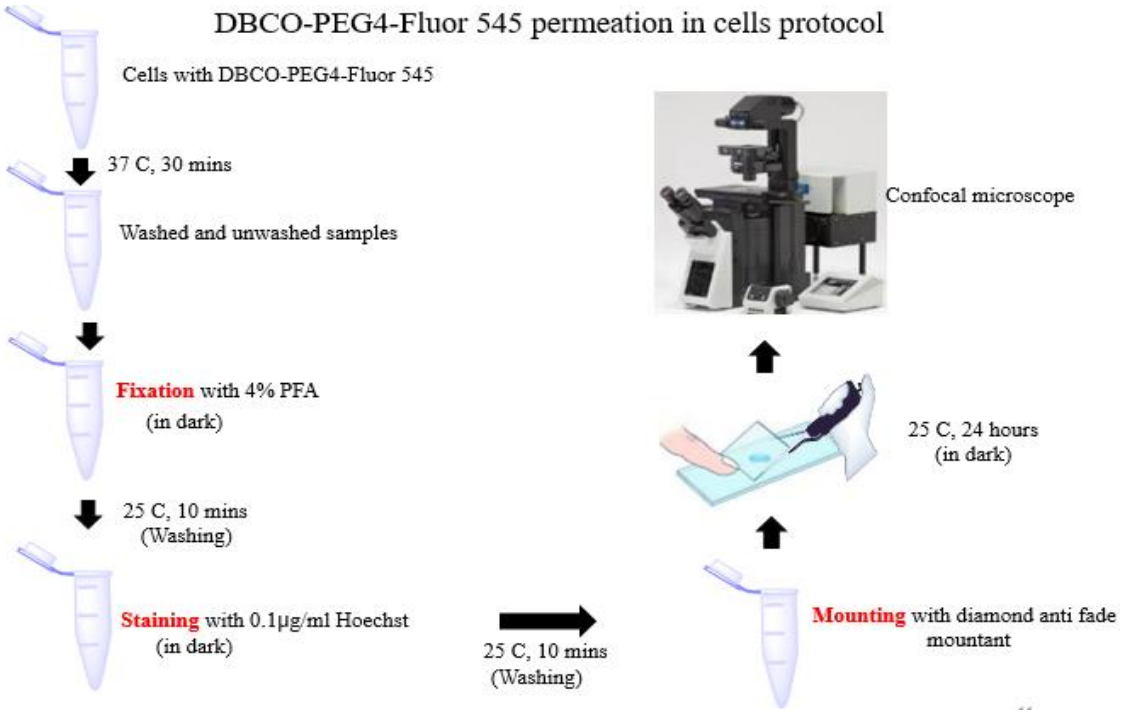


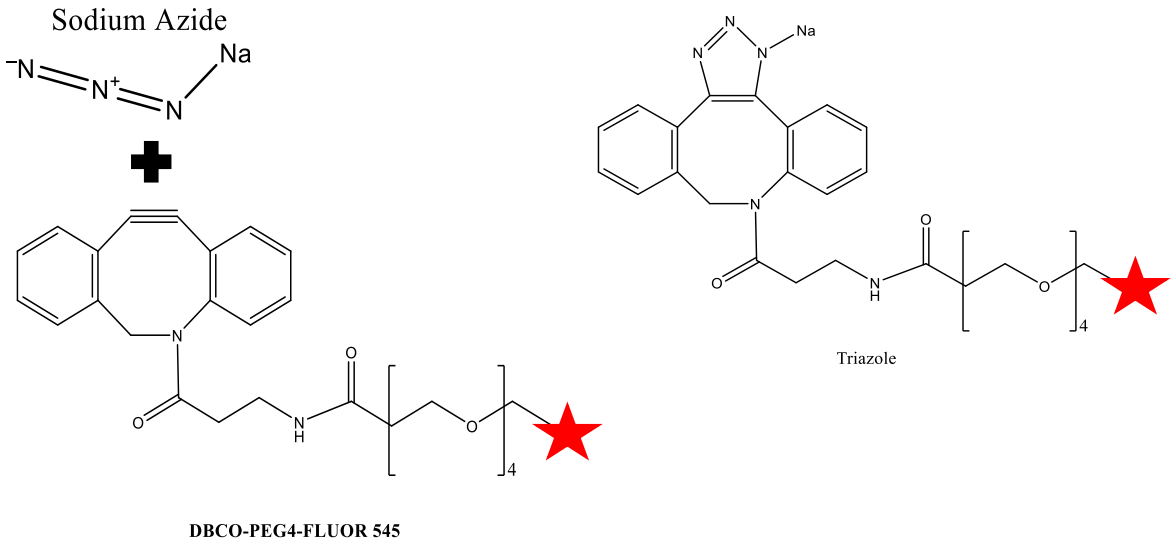
Figure 17: Protocol for sample preparation for confocal microscopy

3.3 Results and Discussion

3.3.1 Absorbance and fluorescence analysis of click chemistry reactions

Absorbance analysis: DBCO-PEG4-FLUOR 545 was reacted with sodium azide and β-D-glucopyransoyl azide separately in 1X PBS buffer pH=7.4 and the reaction mixture UV absorbance spectra were recorded at 309 nm for various time points. It can be observed from Figure 18 that the characteristic wavelength for alkyne absorbance at 309 nm decreases gradually as the reaction is proceeds towards completion. In other words, the

absorbance at 309 nm decreases with the corresponding formation of the triazole group. Popik and co-workers have also reported a similar pattern for disappearance of UV absorbance at 309 nm upon reaction of 0.5 mM ADIBO (also known as DBCO) with 2.5 mM sodium azide in PBS buffer with 5% methanol [56]. This work clearly showed that triazole formation is correlated with a corresponding decrease in the 309 nm UV absorbance. It was also hypothesized that the azide moiety of sodium azide reacts with the alkyne moiety of DBCO-PEG4-Fluor 545 to form a triazole as shown in Scheme 15.



Scheme 15: Hypothesized mechanism of reaction between DBCO-PEG4-Fluor545 with sodium azide (Note: The red star at the end of the structure represents Fluor-545 group)

Reaction kinetics data interpretation was performed by plotting various data points of absorbance vs. time fitted to an exponential decay function (Exp Dec 1) using Origin software with R-squared values around 0.98-0.99 (Figure 18). The Exp Dec 1 function was of the form: $y = A1 * \exp\left(-\frac{x}{t1}\right) + y0$, where y stands for absorbance at 309 nm, x stands for time in minutes, A1 is the slope, y0 is a constant and 1/t1 is the rate constant. This

equation assumes a pseudo-first order reaction kinetics with the rate constant 1/t1. The rate constants for the formation of the triazole moiety during click reaction based on the decrease in UV 309 nm absorbance values were found to be 0.0412±0.006 s^{-1} for sodium azide and 0.043±0.003 s^{-1} for reaction with glucosyl azide. This suggests that the rate of the click reaction is similar irrespective of the type of azide (organic vs. inorganic) taken into consideration.

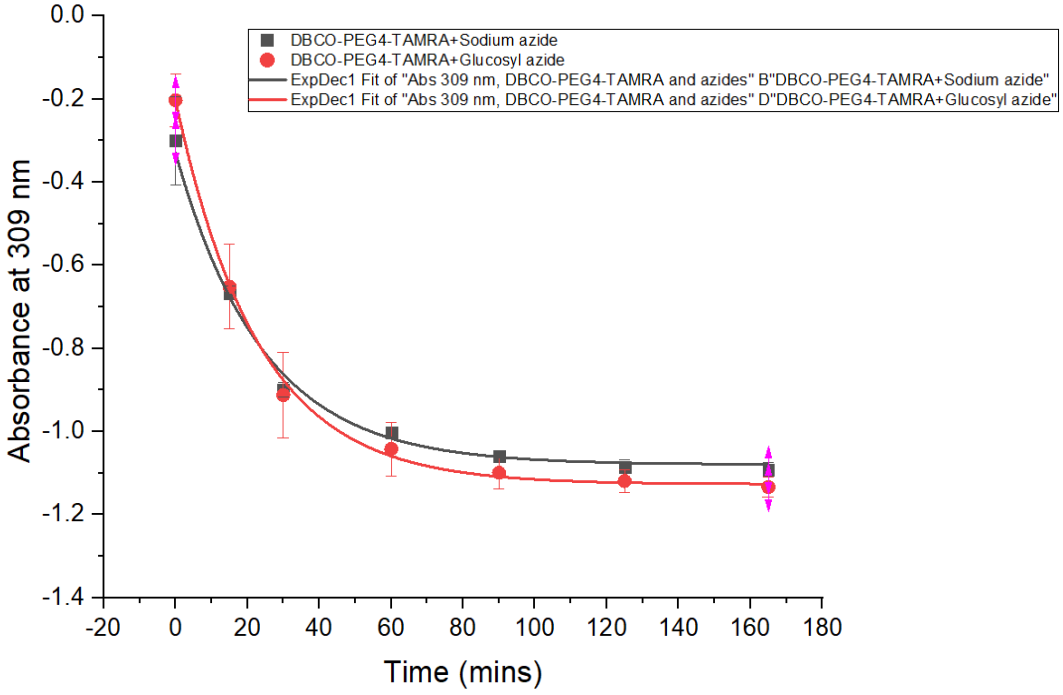


Figure 18: UV spectra showing the loss of absorbance at 309 nm for reaction of DBCO-PEG4-Fluor 545 with sodium azide and β-D-glucopyranosyl azide.

Fluorescence analysis: The fluorescence of the exact same reaction (as mentioned above) was also measured at various time points at 550 nm excitation, 570 nm auto cut off, and 590 nm emission (Section 3.3.1.1). It was observed that the fluorescence of the reaction mixture was decreasing upon formation of the click reaction product (i.e., triazole) (Figure

19). The data points for reaction with sodium azide and glucosyl azide were plotted by subtracting only the DBCO-PEG4-Fluor 545 and azide only controls data from the mixture data. Figure 19 depicts a significant difference between the relative drop in fluorescence during completion of the click reaction with sodium azide and glucosyl azide. These results are clearly in contrast to the previous UV absorbance results that allow a more direct quantitation of the disappearance of the alkyne group.

However, these results also suggest that azide release during the glycosynthase reaction (as free azide) could be detected by monitoring the reaction using DBCO-PEG4-Fluor 545 (click chemistry reaction) and by taking relative fluorescence reduction into consideration as compared to the controls consisting of only DBCO-PEG4-Fluor 545 or unreacted azido sugars. Hence, this could be a crucial parameter for FACS screening of cells expressing active glycosynthases. Here we also fitted the various data points of fluorescence vs. time for click reactions conducted at 37°C fitted to an exponential decay function (Exp Dec 1) using Origin software to get R-squared values around 0.99 for both sodium azide and glucosyl azide. The apparent rate constants for the drop in fluorescence during the click reactions were 0.0340 ± 0.001 and 0.0149 ± 0.0007 for sodium azide and glucosyl azide, respectively. Drop in fluorescence was also monitored for click reactions conducted at two additional temperatures (i.e., 25°C and 10°C) as discussed in the next section.

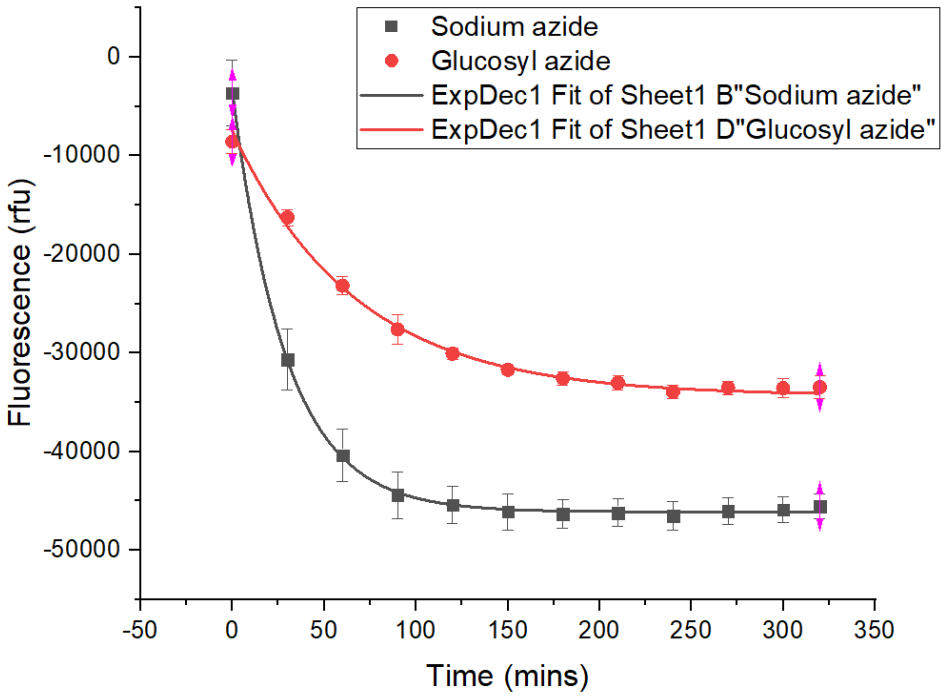


Figure 19: Fluorescence decay curve for the reaction between DBCO-PEG4-FLUOR 545 (200 µM) with sodium azide and β-D-glucopyranosyl azide (400 µM) in 1X PBS buffer pH=7.4 at 37°C measured at 550 nm excitation, 570 nm auto cut off and 590 nm emission

Model	ExpDec1	
Equation	y = A1*exp(-x/t1) + y0	
Plot	Sodium azide	Glucosyl azide
y0	-46107.87052 ± 109.92369	-34319.71384 ± 290.83162
A1	42531.40524 ± 689.81772	26808.75534 ± 624.63194
t1	29.39646 ± 0.95476	67.01805 ± 3.42059
Reduced Chi-Sqr	0.04362	0.39629
R-Square (COD)	0.99785	0.99545
Adj. R-Square	0.99737	0.99444

3.3.2 Click chemistry *in-vitro* temperature optimization results

DBCO-PEG4-Fluor 545 was reacted with sodium azide and β-D-glucopyransoyl azide in the reaction buffer 1X PBS, pH=7.4. This experiment was conducted to observe the change in fluorescence once the click reaction goes to completion at two additional temperatures. Experiments were conducted at two different temperatures: 25°C and 10°C to optimize the temperature for maximum conversion of reactant alkyne to triazole. An exponential decay curve was fitted through the fluorescence-time dataset for click reactions performed at 25°C and 10°C (Figure 20 and Figure 21). R-squared values of approximately 0.98 and 0.99 were obtained for sodium azide and glycosyl azide respectively at both temperatures. The reduction in fluorescence rates at 25°C were 0.013$\pm$ 0.001 s^{-1} and 0.0105$\pm$ 0.002 s^{-1} for sodium azide and glucosyl azide, respectively. At 10°C, the reduction in fluorescence rates were 0.0107$\pm$ 0.001 s^{-1} and 0.0067$\pm$ 0.0005 s^{-1} for sodium azide and glucosyl azide, respectively. This reduction in fluorescence rates at 25°C and 10°C were about 3 times lower than click chemistry reaction rate at 37°C. Hence, 37°C was selected as the optimal temperature for click chemistry reaction. Also, since 37°C is an optimal condition for *E. coli* cell growth, this temperature was used for all future click chemistry experiments.

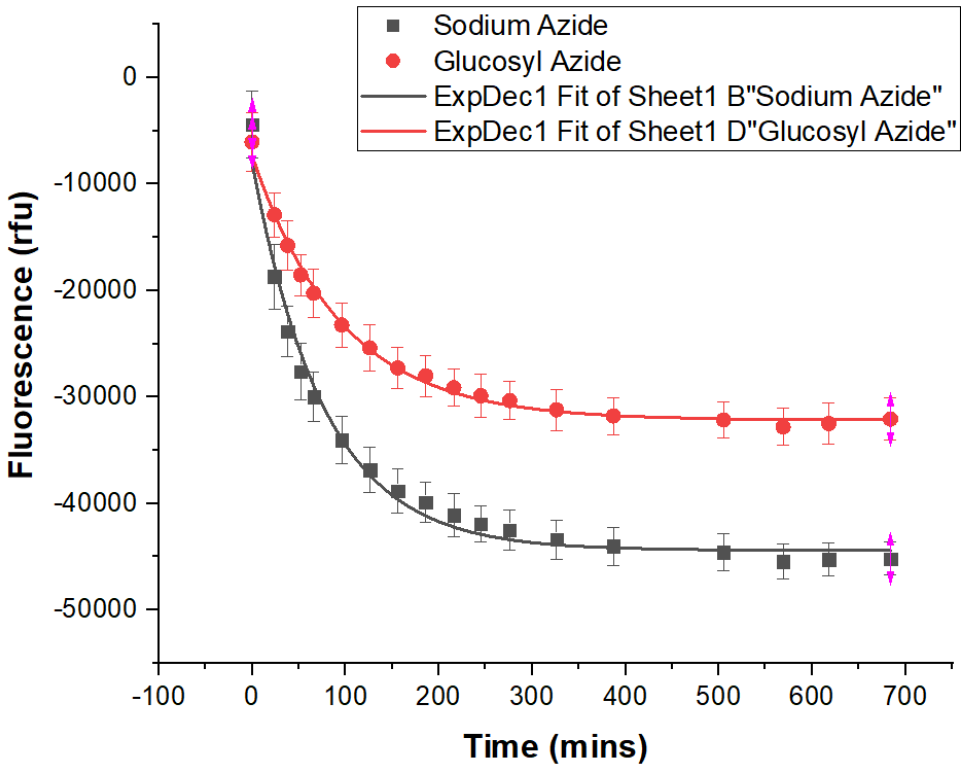


Figure 20: Fluorescence decay curve for the reaction between DBCO-PEG4-FLUOR 545 (200 µM) with sodium azide and β-D-glucopyranosyl azide (400 µM) in 1X PBS buffer pH=7.4 at 25°C measured at 550 nm excitation, 570 nm auto cut off and 590 nm emission

Model	ExpDec1	
Equation	y = A1*exp(-x/t1) + y0	
Plot	Sodium Azide	Glucosyl Azide
y0	-44400.00151 ± 384.89	-32165.41738 ± 196.74
A1	36419.7513 ± 1409.97	24817.06203 ± 473.32
t1	76.93564 ± 5.21935	95.54141 ± 3.83616
Reduced Chi-S	0.34093	0.06236
R-Square(CO	0.9843	0.9956
Adj. R-Square	0.98221	0.99501

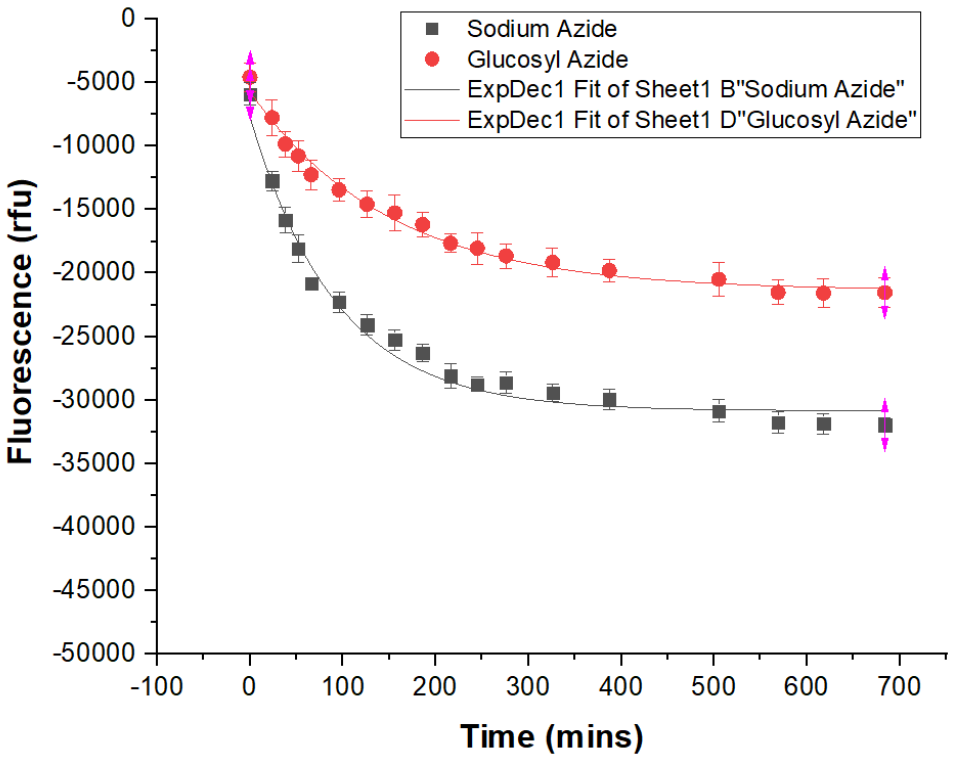


Figure 21: Fluorescence decay curve for the reaction between DBCO-PEG4-FLUOR 545 (200 µM) with sodium azide and β-D-glucopyranosyl azide (400 µM) in 1X PBS buffer pH=7.4 at 10°C measured at 550 nm excitation, 570 nm auto cut off and 590 nm emission

Model	ExpDec1	
Equation	y = A1*exp(-x/t1) + y0	
Plot	Sodium Azide	Glucosyl Azide
y0	-30854.93575 ± 453.588	-21373.81094 ± 329.005
A1	23172.11677 ± 1001.439	15670.19705 ± 440.1439
t1	93.07353 ± 8.38038	150.62191 ± 10.79225
Reduced Chi-Sq	2.18522	0.27379
R-Square(COD)	0.97715	0.9883
Adj. R-Square	0.9741	0.98675

3.3.3 Comparison of rate of reaction of glucosyl azide and fucosyl azide:

Click chemistry reaction was performed at 37°C using DBCO-PEG4-Fluor 545 with glucosyl azide and fucosyl azide for 350 minutes. It was suspected that fucosyl azide would

show a similar behavior to glucosyl azide because they are both organic azides and are structurally similar (Figure 22 and Figure 23).

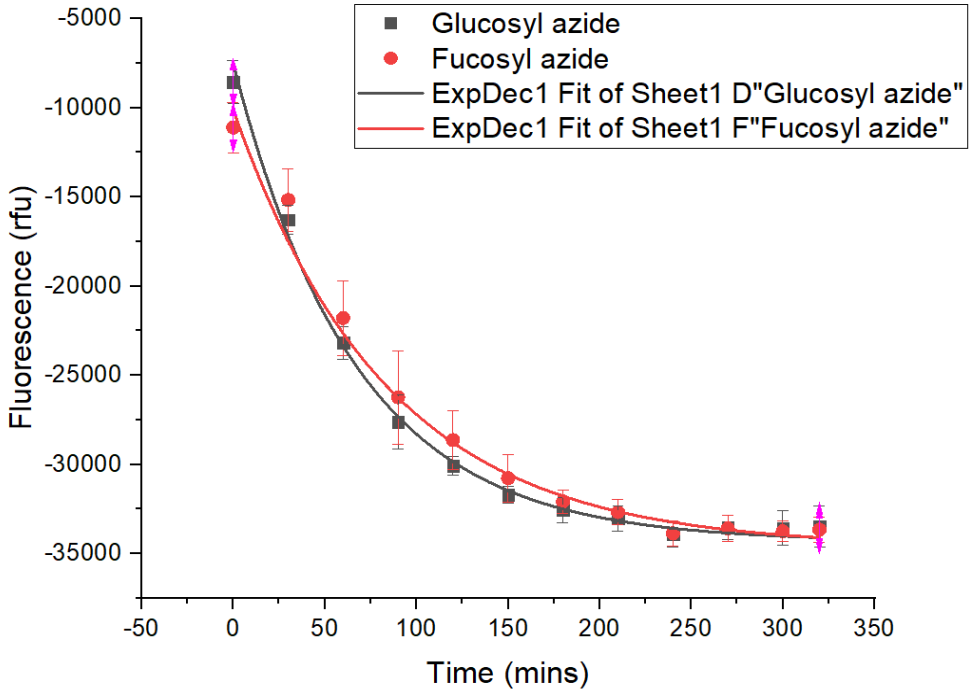
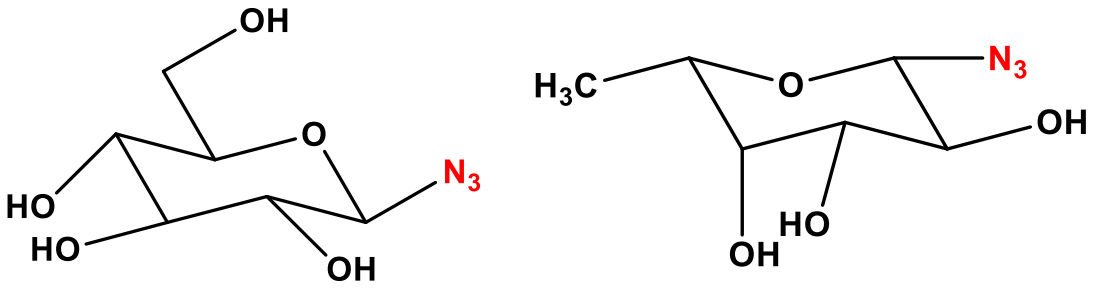


Figure 22: Fluorescence decay curve for the reaction between DBCO-PEG4-FLUOR 545 (200 µM) with β-D-glucopyranosyl azide (400 µM) and β-L-fucopyranosyl azide in 1X PBS buffer pH=7.4 at 37°C measured at 550 nm excitation, 570 nm auto cut off and 590 nm emission

Model	ExpDec1	
Equation	y = A1*exp(-x/t1) + y0	
Plot	Glucosyl azide	Fucosyl azide
y0	-34319.71384 ± 290.83162	-34637.96295 ± 397.02075
A1	26808.75534 ± 624.63194	24523.8233 ± 851.43823
t1	67.01805 ± 3.42059	83.88465 ± 7.22576
Reduced Chi-Sqr	0.39629	0.44343
R-Square (COD)	0.99545	0.98929
Adj. R-Square	0.99444	0.98691

We found that click reaction with glucosyl azide versus fucosyl azide showed a very similar drop in fluorescence that was expected due to similarity in the molecular structures of the formed triazole product. This triazole moiety is expected to be clearly distinct from the one

formed from inorganic azides and hence that could explain why the sugar azides behave similarly based on the change in fluorescence associated with the Fluor545 group.



β-D- glucopyranosyl azide

β-L- fucopyranosyl azide

Figure 23: Molecular structures of β-D- glucopyranosyl azide and β-L- fucopyranosyl azide

It

Click reaction between DBCO-PEG4-FLUOR 545 and various azides results in decay of fluorescence over time. Here are a few hypotheses that could explain this phenomenon:

1. The free azide in the reaction mixture interacts with the fluorophore group and causes a reduction in fluorescence. Though this is unlikely to be the case since the drop in fluorescence is seen even when the free azide is fully consumed at the end of the click reaction.
2. The triazole moiety formed during the click reaction interacts with the fluorophore group to quench its fluorescence as compared to when no triazoles are present along with the fluorophore group.
3. Lastly, it is possible that the PEG-4 linker keeps the DBCO moiety (after reaction with azide to form triazoles) and the FLUOR-545 moiety in a particular conformation where they interact more readily in some specific manner and hence

cause reduction in fluorescence. This is an extension of the second hypothesis listed above. We tested all these hypotheses as outlined below.

3.3.4 Effect of free azide groups on the fluorophore group

From Section 3.3.2, we observed that the fluorescence of the FLUOR-545 moiety decreases upon formation of a triazole product. There are no previous studies that have tested fluorescence of DBCO-PEG4-Fluor 545 with azides. We hypothesized that the fluorescence reduction was maybe due to interaction of azides with the fluorophore.

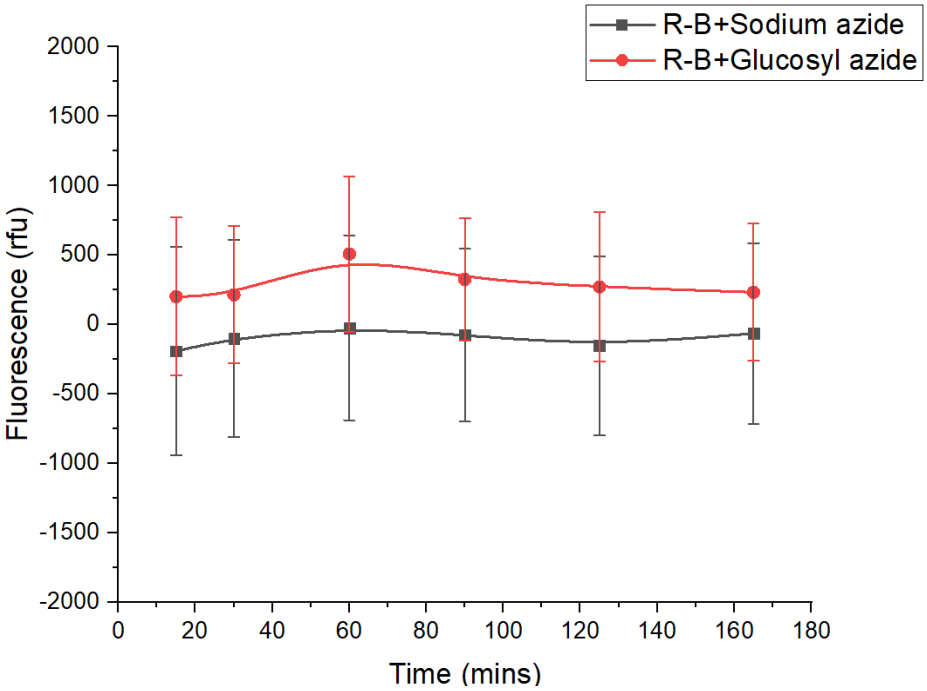


Figure 24: Fluorescence curve for the reaction between Rhodamine-B (200 µM) with sodium azide and β-D-glucopyranosyl azide (400 µM) in 1X PBS buffer pH=7.4 at 37°C measured at 550 nm

So, this experiment was conducted to check if azides affect fluorophore in some way that might cause reduction in fluorescence of the final reaction mixture. Since, Rhodamine-B and Fluor-545 dyes are structurally similar (Figure 16), the experiment was conducted using Rhodamine-B as the fluorophore. Respective azides were mixed with Rhodamine-B

and the mixture fluorescence was recorded at various time points at 550 nm excitation, 570 nm auto cutoff and 590 nm emission using UV spectrophotometer Spectra Max M5e. From Figure 24, we can clearly observe that there is no significant change in fluorescence in the presence of either organic or inorganic azides. In other words, we can say that excess azide present in click reaction likely does not impact the fluorophore.

3.3.5 Effect of triazole moiety formed on the fluorophore

This experiment was performed to check if the triazole moiety formed in the click chemistry reaction affects the fluorescence (reduces the fluorescence observed from free dye added to solution) of the final mixture. Initially, DBCO-NHS was mixed with respective azides (sodium azide and β-D- glucopyranosyl azide). DBCO-NHS controls without azides and azide controls without DBCO-NHS were also taken in the Matrical-plate and incubated at 37°C for 200 minutes. The triazole formation was first confirmed during the click reaction by monitoring the change in absorbance at 309 nm for various time points as shown in Figure 25 (We can observe that the rate in absorbance decrease is slightly different for DBCO-PEG4-Fluor 545 (Figure 18) and DBCO-NHS (Figure 25). This might be due to the presence of the PEG linker in DBCO-PEG4-Fluor 545 which facilitates the interactions of DBCO and Fluor-545 group over the course of the reaction. These interactions might lead to similar reaction rates for both inorganic and organic click chemistry reactions). Once the triazole formation was confirmed to be completed after 200 mins total reaction time, Rhodamine-B dye was added to all the wells including the controls and incubated at 37°C while constantly measuring fluorescence at 550 nm excitation, 570 nm auto cutoff and 590 nm emission using a spectrophotometer SpectraMax M5e for various incubation times ranging from 0-120 minutes from the point of addition of the dye.

The Rhodamine-B controls in PBS buffer were subtracted from the readings obtained for all wells and the graph was plotted for fluorescence vs. time as shown in Figure 26. From

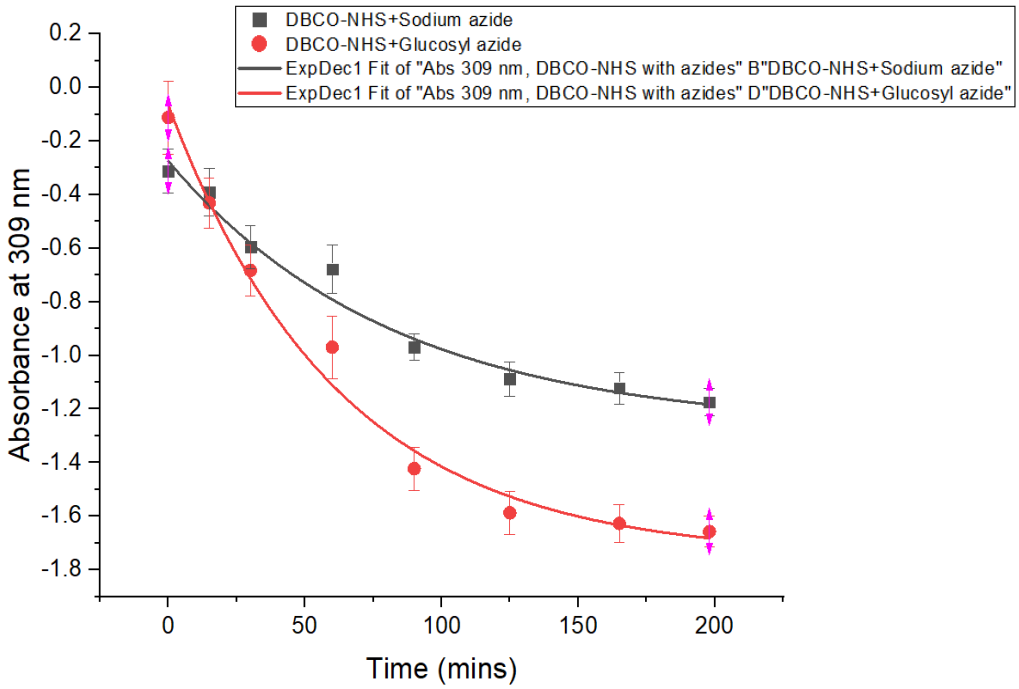


Figure 25: UV spectra showing the loss of absorbance at 309 nm for reaction of DBCO-NHS with sodium azide and β-D-glucopyranosyl azide.

Figure 26, we observed that the fluorescence of the solutions containing the triazole based click reaction products (for both sodium azide and glucosyl azide) was not significantly different when compared to sample containing only DBCO-NHS. Also, there was no significant difference seen between sodium and glucosyl azide as confirmed by student's t-test analysis for p= 0.1, 0.05, 0.01 (data not shown here). It can be inferred from these results that the PEG linker between the triazole product and fluorophore is very important to cause a decrease in the fluorescence.

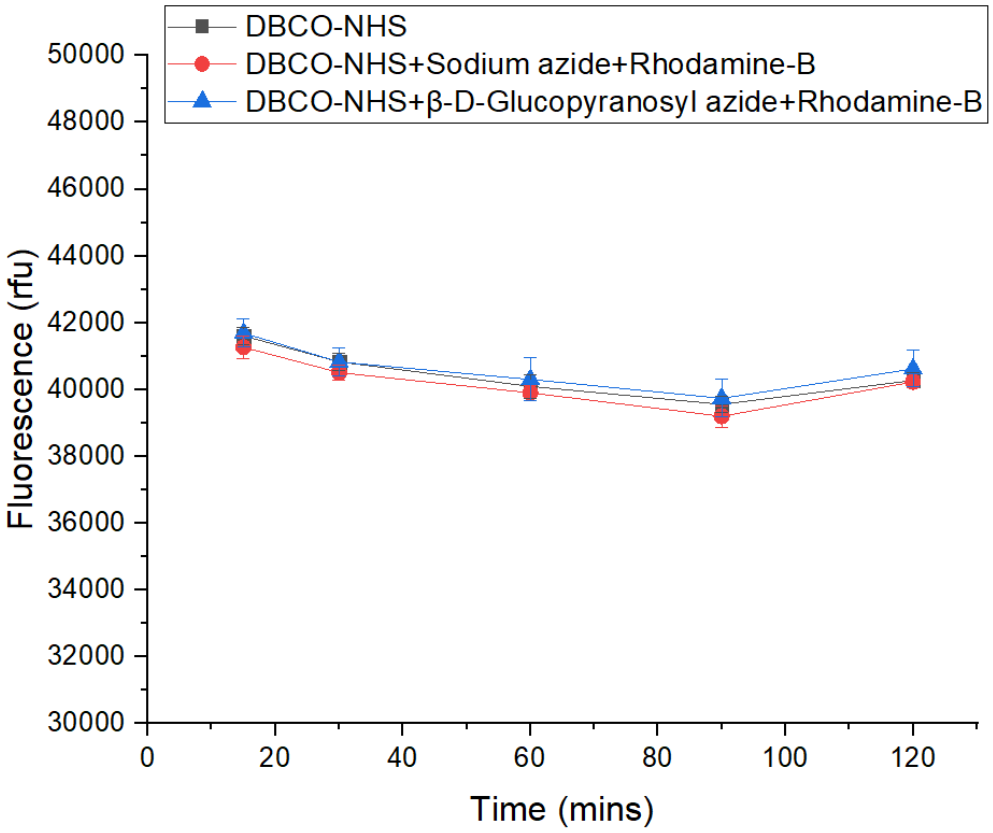


Figure 26: DBCO-NHS reaction with azides at 37°C results in formation of triazole products that reduces Rhodamine-B fluorescence

It is likely that the presence of a PEG linker localizes the fluorophore and the triazole product in close proximity influencing some unique intermolecular interactions that are not seen for intermolecular interactions of the free fluorophore with a free triazole product. For achieving a maximum Förster resonance energy transfer (FRET) signal it is generally observed that the distance between appropriately structured donor and acceptor fluorophores cannot exceed 10 nm [57]. The difference in triazole structure formed for inorganic azide versus glycosyl-azide could potentially give a distinct quenching of the emitted fluorescence signal from the Rhodamine-B dye if the triazole is within a certain

distance and orientation of the fluorophore. However, more work is needed to clearly understand this phenomenon.

3.3.6 Click chemistry reaction with mixed azides

This experiment was performed to evaluate the effects of click chemistry reaction products on the dye molecule fluorescence at 550 nm excitation, 570 nm auto cutoff and 590 nm emission when a equimolar combination of 50% sodium azide and 50% β-D-glucopyransoyl azide was reacted with DBCO-PEG4-FLUOR 545. From Section 3.3.2, we found that fluorescence of the click reaction of DBCO-PEG4-FLUOR 545 with sodium azide was 30% lower than with β-D-glucopyransoyl azide.

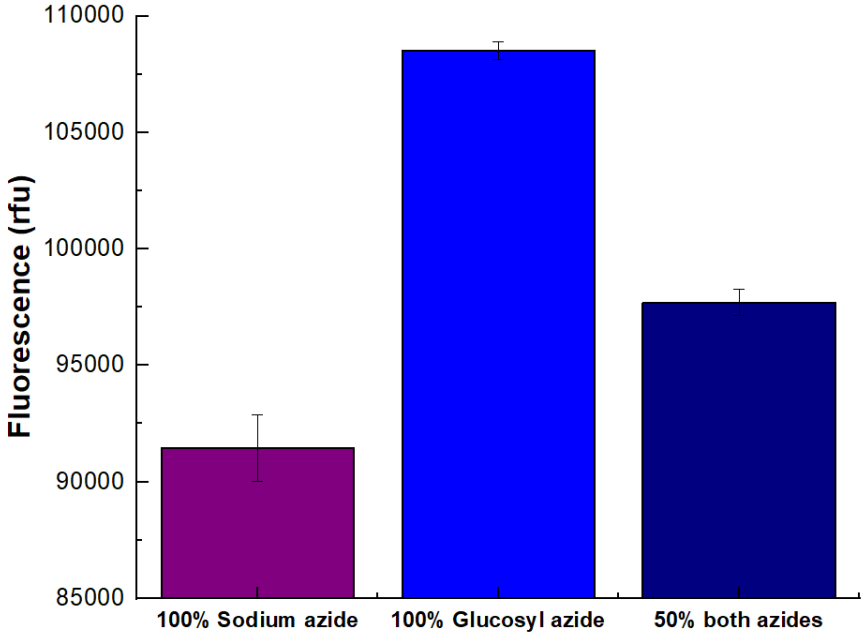


Figure 27: Click chemistry results with DBCO-PEG4-Fluor 545 and a combination of azides.

It was expected that the fluorescence of click chemistry reaction with a combination of both azides would be lying somewhere between results observed for sodium azide and β-D-glucopyranosyl azide alone. This was indeed observed and can be seen in Figure 27.

This clearly shows that if we have mixed concentrations of the organic and inorganic azide formed by glycosynthases of varying degrees of catalytic efficiency, we should still be able to differentiate amongst the products formed and substrates left using this click chemistry method.

3.3.7 Confocal Microscopy

Confocal microscopy was performed to check if the fluorescent alkyne used for click chemistry (DBCO-PEG4-FLUOR 545) could readily permeate through the cell wall of *E. coli*. Confocal microscopy samples were prepared with and without washing of excess DBCO-PEG4-FLUOR 545. Cell fixation was performed with 4% paraformaldehyde (polyoxymethylene).

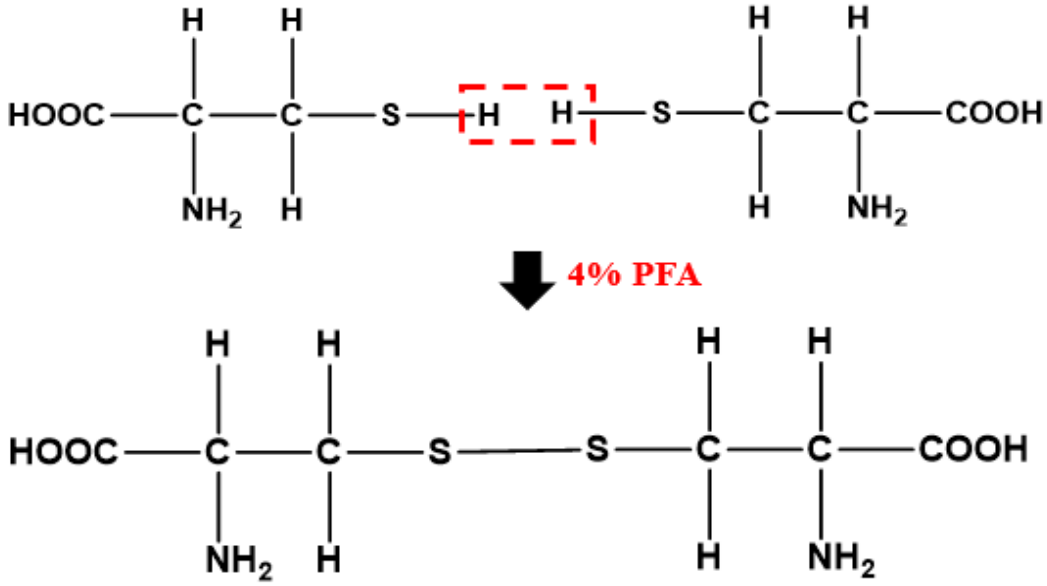


Figure 28: Cross-linking of molecules in presence of paraformaldehyde.

Fixation causes cross-linking between molecules (Figure 28) and is performed to make the cells permeable to allow antibodies to access intracellular structures, protect sample against microbial contamination, to inactivate the proteolytic enzymes that could degrade the

sample, and to strengthen the sample to withstand further processing and staining. The samples were then washed twice with 1X PBS buffer, pH=7.4 to remove the unreacted fixative and any byproducts of the fixative reaction itself. Furthermore, staining was performed using 0.1 µg/ml Hoechst 33342 (2'-[4-ethoxyphenyl]-5-[4-methyl-1-piperazinyl]-2,5'-bi-1H-benzimidazole trihydrochloride trihydrate). Hoechst 33342 is a cell-permeable DNA stain that is excited by UV light and emits blue fluorescence at 460 to 490 nm. Hoechst 33342 binds preferably to the A-T regions that greatly increase the fluorescence of the stained cells.

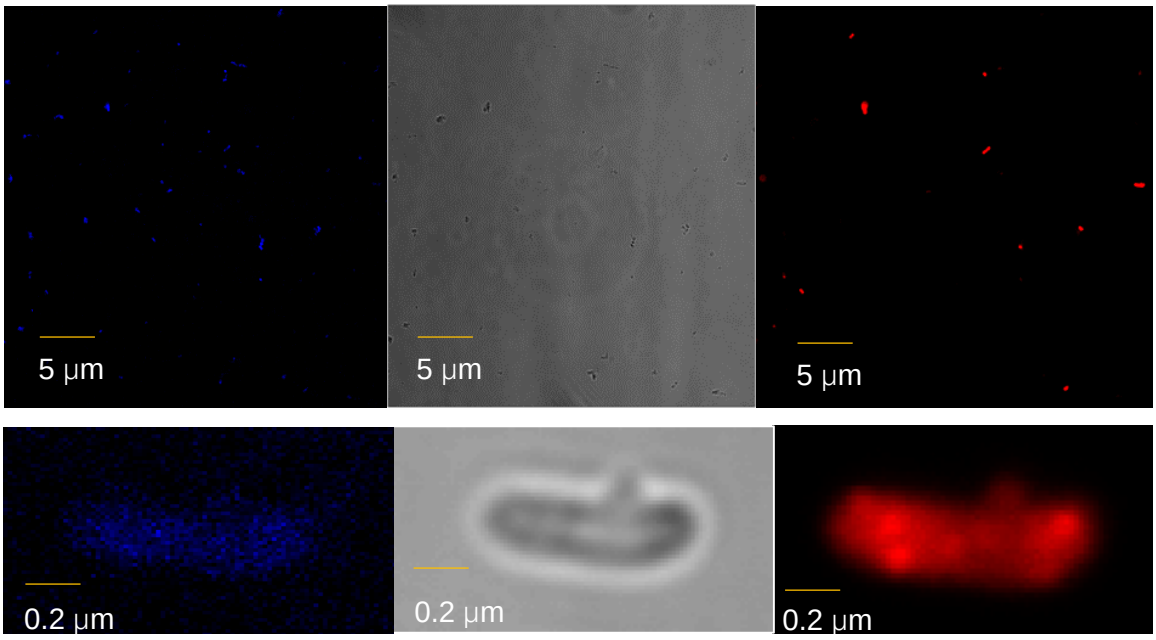


Figure 29: DBCO-PEG4-FLUOR 545 seems to be able to permeate into *E. coli* cells as seen with confocal microscopy showing co-localization of the dye (red) with the host DNA (blue) inside the bacterial cells (greyscale).

Also, addition of Hoechst 33342 makes the cells more permeable, is less toxic than DAPI (4′,6-diamidino-2-phenylindole) and ensures the viability of cells. The samples were then

washed twice with 1X PBS buffer pH=7.4 to remove the unreacted Hoechst 33342 from the samples. Mounting of the samples was performed using gold antifade diamond mountant and the samples were incubated at 25°C for 24 hours and visualized using a confocal microscope under bright field and wide-field/confocal fluorescence imaging modes. Mounting media is a mixture of 9 parts of glycerol and 1 part of PBS with a pH=8.5-9. Mounting media is added to avoid the photo bleaching of the sample. Since, mounting media is an anti-oxidant, it reduces the amount of oxygen available for photo-oxidation reactions, thus reducing photo bleaching across entire visible and UV spectrums. Mounting media is also crucial to hold the cells in place intact while imaging, to preserve sample for long-term storage and to prevent the sample from drying out. Figure 29 provides representative z-stack images of unwashed cells (excess DBCO-PEG4- FLUOR 545 not removed by washing). From Figure 29, it suggests that DBCO-PEG4-FLUOR 545 is able to permeate inside the cells.

Chapter-4: Random mutagenesis and screening of mutant libraries using FACS

4.1 Introduction

Directed evolution based approaches have tremendously advanced the field of protein engineering and our understanding of enzyme structure-function relationships [25]. In this particular approach, there are two iterative steps that include: (1) random mutagenesis of wild-type DNA to generate a diverse library of mutants, and (2) screening the library to identify the positive mutants that give desired function. There are various known techniques available in the literature for random mutagenesis among which error-prone PCR is widely used. Error-prone PCR (epPCR) is performed by taking advantage of low-fidelity nature of Taq DNA polymerase in the presence of excess divalent cations [58]. Taq DNA polymerase lacks proof reading mechanism unlike the T7 polymerases. In the presence of divalent metal cations, the rate of the reaction is increased which hence also increases the probability of causing errors during nucleotide insertion during DNA amplification. In this chapter, we will randomly mutagenize Tm0306_D224G gene using Taq polymerase in the presence of manganese chloride and magnesium chloride to create a large library of variants to screen and identify a more efficient fucosynthase. Furthermore, these variants will be run through a flow cytometer based sorter instrument to isolate the potentially improved glycosynthase. However, prior to running samples through the FACS machine, the sample preparation methods and incubation periods for various reactions were optimized using a flow cytometer.

4.2 Materials and methods

4.2.1 Generation of random library of mutants

Error-prone PCR setup via sequence ligation independent cloning (SLIC) approach

For insert PCR, 0.5 µM of forward and reverse primers (indicated in Table 4) were mixed with 20 ng of plasmid DNA of Tm0306_WT with 0.2 mM of dATP and dGTP, 1 mM of dCTP and dTTP in a 100 µl total reaction volume. The reaction was performed in 1X Taq buffer with 1.25 U of Taq DNA polymerase. 0.1 mM and 0.5 mM $MnCl_2$ was taken in different tubes with (labeled as I1 and I2) and without (labeled as I3 and I4) 1.5 mM and 7 mM $MgCl_2$. The PCR conditions used for insert PCR product amplification are indicated in Table 5. For Vector PCR products, 0.5 µM of forward and reverse primers (indicated in Table 4) were mixed with 20 ng of plasmid DNA of Tm0306_WT in 1X Phusion Master mix in a 50 µl total reaction volume (labeled as V1 and V2). The PCR conditions used for Vector PCR are indicated in Table 6.

Primer name	Primer sequence	Melting temperature
Tm0306_WT_epPCR_Vector_Forward	GAATAAGGATCCTCTAGAGTCGAC	57.5°C
Tm0306_WT_epPCR_Vector_Reverse	CATGGCGATCGCCTGG	57.7°C
Tm0306_WT_epPCR_Insert_Forward	CCAGGCGATCGCCATG	57.7°C
Tm0306_WT_epPCR_Insert_Reverse	GTCGACTCTAGAGGATCCTTATTC	57.5°C

Table 4: Primer sequences of vector and insert for error prone PCR

Process	Temperature	Time
Initial Denaturation	95°C	60 s
Denaturation	95°C	30 s
Annealing	60°C	30 s
Extension	68°C	3 min
Final Extension	68°C	5 min
Hold	10 °C	∞
No. of cycles	20	

Table 5: Reaction conditions for Insert PCR

Process	Temperature	Time
Initial Denaturation	98°C	30 s
Denaturation	98°C	10 s
Annealing	60°C	30 s
Extension	72°C	3 min
Final Extension	72°C	5 min
Hold	10 °C	∞
No. of cycles	30	

Table 6: Reaction conditions for Vector PCR

DNA gel for PCR amplification check, PCR product purification

Once PCR is complete, 2 µl of the PCR product was mixed with 3 µl PCR water and 1 µl of the Purple loading dye and run in SYBR safe DNA gel alongside 5 µl of DNA ladder at

120 V for 40 minutes. With the remaining PCR products, PCR product purification was performed using PCR extraction kit from IBI Scientific.

DNA gel for purified PCR product, DNA concentration measurement

Next, 2 µl of the purified PCR product was mixed with 3 µl of PCR water and 1 µl of Purple loading dye alongside 5 µl of DNA ladder and run at 120 V for 40 minutes. The gel was imaged using Gel Doc EZ Imager. The concentration of the purified PCR products was calculated using the gel image.

Dpn1 digestion, SLIC and transformation

Reaction mixtures were prepared for Dpn1 digestion. Next, 100 ng of V1 was taken without insert as a control (Reaction-1), 100 ng of V1 was taken with I1 in the Vector: Insert ratios of 1:2.5, 1:5 and 1:10 (Reactions 2,3 and 4 respectively), 100 ng of V1 was taken with I2 in the Vector: Insert ratios of 1:2.5, 1:5 and 1:10 (Reactions 5,6 and 7 respectively), 100 ng of V1 was taken with I4 in the Vector: Insert ratios of 1:2.5, 1:5 and 1:10 (Reactions 8,9 and 10 respectively) in 1X Cut smart buffer in a 10 µl total reaction volume and were digested using 20U of Dpn1 at 37°C for 1 hour. After DPn1 digestion, 1.5U of T4 DNA Polymerase in NEB buffer 2.1 was added to the PCR reaction mixture in a total reaction volume of 20 µl and incubated at 25°C for 5 minutes for SLIC (Sequence Ligation Independent Cloning). The PCR products were incubated on ice immediately after the SLIC run and transformed into E.cloni 10 g cells and incubated at 37°C for 2 hours. The transformation mixture was plated on LB-agar plate with 50 µg/ml kanamycin and incubated at 37°C for 16 hours. Several colonies were observed on the LB agar plates and colony screening was performed to figure out the right colonies.

Colony Screening

For colony screening, 30 random colonies were picked from Insert plate (Reaction 3), 30 random colonies were picked from Insert plate (Reaction 9), 5 random colonies were picked from Vector plate (Reaction 1) and transferred to a PCR plate (PCR plate-1) with 5 µl PCR water and incubated at 95°C for 5 minutes. Also, the tip which was used to pick up a particular colony was transferred to LB media with 50 µg/ml kanamycin and incubated at 37°C for 14 - 15 hours. 1 µl of colony from the PCR plate 1 was added to 0.5 µM NcoI forward (TTGCTTTGTGAGCGGATAAC) and 0.5 µM T7 terminator reverse (GCTAGTTATTGCTCAGCGG) primers. The reaction was performed in 1X Master mix in total reaction volume of 40 µl in PCR Plate-2. The reaction conditions for colony screening are indicated in Table 7.

Process	**Temperature**	**Time**
Initial Denaturation	98°C	30 s
Denaturation	98°C	10 s
Annealing	50°C	30 s
Extension	72°C	3 min
Final Extension	72°C	5 min
Hold	10 °C	∞
No. of cycles	30	

Table 7: Colony screening PCR conditions

After colony screening PCR was complete, 2 µl of the PCR reaction mixture was added to 3 µl PCR water and 1 µl of the Purple loading dye alongside 5 µl of the DNA Ladder and

loaded onto a DNA gel and run at 120 V for 40 minutes. The DNA gel was imaged using Gel Doc EZ Imager and the positive colonies were identified. The positive colonies were purified using PCR extraction kit and sent for DNA sequencing. The grown colonies were also sent for DNA sequencing after performing mini-prep plasmid extraction.

4.2.2 Glycosynthase reaction *in-vivo* coupled with click chemistry for flow cytometry

For evaluating the glycosynthase activity inside the cells, starter cultures were always grown from freshly transformed colonies. The colonies were picked and inoculated in LB media for the following: Tm0306_WT (as a control) with 50 µg/ml kanamycin and chloramphenicol and Tm0306_D224G (as the sample) with kanamycin and chloramphenicol. The starter cultures were incubated at 37°C for 16 hours. Next, 500 µl starter cultures (5% of the total volume of 10 ml sized cultures) were transferred to 10 ml cultures with the same antibiotics and incubated at 37°C for around 4 hours. Then, 2 ml of the sample from the 10 ml cultures were taken in cuvettes with only LB media+antibiotics as control and OD600 measured using the UV spectrophotometer. It was confirmed that the cells were in exponential phase (OD600=0.4-0.8) before 0.5 mM IPTG was used for protein expression induction of all cultures (Tm0306_WT, Tm0306_D224G) at 25°C for 20-24 hours. The cell cultures are then centrifuged at 8,000 rpm for 15 minutes at 4°C and supernatants discarded. Even after removing the supernatant, there are some impurities that are stuck to the outside of the cells which are removed by washing the cells twice with 1X PBS buffer pH=7.4. The samples were then re-suspended in 2 ml of 1X PBS buffer pH=7.4. Next, 2 ml of the re-suspended cultures were divided into various aliquots for the reaction. The following samples are taken in triplicates for this particular reaction: WT/D224G + pNP-X (Acceptor sugar Control), WT/D224G + β-L-Fucosyl azide (Donor sugar Control),

WT/D224G (Enzyme control), WT/D224G + β-L-Fucosyl azide + pNP-X (desired reaction mixture). Every control/reaction mixture consists of 250 µl cells, 250 µl of 100 mM pNP-X (wherever needed or DI water) and 50 µl of 100 mM β-L-Fucosyl azide (wherever needed or DI water). These controls/reactions are incubated at 37°C for around 20 hours in the dark and centrifuged at 13,500 rpm for 5 minutes at room temperature and supernatants were discarded. Another round of centrifugation was performed with the same conditions to remove excess supernatant. The cell pellets obtained were re-suspended in 400 µl of 1X PBS pH=7.4. Next, 50 µM of DBCO-PEG4-Fluor 545 was added to the reaction mixture and incubated at 37°C for 30 minutes. After 30 minutes, the samples were centrifuged at 13,500 rpm for 15 minutes at room temperature and resulting supernatants were removed. Another round of centrifugation was performed with the same conditions to remove excess supernatant. The samples were re-suspended in 400 µl of 1X PBS buffer pH=7.4 and run fresh (immediately) in a Guava easyCyte flow cytometer (EMD Millipore) with blue laser (488 nm excitation, 595 nm emission).

4.2.3 Optimization of DBCO-PEG4-Fluor 545 concentration and flow cytometer gain settings

The experimental method used for this Section is same as Section 4.2.2. The only difference being three different concentrations used for optimization of DBCO-PEG4-Fluor 545 concentration inside the cells: 5 µM, 25 µM, 50 µM. For adjustment of gain settings, the same experiment was conducted and the flow cytometer experiment was run thrice on the same samples at three different gain settings: 56.2, 86.5, and 112.

4.2.4 FACS based analysis of Wild type and D224G enzyme expressing cells

Plasmid DNA pEC_Tm0306_WT and pEC_Tm0306_D224G was transformed into BL21 (DE3) cells and colonies were observed in a span of 12-16 hours. Fresh colonies were used for the inoculation of starter cultures of pEC_Tm0306_WT and pEC_Tm0306_D224G in LB media with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol. The starter cultures were incubated at 37°C for 16 hours. 1 ml starter cultures (5% of the total volume of larger cultures) were transferred to 20 ml larger volumed cultures in conical flasks with the same antibiotics and incubated at 37°C for around 3-4 hours until OD600 reached the exponential phase (OD600=0.6-0.8). Next, 0.5 mM IPTG expression was induced in the cultures and incubated at 37°C for 1 hour for protein expression. OD600 was measured after one hour of IPTG induction. The cell cultures were then centrifuged at 8,000 rpm for 10 minutes at 4°C and their supernatants are discarded. Even after removing the supernatant, there are some impurities that are stuck to the outside of the cells which are removed by washing the cells twice with 1X PBS buffer pH=7.4. The OD600 of the samples was adjusted by re-suspending them in minimal media, pH= 7.03 (since this solution contains approximately 22 mM glucose necessary for cell survival during the course of the glycosynthase reaction). The amount of minimal media used to re-suspend the cultures was dependent on the OD600 value after IPTG induction. By re-suspending in minimal media, the amount of cells in each culture was approximately the same and it was confirmed by OD600 measurement after re-suspension. The following samples were prepared for *in-vivo* GS reaction: Tm0306_WT/D224G + 10 mM β-L-Fucosyl azide + 25 mM pNP-X (reaction mixture) and Tm0306_WT/D224G + water (control). These samples were incubated at 37°C for 12 hours in dark and centrifuged at 10,000 rpm for 3 minutes at room temperature and resultant

supernatants were discarded. Another round of centrifugation was performed with the same conditions to remove excess supernatant. The cell pellets obtained were re-suspended in equal amounts of 1X PBS pH=7.4 for both cell cultures. Finally, 50 µM of DBCO-PEG4-FLUOR 545 was added to the reaction mixture and incubated at 37°C for 30 minutes. After 30 minutes, the samples were centrifuged at 10,000 rpm for 3 minutes at room temperature and all supernatants were removed. Another round of centrifugation was performed with the same conditions to remove excess supernatant. Unstained samples were also taken as controls. The samples were then re-suspended in equal amounts of 1X PBS buffer pH=7.4 and immediately run on a FACS instrument (Beckman Coulter MoFlo XDP Cell Sorter) with 488 nm laser used for FACS (FL2) sorting.

4.2.5 FACS sorting of epPCR library

The error-prone PCR library was generated as detailed in Section 4.2.1. After the PCR reaction, the PCR mixture was run on a DNA gel and the products were extracted using the gel extraction protocol. After gel extraction, the PCR products were purified using the PCR clean-up kit from IBI Scientific. DPn1 digestion was performed at 37°C for 1 hour and SLIC was performed at 25°C for 5 minutes on the extracted products. The SLIC reaction mixture was transformed into E.cloni 10 g cells and incubated at 37°C for 2 hours in SOC media for recovery. After 2 hours, the transformation mixtures were directly transferred to 5 ml LB media as inoculum and grown at 37°C for 16 hours. Next, 1 ml starter cultures (5% of the total volume of larger cultures) were transferred to 20 ml volume cultures in conical flasks with the same antibiotics and incubated at 37°C for around 2-3 hours until OD600 reached the exponential phase (OD600=0.4-0.8). Then, 1 mM IPTG expression was induced in the cultures and incubated at 37°C for 1 hour for protein

expression. OD600 was measured after one hour of IPTG induction and 1 ml of the cell cultures were taken out in sterile micro-centrifuge tubes and centrifuged at 10,000 rpm for 3 minutes at 25°C to remove supernatants. Even after removing the supernatant, there are some impurities that are stuck to the outside of the cells which are removed by washing the cells twice with 1X PBS buffer pH=7.4. The samples were then re-suspended in 60 µl of 1X PBS pH 7.4 and 10 mM β-L-Fucosyl azide and 25 mM pNP-Xylose were added to makeup a total reaction volume of 150 µl. This solution was then incubated at 37°C for 2 hours for the glycosynthase reaction to take place. After 2 hours, the samples were centrifuged at 10,000 rpm for 3 minutes at 25°C and all supernatants were discarded. The cells were centrifuged again at the same conditions to remove excess supernatant. The samples were re-suspended in PBS buffer and then 50 µM DBCO-PEG4-Fluor 545 was added into a total reaction volume of 150 µl and incubated at 37°C for 30 minutes. After 30 minutes, the samples were centrifuged at 10,000 rpm for 3 minutes at room temperature and all supernatants removed. Another round of centrifugation was performed with the same conditions to remove excess supernatant. Unstained samples and D224G (the template DNA) were also taken as controls. The samples were then re-suspended in 1 ml of 1X PBS buffer pH=7.4, filtered using 40 µm filter and run on a FACS instrument (BD Influx High Speed Sorter) with 561 nm excitation laser.

4.3 Results and discussion

4.3.1 PCR amplification confirmation, DNA concentration calculation using gel

The PCR amplification was confirmed using 1% agarose gel as shown in Figure 30A. I1, I2, I4 showed band around 1500 base pairs which corresponds to the length of the insert in Tm0306_WT. I3 did not show amplification of the product, hence it was not used for any future trials. V1 and V2 showed band around 4000 base pairs, which corresponds to the length of the vector pEC in pEC_Tm0306_WT.

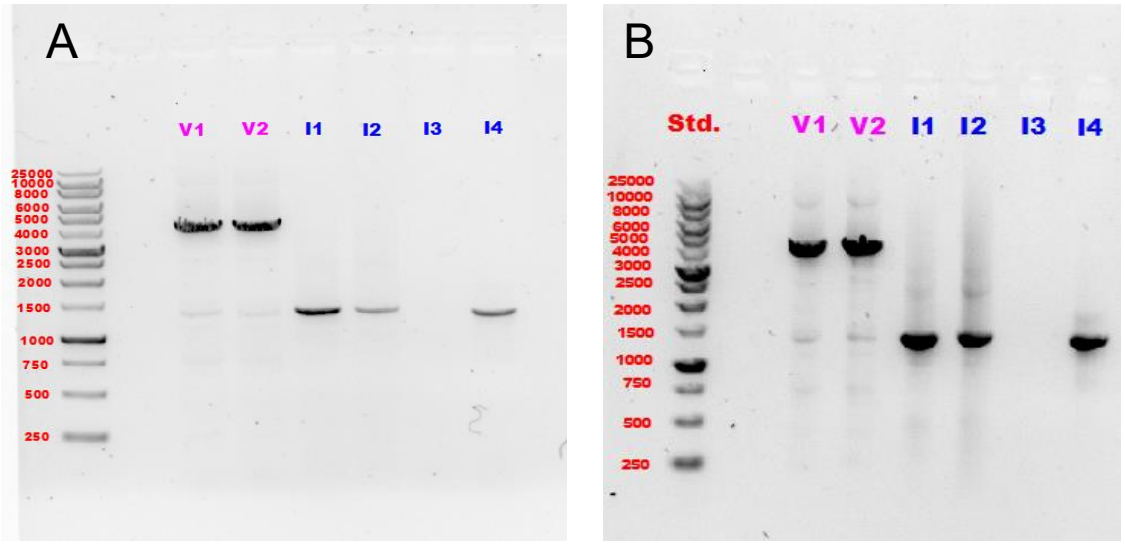


Figure 30: A.) DNA gel image of Insert and Vector PCR confirming amplification of the products B.) DNA gel image of the purified PCR products

The remaining PCR products were purified and DNA gel was run again. The concentrations of insert and vector purified PCR product was calculated using the DNA gel (Figure 30B). Furthermore, Dpn1 digestion and SLIC was performed on the remaining PCR mixtures and transformed into E.cloni 10 g cells and plated on LB agar plate and multiple colonies were observed.

4.3.2 epPCR test library

30 colonies labeled as test library were picked from the plates and colony PCR was performed. The colony PCR mixture was run on 1% agarose gel and positive colonies were identified. The plasmid DNA from the positive colonies were extracted by growing a small culture overnight and the samples were sent for DNA sequencing. From the DNA sequencing results, all the point mutations were found and tabulated in Table 8.

Matrix of point mutations identified in Tm0306 from constructs									
		Mutation to							
		0.1 mM Mn^{+2} and 7 mM Mg^{+2}				**0.5 mM Mn^{+2}**			
		T	**C**	**A**	**G**	**T**	**C**	**A**	**G**
Mutation from	**T**	0	9	5	0	0	6	6	1
	C	6	0	0	3	8	0	2	1
	A	4	1	0	0	10	0	0	16
	G	0	0	2	0	1	0	9	0
Total mutations		30				60			

Table 8: Matrix of point mutations identified in Tm0306_WT

Also, the DNA sequencing results show that no insertions or deletions had arisen in the gene of interest during the error prone PCR reaction. Table 8 was used to calculate the overall mutation rate and to estimate biases in the mutation spectrum of the error prone PCR library.

Mutational spectrum of the Tm0306 error prone PCR library				
	0.1 mM Mn^{+2} and 7 mM Mg^{+2}		**0.5 mM Mn^{+2}**	
Types of mutation	**Frequency**	**Total Prop**	**Frequency**	**Total Prop**
Transitions				
A → G, T → C	8	26.67%	22	36.67%
G → A, C → T	4	13.33%	17	28.33%
Transversions				
A → T, T → A	9	30.00%	10	16.67%
A → C, T → G	0	0.00%	1	1.67%
G → C, C → G	2	6.67%	1	1.67%
G → T, C → A	1	3.33%	3	5.00%
Summary of bias				
Transitions/transversions				
AT → GC/GC → AT	2		1.29	
A → N, T → N	23	76.67%	39	65.00%
G → N, C → N	7	23.33%	21	35.00%
Mutation rate				
Mutations per kb	2.22		2.78	
Mutations per Tm0306 gene	3		3.75	

Table 9: Mutational spectrum of the error-prone PCR library

There are three key indicators of bias: a) the ratio of transition (Ts) to transversion (Tv) mutations, b) the ratio of AT to GC transitions, and c) the frequency of mutations at A:T base pairs, to mutations at G:C base pairs. The mutation rate and bias measures for the

epPCR library are shown in Table 9. Transitions are interchanges of two-ring purines (A ↔ G) or one-ring pyrimidine's (C ↔ T), as they involve bases of similar shape. Transversions are interchanges of purine for pyrimidine bases, which involve exchange of one-ring and two-ring structures as shown in Figure 31 [59].

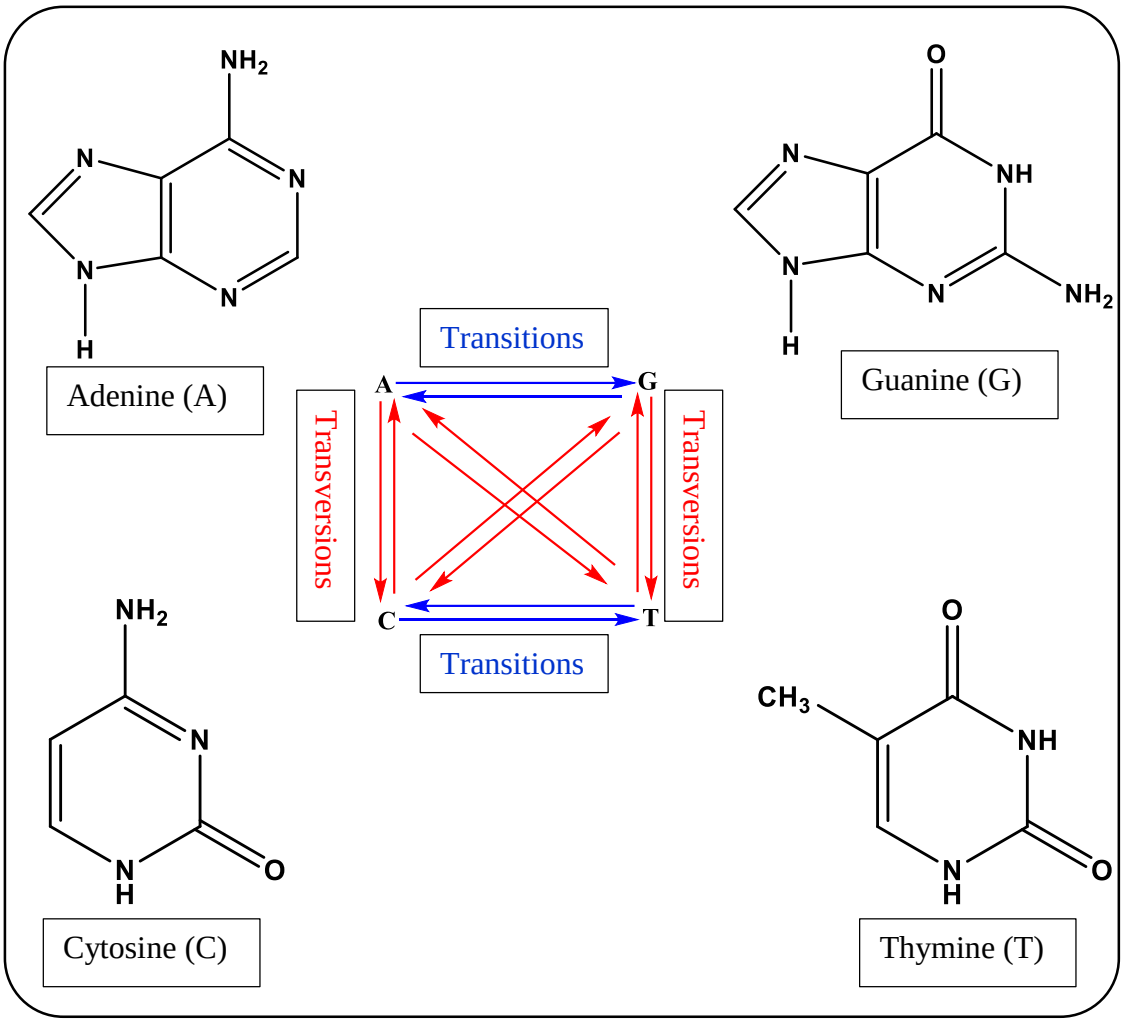


Figure 31: Molecular structures of Adenine (A), Guanine (G), Cytosine (C), Thymine (T) and transitions and transversions

Its important to calculate the PCR efficiency because it allows analysis of library composition. When the total product yield and the starting template are known, the number of doublings in the PCR, d, can be calculated using the following formula:

$$d = \frac{\log\left(\frac{Product}{Template}\right)}{log2}$$

PCR efficiency can be calculated by the following formula:

$$eff = 2^{\frac{d}{n}} - 1$$

where n is the number of cycles.

Parameter	0.1 **mM Mn^{+2} and 7 mM Mg^{+2}**	**0.5 mM Mn^{+2}**
Number of doublings	8.06	7.14
PCR efficiency	0.32	0.28

The epPCR library utility was analyzed using PEDEL-AA software (http://guinevere.otago.ac.nz/stats.html). PEDEL-AA is an easy to use web-interface that requires the following input parameters:

a) Nucleotide sequence

ATGTCGGGAAAATTGACCGTAATCACGGGTCCCATGTACTCCGGAAAGACAACCGAGCTTCTCTCCTTTGTG
GAAATATACAAACTGGGAAAGAAAAAAGTCGCTGTTTTTAAACCAAAAATCGACAGCAGATATCACTCCAC
CATGATCGTCTCTCATTCTGGAAACGGTGTTGAAGCACACGTGATAGAACGTCCTGAAGAAATGCGAAAGTA
CATCGAAGAAGACACCCGGGGAGTTTTCATAGACGAGGTGCAGTTCTTCAATCCGAGTTTGTTTGAAGTGGT
GAAAGATCTTCTCGACAGAGGAATAGATGTCTTCTGTGCAGGGCTCGATCTCACACACAAGCAAAATCCCTT
TGAAACAACCGCTCTGCTCCTCAGTCTCGCCGACACCGTGATCAAAAAGAAGGCAGTCTGTCACAGATGCGG
TGAGTACAACGCTACTCTCACACTAAAAGTAGCAGGAGGTGAAGAAGAAATAGACGTTGGGGGACAGGAAA
AATACATTGCGGTCTGCAGAGACTGTTACAACACCCTCAAAAAACGAGTTGGAACTAGT

b) Library Size (if we assume 10^6)

c) Nucleotide mutation matrix

d) Whether the nucleotide mutation matrix consists of normalized or non-normalized values (Note that we had non-normalized values)

e) Mean number of nucleotide substitutions per daughter sequence

f) Whether Poisson or both Poisson and PCR distributions have to be used (we used both Poisson and PCR distributions)

g) Number of PCR cycles (we had used 20)

h) PCR efficiency

i) Mean number of insertions and deletions per daughter sequence

Summary of library characteristics	**0.1 mM Mn^{+2} and 7 mM Mg^{+2}**	**0.5 mM Mn^{+2}**
Total library size	1.00E+06	1.00E+06
Number of variants with no indels or stop codons	2.82E+05	8.67E+05
Mean number of amino acid substitutions per variant	1.343	2.625
Non-mutated (wild-type) sequences (% of library; PCR est.)	8.54%	11.49%
Number of distinct full-length proteins in the library (Poisson est.)	1.03E+05	5.71E+05
Number of distinct full-length proteins in the library (PCR est.)	1.01E+05	5.23E+05

Table 10: Summary of library characteristics

After providing all the necessary information as mentioned in points from a) to i), the web interface analyzes and gives output as shown in Table 10.

4.3.3 Optimization of DBCO concentration

The efficiency of the *in-vivo* click chemistry reaction depends largely on the uptake of the substrates by the cells. Various concentrations of DBCO-PEG4-FLUOR 545 were thus incubated with WT cells, which showed that increasing the concentration of the fluorescent dye gave a better signal in the flow cytometer (Figure 32). Since, 50 µM concentration of DBCO-PEG4-FLUOR 545 gave a higher signal it was hence chosen for all future studies.

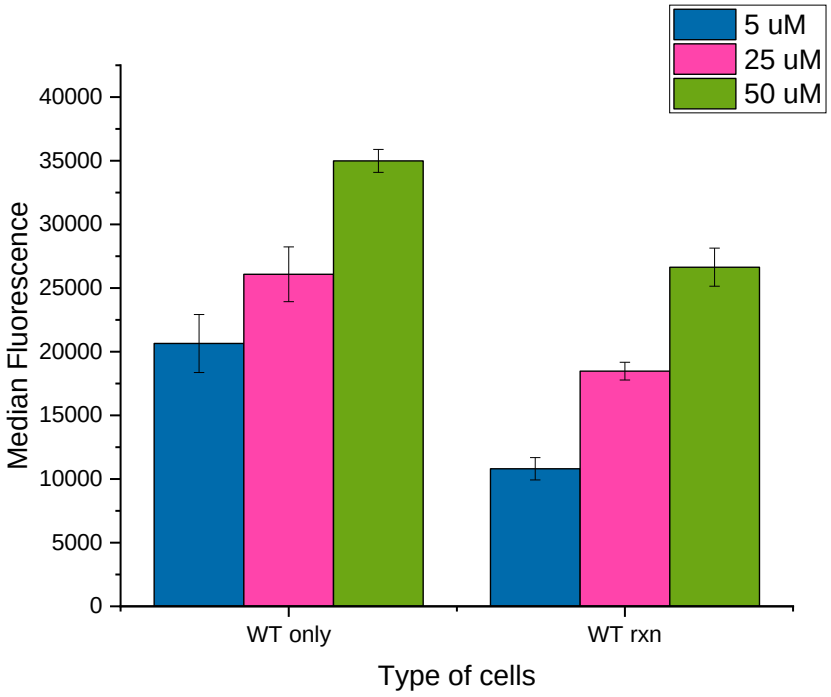


Figure 32: Optimization of DBCO-PEG4-Fluor 545 concentrations for *in-vivo* click reaction

4.3.4 Optimization of flow cytometer gain settings

The *in-vitro* click chemistry fluorescence was measured at 550 nm excitation, 590 nm emission which are the optimal wavelengths for Fluor-545 group. However, the lasers used in the flow cytometry instrument were 488 nm excitation and 585 nm emission. Since, the

laser used was not the optimal excitation laser, the emission intensities were expected to be reduced. To increase the signal obtained from the instrument, higher gain was needed to be used. Different gains were tried which could give an optimal detection range (not near the instruments noise and detection limit). From Figure 33, we concluded that Gain 56.2 was ideal for our experimental setup.

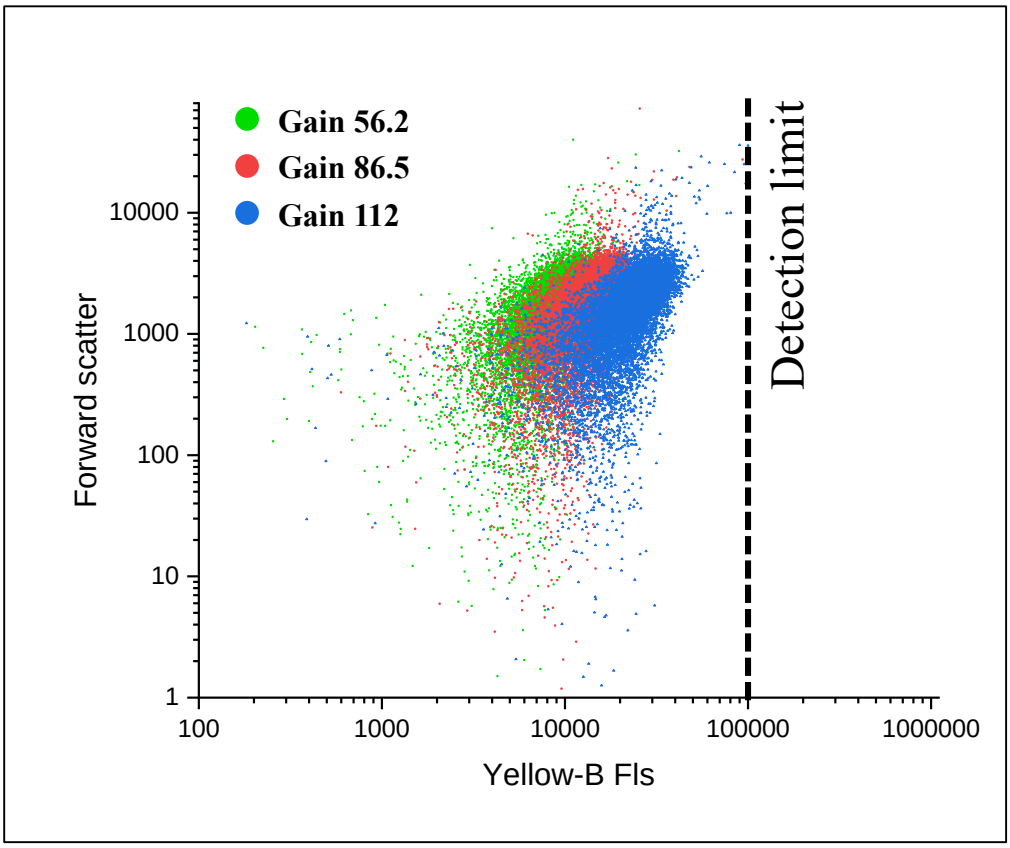


Figure 33: Effect of various gains on the fluorescence signal

4.3.5 Analysis of Tm0306-WT and Tm0306-D224G population after GS reaction

The *in-vivo* GS reaction carried out in Tm0306_D224G cells and control Tm0306_WT cells were stained with DBCO-PEG4-Fluor 545 using click chemistry. The mixture of

cells were analyzed using a Guava flow cytometer and Beckman Coulter MoFlow Cell Sorter to observe the fluorescence spectrum distribution of cell population.

Flow cytometer analysis:

In the flow cytometer instrument, SSC (side scatter) vs. FSC (forward scatter) plot is used to determine the cell population. The gate, labeled as R1 in Figure 34 can be adjusted according to the cell population. FSC provides the size of the cells and SSC decides the granularity and refractivity of the cells. In other words, SSC provides information on the morphology of the cells. The plot of FSC vs SSC is shown in Figure 34.

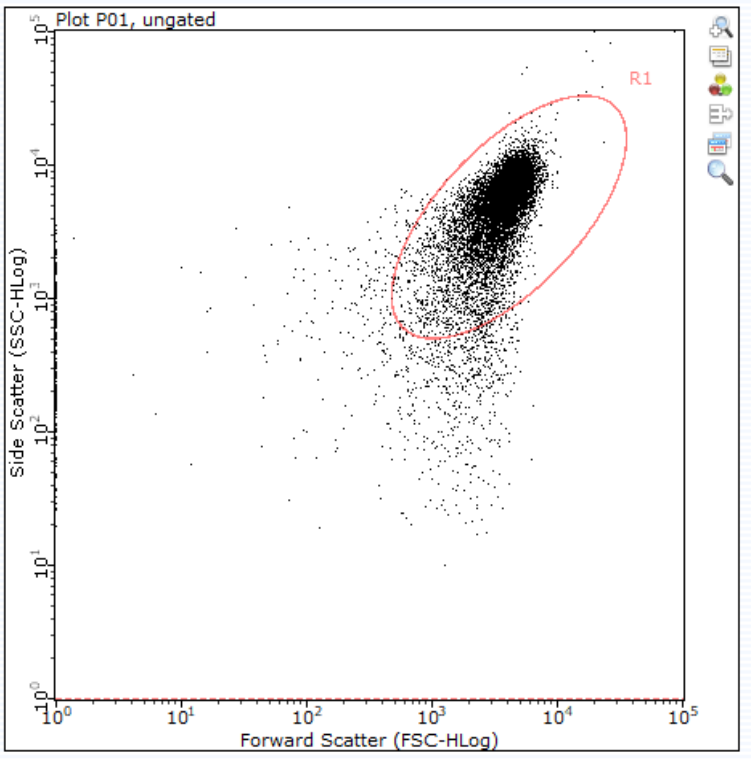


Figure 34: Plot of FSC vs. SSC

Once the cell population region was decided, the samples were run in the flow cytometer by capturing 10,000 events for 60 seconds and plot of Yellow-B fluorescence vs. SSC and histogram of cell count vs. Yellow-B fluorescence were obtained for every sample. An

illustrative plot of Yellow-B fluorescence and Side-scatter containing the gate R3 is shown in Figure 35 and histogram plot containing the gate R4 is shown in Figure 36.

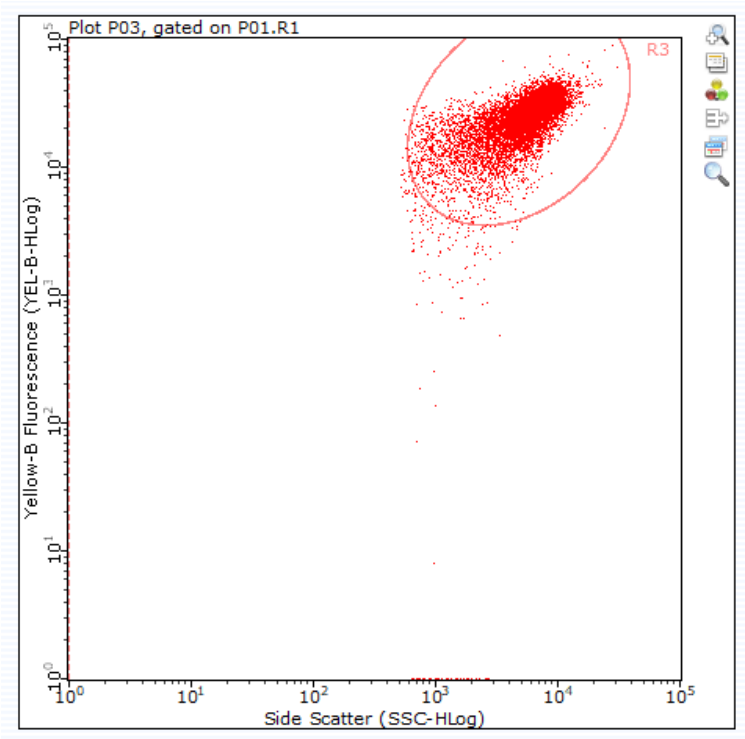


Figure 35: Plot of Yellow-B fluorescence vs. SSC

Once the fluorescence intensities of all the samples are measured using the flow cytometer, Guava Soft 3.3 is used to export the median fluorescence data to excel for further analysis. The median fluorescence intensities of all controls with only cells and the reaction mixtures were compared. It was observed that the median fluorescence was not indicative of the true population of the spread and hence, further median fluorescence analysis was not taken into consideration. Therefore, the raw data was extracted to plot the entire analyzed population to overlay the WT and D224G signals.

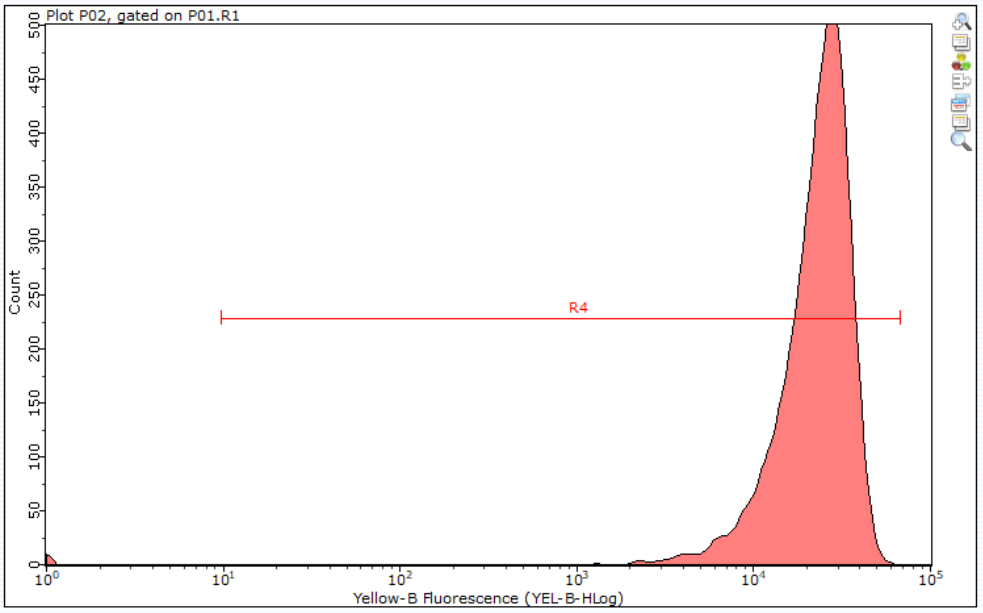


Figure 36: Plot of cell count vs. Yellow-B fluorescence

As seen in Figure 37, the population of D224G had marginally shifted when compared to the WT population. This marginal shift in fluorescence signals could be attributed to the release of free azide in D224G cells. The released free azide had likely reacted with the DBCO-PEG4-Fluor 545 present in the reaction mixture and decreased the Fluor 545 fluorescence signal. This experimental data provided a proof of concept in-vivo validation for difference in signals obtained for an active glycosynthase vs. an inactive enzyme control, although the signal was low due to the poor activity of D224G.

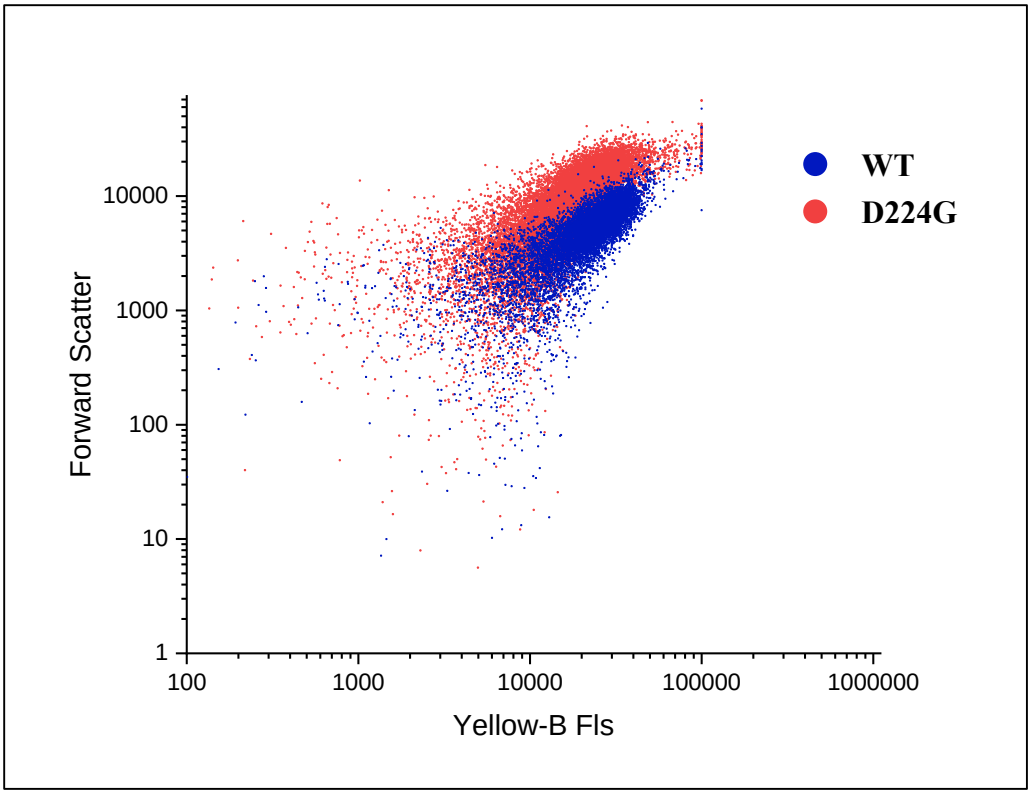


Figure 37: WT and D224G populations in flow cytometer

FACS analysis:

The glycosynthase reaction was carried out inside the cells to facilitate azide release in the D224G reaction as it was supposed to be an active glycosynthase. On the other hand, WT was not expected to release any free azide in the reaction mixture. DBCO-PEG4-Fluor 545 was added to react with the free azide and the organic azide to give different fluorescence intensities for WT and D224G. The samples were run on the FACS instrument and various curves were obtained. FSC (Forward Scatter) vs. FL2 were plotted to check the size and the florescence of the cells. Also a histogram was plotted for Cell count vs. FL2 to more quantivatively analyze the cell population shift of WT versus D224G. The histograms obtained from FACS were used to extract the raw data from an online software called Web Plot Digitizer. With the help of this software, individual data points were extracted for

unstained cells, WT, and D224G and plotted on the same graph for cell count vs. FL2 and overlayed to clearly observe the fluorescence shift. This behavior can be seen in Figure 38. From Figure 38, we can observe that the cell population of D224G has shifted towards lower fluorescence but the mean difference is not very drastic. This is expected since D224G is not a highly efficient glycosynthase. Nevertheless, the slight difference in the fluorescence of WT and D224G can be used for sorting, although a mixture of both populations will be expected to be found in significant amounts.

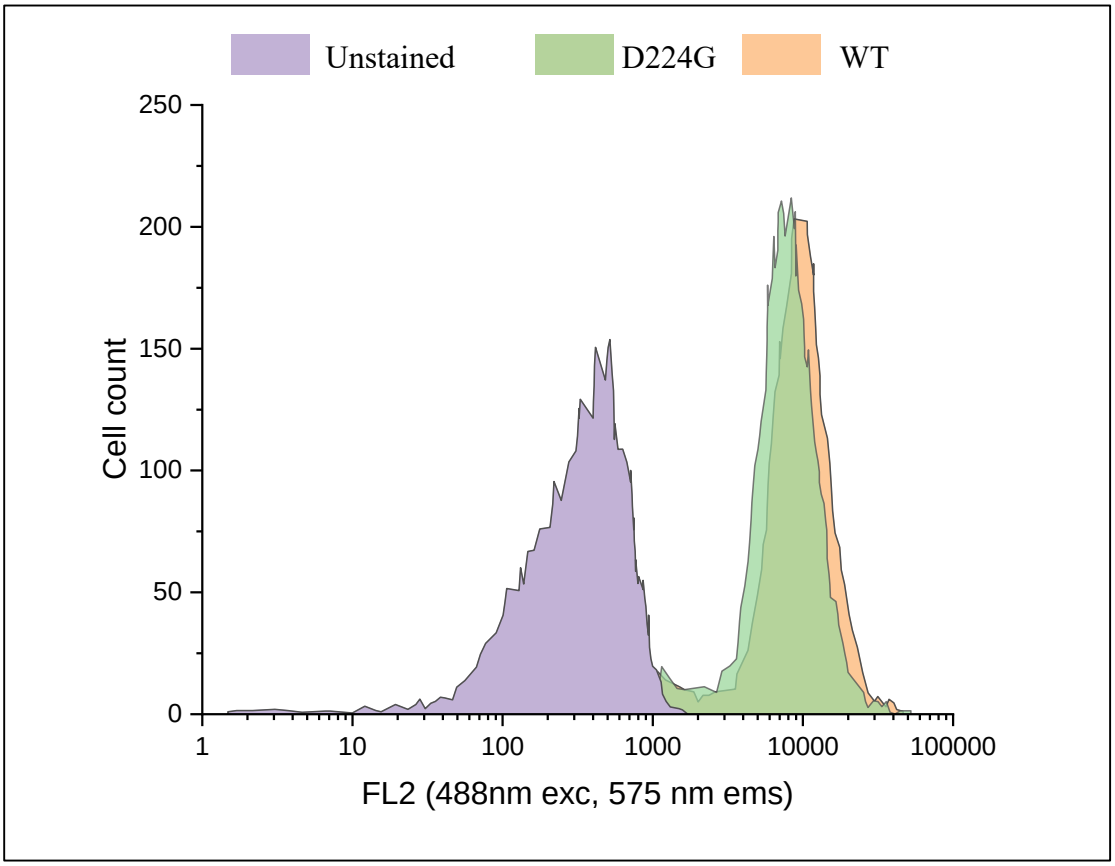


Figure 38: Fluorescence shifts in WT and D224G cells population as observed in FACS machine

4.3.6 Error-prone PCR library and sorting using FACS

Using D224G gene as a template, we generated an epPCR mutant library which were sorted through FACS. Two different types of FACS instruments (with different excitation lasers) were used for sorting of the epPCR library.

FACS sorting using 488 nm excitation laser:

The epPCR library was subjected to similar growth and protein expression conditions. The expressed cultures were used to run glycosynthase reaction for 2 hours at 37°C and subsequently click reaction was performed to prepare the samples for sorting. The stained cells were run through the FACS machine for analysis which showed two different populations as seen in Figure 39**.** The population of cells which showed low fluorescence signal on x-axis were gated as Gate 1 and cells with higher fluorescence were gated as Gate 2 in Figure 39**.** The Gate 2 population was identical to the fluorescent signals from wild type and D224G cells and population in Gate 1, which likely indicates the presence of improved glycosynthase mutants. The samples from Gate1 and Gate2 were appropriately sorted into two different tubes. We sorted 200,000 cells from Gate 2 and 5000 cells from Gate 1 into 1 ml PBS buffer. Next, these 5000 cells were grown and characterized for improved glycosynthase activity. In the event of multiple false positive, a further sorting of these 5000 cells was necessary to be carried out. A second and if necessary a third round of sorting could be further performed. After every round of sorting, the cells would need to grow again with antibiotics as starter culture which would later be transferred to larger cultures. IPTG expression would be done in cultures and with required amount of washing steps and re-suspension of cells, GS reaction would be performed inside the cells. Ideally, in second or third round of sorting, one would be able to readily identify a mutant that is

an active glycosynthase. After the third round of sorting, glycosynthase activity can be confirmed by performing chemical rescue activity of the culture followed by extracting the isolated cell plasmids and performing DNA sequencing. The novel mutants identified can be expressed and *in-vitro* GS reaction of the purified protein can be conducted, along with detailed HPLC/TLC analysis, to confirm improved glycosynthase activity.

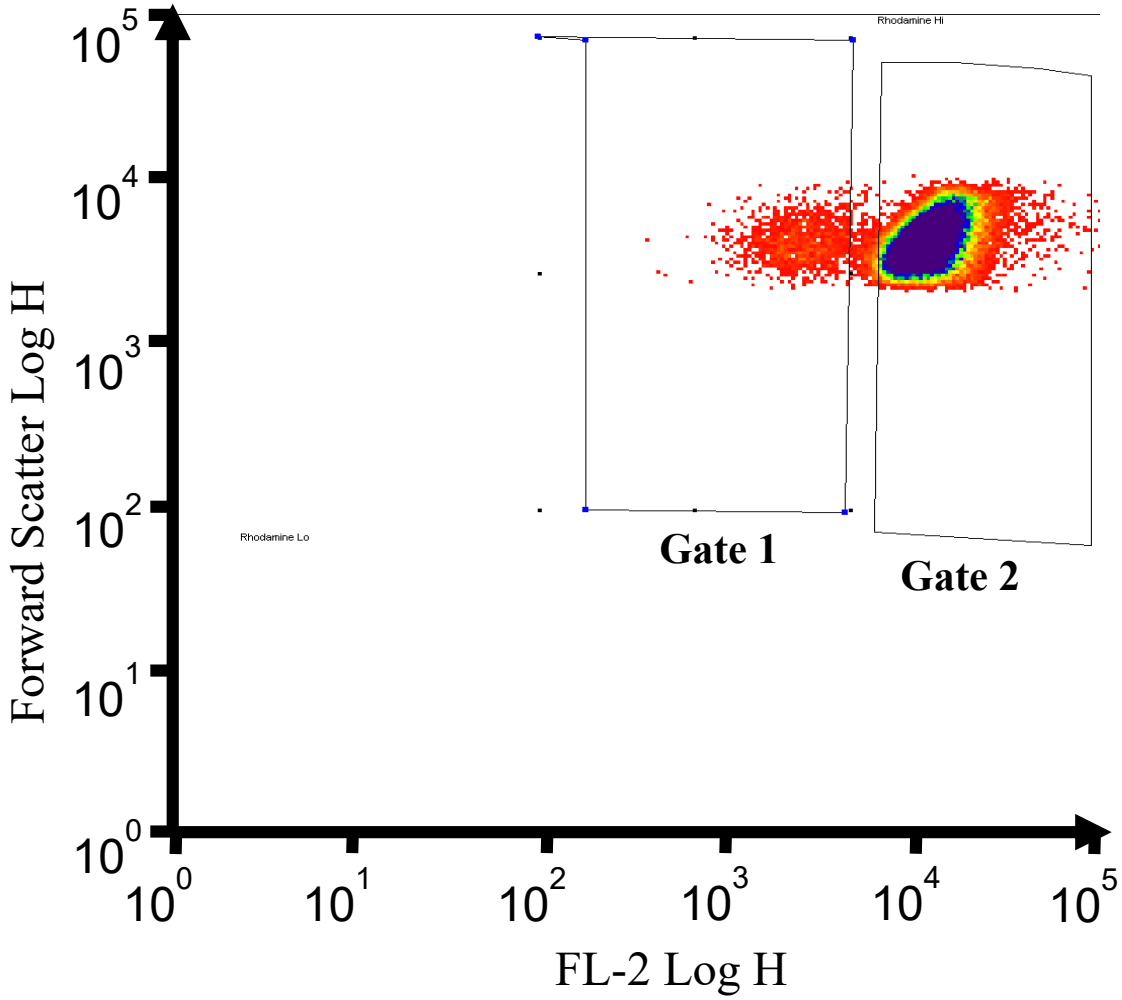


Figure 39: Population gates for sorting epPCR generated mutant libraries

FACS sorting using 561 nm excitation laser:

The epPCR library was subjected to same growth and protein expression conditions as indicated in the materials and methods section earlier. The expressed cultures were used to

run glycosynthase reaction for 2 hours at 37°C and subsequently click reaction was performed to prepare the samples for sorting. The stained cells were run through the FACS machine for analysis (using 561 nm excitation laser) which again showed two different populations as seen in Figure 40. The cell population in gate-high is indicative of the population of cells with higher fluorescence intensities. Gate-low is indicative of the population of cells with lower fluorescence intensities and thus is likely enriched in a library with mutations that help increase glycosynthase activity.

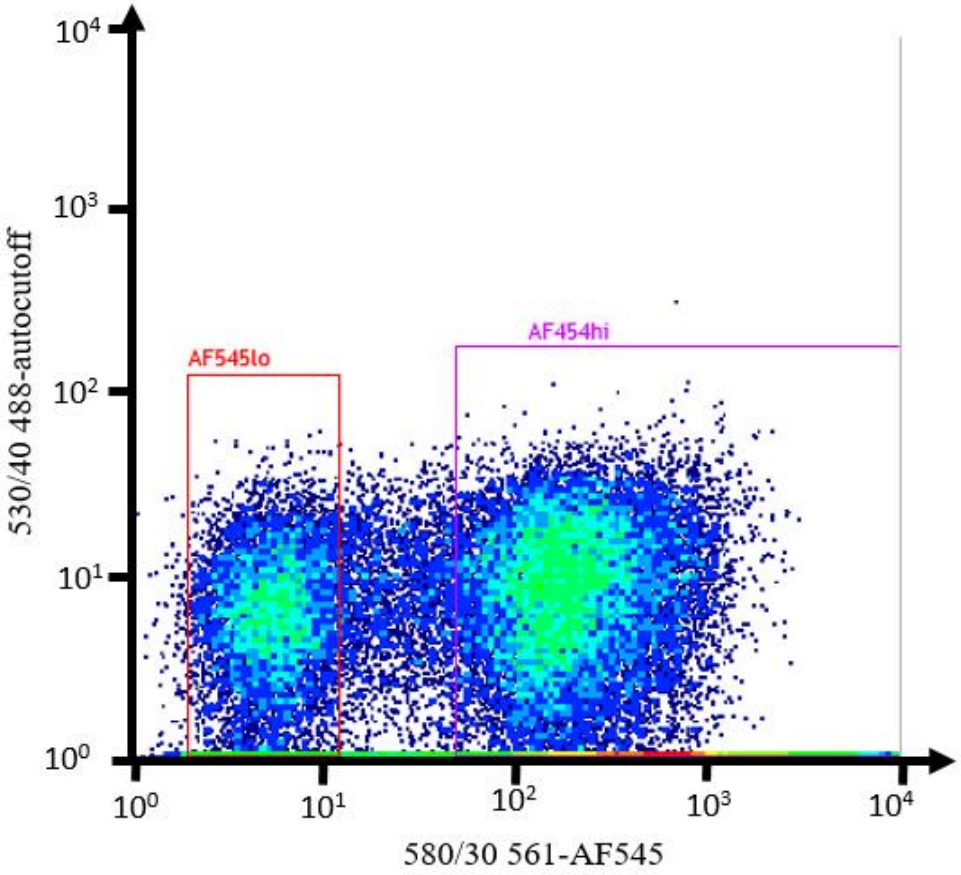


Figure 40: Population gates low and high for first round of sorting epPCR generated mutant libraries with 561 nm excitation laser.

These populations were sorted in 1 ml LB media with respective antibiotics. These samples were then grown as starter cultures at 37°C for 16 hours for second round of sorting. Next,

500 µl of the starter cultures were transferred to 10 ml bigger cultures in LB media with respective antibiotics and incubated at 37°C until OD600=0.4. Once the cells reached an exponential phase, 1 mM IPTG induction was done at 37°C for 1 hour. After 1 hour, the cells were washed twice with 1 X PBS pH 7.4.

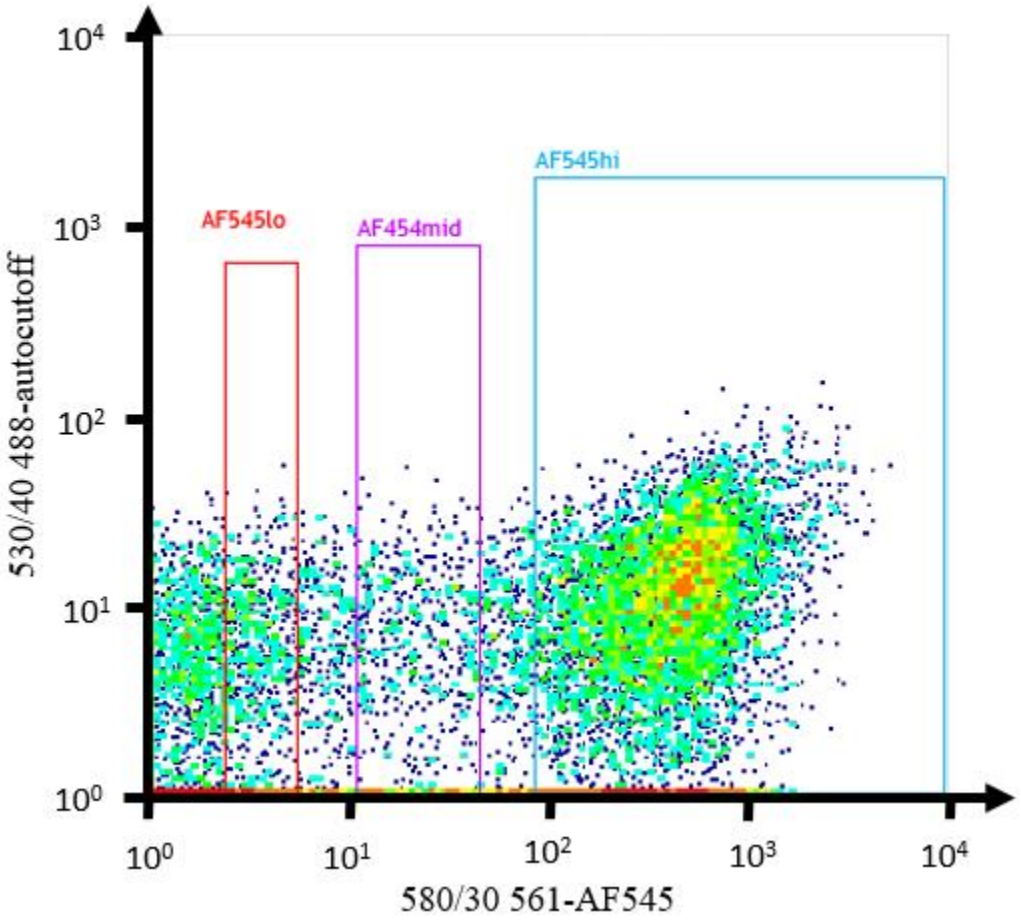


Figure 41: Population gates low, medium and high for second round of sorting epPCR generated mutant libraries with 561 nm excitation laser.

Here, the exact same protocol from the first round of sorting was followed to prepare samples for the second round and the samples were run on the same FACS instrument with identical detection and sorting protocol. We observed that the fluorescence of the cells was distributed over a wide range and gated as low, medium, and high fluorescence populations. The cell populations from the various gates shown in Figure 41 were sorted

as individual cells in a 96-well plate in 100 µl LB media with respective antibiotics. To determine whether an improved glycosynthase was present in these collected cell populations, further processing of the samples was performed as follows: The 96-well plate containing individual FACS sorted cells of low fluorescence intensities were first grown as starter culture at 37°C for 16 hours. Next, 50 µl of the starter culture was transferred to 1 ml LB media with respective antibiotics in a 2 ml deep-well plate. The cultures were grown at 37°C until OD600 reached an exponential phase. Next, 1 mM IPTG expression was performed at 25°C for around 20 hours. The cells with expressed proteins were tested for their chemical rescue activity by adding 50 µl cells into a solution containing 2.5 mM pNP-Fucose and 1 M sodium azide. The chemical rescue experiment was performed at 60°C for 1 hour. After 1 hour, the 96-well plate was centrifuged and 10 µl of the supernatant was taken in a new micro-well plate and mixed with 90 µl DI water and 100 µl of 0.1 M NaOH and absorbance was measured at 405nm. Various sorted cell populations showed chemical rescue activity and are currently being sequenced to determine exact mutations responsible for potentially increased glycosynthase activity.

Chapter-5: Conclusions and future work

Rational engineering and directed evolution of glycosynthases has been used over the last two decades to engineer better glycan synthesizing enzymes. However, previous studies have mostly used low-throughput screening techniques such as LB agar plate based colorimetric screening or microplate based screening based on fluoride sensors. Although, these techniques have successfully identified significantly improved glycosynthase mutants, as these screening methods are very high-throughput or broadly applicable for diverse glycosynthase families. Since positive mutants are often in low abundance, such rare mutants could be difficult to identify if the size of the library is small. Hence, it is essential to increase the library size and have a suitable high-throughput screening method for identifying a improved glycosynthase mutants.

In this study, we aimed to develop a universal azide based screening technique using a novel click chemistry based product detection method. This screening technique was validated for a model GH29 based fucosynthase to establish the proof-of-concept technology for directed evolution of glycosynthases. We also successfully replicated the *Thermotoga maritima* Tm0306 GH29 construct based fucosynthase work reported earlier [19]. The D224G mutant of Tm0306 was characterized and shown to synthesize two glycosynthase products using β-L-fucopyranosyl azide and pNP-β-D-Xylose as donor and acceptor sugars, respectively. We used copper free click chemistry to label the azides and interestingly observed a clear difference between the inorganic product free azide and the organic substrate azide which was further used for in-vitro and in-vivo differentiation between an active or inactive glycosynthase mutant. A detailed analysis of the click chemistry products at various temperatures was performed and reaction kinetics analysis

was conducted. A mechanistic analysis revealed that the triazole moiety formed during the Click reaction had a direct influence on the fluorescence of the fluorophore.

Flow cytometry was used as a tool for analyzing the *in-vivo* click chemistry and glycosynthase reactions. A systematic protocol for capturing and analyzing the cell population was established and fluorescence measurements were recorded. A distinct population of cells was observed for the populations incubated for click reaction as compared to the cells incubated only with the fluorescent alkyne. During the analysis of the cells for glycosynthase reactions, we observed that there was a large overlap in the population between WT and D224G cells. It was hypothesized that this lower gap was due to the lower glycosynthase activity of D224G.

In order to improve the glycosynthase activity of D224G, the template construct DNA was subjected to random mutagenesis using error-prone PCR to generate a large library of random mutants. This full-scale epPCR library was subjected to the same experimental conditions that were optimized during flow cytometer analysis and cells sorted using a FACS instrument. The samples which were incubated for 2 hours of glycosynthase reaction showed a clear spread in cell population on fluorescence axis. The isolated constructs with improved activity would be potentially used for synthesizing fucosylated glycans.

The protocol for screening the glycosynthase mutants was setup using pNP-Xylose as an acceptor sugar group. However, this protocol could be readily applied for various other acceptors such as lactose, galactose, glucose, N-acetyl galactosamine to isolate respective fucosynthase mutants. The improved constructs can be used for synthesizing fucosylated sugars that are a critical component of human milk oligosaccharides. Furthermore, the

developed protocol can be readily extended to engineer other families of glycosyl hydrolases to glycosynthases that use azide based donor sugars.

Bibliography

[1] N. J. Agard, “Chemical approaches to glycobiology,” *ACS Symp. Ser.*, vol. 990, pp. 251–271, 2008.

[2] J. A. Prescher and C. R. Bertozzi, “Chemical Technologies for Probing Glycans,” *Cell*, vol. 126, no. 5, pp. 851–854, 2006.

[3] L. R. Ruhaak, S. Miyamoto, and C. B. Lebrilla, “Developments in the Identification of Glycan Biomarkers for the Detection of Cancer,” *Mol. Cell. Proteomics*, vol. 12, no. 4, pp. 846–855, 2013.

[4] J. E. Hudak and C. R. Bertozzi, “Glycotherapy: New advances inspire a reemergence of glycans in medicine,” *Chem. Biol.*, vol. 21, no. 1, pp. 16–37, 2014.

[5] O. J. Plante, E. R. Palmacci, and P. H. Seeberger, “Automated solid-phase synthesis of oligosaccharides,” *Science (80-.).*, vol. 291, no. 5508, pp. 1523–1527, 2001.

[6] J. D. C. Codée, L. J. Van Den Bos, R. E. J. N. Litjens, H. S. Overkleeft, J. H. Van Boom, and G. A. Van Der Marel, “Sequential one-pot glycosylations using 1-hydroxyl and 1-thiodonors,” *Org. Lett.*, vol. 5, no. 11, pp. 1947–1950, 2003.

[7] C.-W. Cheng *et al.*, “Hierarchical and programmable one-pot synthesis of oligosaccharides,” *Nat. Commun.*, vol. 9, no. 1, p. 5202, 2018.

[8] L. X. Wang and B. G. Davis, “Realizing the promise of chemical glycobiology,” *Chem. Sci.*, vol. 4, no. 9, pp. 3381–3394, 2013.

[9] L. Krasnova and C.-H. Wong, *Understanding the Chemistry and Biology of Glycosylation with Glycan Synthesis*, vol. 85, no. 1. 2016.

[10] E. C. O’Neill and R. A. Field, “Enzymatic synthesis using glycoside phosphorylases,” *Carbohydr. Res.*, vol. 403, pp. 23–37, 2015.

[11] C. Breton, L. Šnajdrová, C. Jeanneau, J. Koča, and A. Imberty, “Structures and mechanisms of glycosyltransferases,” *Glycobiology*, vol. 16, no. 2, pp. 29–37, 2006.

[12] L. L. Lairson, B. Henrissat, G. J. Davies, and S. G. Withers, “Glycosyltransferases: Structures, Functions, and Mechanisms,” *Annu. Rev. Biochem.*, vol. 77, no. 1, pp. 521–555, 2008.

[13] F. Xue *et al.*, “Expression of Codon-Optimized Plant Glycosyltransferase UGT72B14 in Escherichia coli Enhances Salidroside Production,” *Biomed Res. Int.*, vol. 2016, 2016.

[14] T. J. Boltje, T. Buskas, and G. J. Boons, “Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research,” *Nat. Chem.*, vol. 1, no. 8, pp. 611–622, 2009.

[15] S. Ruiz-Moyano *et al.*, “Variation in consumption of human milk oligosaccharides by infant gut-associated strains of bifidobacterium breve,” *Appl. Environ. Microbiol.*, vol. 79, no. 19, pp. 6040–6049, 2013.

[16] H. Wang *et al.*, “High-temperature enzymatic breakdown of cellulose,” *Appl. Environ. Microbiol.*, vol. 77, no. 15, pp. 5199–5206, 2011.

[17] D. Fleming, L. Chahin, and K. Rumbaugh, “Glycoside hydrolases degrade polymicrobial bacterial biofilms in wounds,” *Antimicrob. Agents Chemother.*, vol. 61, no. 2, pp. 1–9, 2017.

[18] B. David *et al.*, “Internal Water Dynamics Control the Transglycosylation/Hydrolysis Balance in the Agarase (AgaD) of Zobellia galactanivorans,” *ACS Catal.*, vol. 7, no. 5, pp. 3357–3367, May 2017.

[19] B. Cobucci-Ponzano *et al.*, “β-Glycosyl Azides as Substrates for α-Glycosynthases: Preparation of Efficient α-L-Fucosynthases,” *Chem. Biol.*, vol. 16, no. 10, pp. 1097–1108, 2009.

[20] C. Mayer *et al.*, “Directed evolution of new glycosynthases from Agrobacterium β-glucosidase: a general screen to detect enzymes for oligosaccharide synthesis,” *Chem. Biol.*, vol. 8, no. 5, pp. 437–443, May 2001.

[21] M. R. Hayes, K. A. Bochinsky, L. S. Seibt, L. Elling, and J. Pietruszka, “Development of a colourimetric assay for glycosynthases,” *J. Biotechnol.*, vol. 257, pp. 162–170, 2017.

[22] E. Andrés, H. Aragunde, and A. Planas, “Screening glycosynthase libraries with a fluoride chemosensor assay independently of enzyme specificity: identification of a transitional hydrolase to synthase mutant,” *Biochemical Journal*, vol. 458, no. 2. pp. 355–363, 2014.

[23] A. Ben-david, G. Shoham, and Y. Shoham, “Article A Universal Screening Assay for Glycosynthases : Directed Evolution of Glycosynthase XynB2 (E335G) Suggests a General Path to Enhance Activity,” vol. 2, no. June, pp. 546–551, 2008.

[24] J. H. Lee *et al.*, “Quantified High-Throughput Screening of Escherichia coli Producing Poly (3-hydroxybutyrate) Based on FACS,” pp. 1767–1779, 2013.

[25] A. Aharoni *et al.*, “High-throughput screening methodology for the directed evolution of glycosyltransferases,” *Nat. Methods*, vol. 3, no. 8, pp. 609–614, 2006.

[26] G. Yang, J. R. Rich, M. Gilbert, W. W. Wakarchuk, Y. Feng, and S. G. Withers, “Fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases,” *J. Am. Chem. Soc.*, vol. 132, no. 30, pp. 10570–10577, 2010.

[27] T. Bernard, B. L. Cantarel, B. Henrissat, V. Lombard, P. M. Coutinho, and C. Rancurel, “The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics,” *Nucleic Acids Res.*, vol. 37, no. Database, pp. D233–D238, 2008.

[28] H. Wolf *et al.*, “A mouse model for fucosidosis recapitulates storage pathology and neurological features of the milder form of the human disease,” *Dis. Model. Mech.*, vol. 9, no. 9, pp. 1015–1028, 2016.

[29] P. M. Hoogerbrugge *et al.*, “Allogeneic bone marrow transplantation for lysosomal storage diseases,” *Lancet*, vol. 345, no. 8962, pp. 1398–1402, 1995.

[30] M. Miano *et al.*, “Four year follow-up of a case of fucosidosis treated with unrelated donor bone marrow transplantation,” *Bone Marrow Transplant.*, vol. 27, no. 7, pp. 747–751, 2001.

[31] D. A. Sela *et al.*, “The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptions for milk utilization within the infant microbiome,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 48, pp. 18964–18969, 2008.

[32] G. Sulzenbacher *et al.*, “Crystal Structure of Thermotoga maritima ⎵ - L - Fucosidase,” vol. 279, no. 13, pp. 13119–13128, 2004.

[33] B. Cobucci-ponzano *et al.*, “Accelerated Publications Identification of the Catalytic Nucleophile of the Family 29 R - L -Fucosidase from Sulfolobus solfataricus via Chemical Rescue of an Inactive Mutant †,” pp. 9525–9531, 2003.

[34] “Novel α-L-Fucosidases from a Soil.pdf.” .

[35] S. Liu *et al.*, “Identification of Essential Residues of Human R - L -Fucosidase and Tests of Its,” pp. 110–120, 2009.

[36] T. Li *et al.*, “cro Identification and characterization of a core fucosidase from the bacterium Elizabethkingia meningoseptica,” vol. 293, pp. 1243–1258, 2018.

[37] Y. Qiao, “Chemical Rescue of a Mutant Enzyme in Living Cells,” *Science (80-.).*, vol. 311, no. 5765, pp. 1293–1297, Mar. 2006.

[38] M. Faijes, A. Planas, J. Romero-García, T. Pozzo, and E. Nordberg Karlsson, “Rational design of a thermostable glycoside hydrolase from family 3 introduces β-glycosynthase activity,” *Glycobiology*, vol. 27, no. 2, pp. 165–175, 2016.

[39] K. Islam, “The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics,” *Cell Chem. Biol.*, vol. 25, no. 10, pp. 1171–1184, Oct. 2018.

[40] V. K. Tiwari, B. B. Mishra, K. B. Mishra, N. Mishra, A. S. Singh, and X. Chen, “Cu-Catalyzed Click Reaction in Carbohydrate Chemistry,” *Chem. Rev.*, vol. 116, no. 5, pp. 3086–3240, 2016.

[41] J. E. Hein and V. V Fokin, “NIH Public Access,” vol. 39, no. 4, pp. 1302–1315, 2011.

[42] M. G. Hong, V., Persoski, S., Ma, C., Finn, “NIH Public Access,” *Angew. Chem. Int. Ed. Engl.*, vol. 48, no. 52, pp. 9879–9883, 2012.

[43] Q. Wang, T. R. Chan, R. Hilgraf, V. V Fokin, K. B. Sharpless, and M. G. Finn, “Bioconjugation by Copper (I) -Catalyzed Azide-Alkyne [3 + 2] Cycloaddition,” no. I, pp. 3192–3193, 2003.

[44] C. A. Cycloaddition, L. Li, and Z. Zhang, “Development and Applications of the,” no. Figure 1, pp. 1–22, 2016.

[45] A. Manuscript, “NIH Public Access,” vol. 39, no. 4, pp. 1272–1279, 2010.

[46] J. M. Baskin *et al.*, “Copper-free click chemistry for dynamic in vivo imaging,” no. 11, 2007.

[47] V. A. Online, K. Kalidasan, Y. Su, X. Wu, S. Q. Yao, and M. Uttamchandani, “Fluorescence-activated cell sorting and directed evolution of a - N - acetylgalactosaminidases using a quenched activity-based probe (q ABP) †,” no. C, 2013.

[48] S. Lim, S. P. S. Chundawat, and B. G. Fox, “Expression, purification and characterization of a functional carbohydrate-binding module from Streptomyces sp. SirexAA-E,” *Protein Expr. Purif.*, vol. 98, pp. 1–9, Jun. 2014.

[49] B. Cobucci-Ponzano *et al.*, “beta-Glycosyl azides as substrates for alpha-glycosynthases: preparation of efficient alpha-L-fucosynthases.,” *Chem. Biol.*, vol. 16, no. 10, pp. 1097–108, 2009.

[50] D. Gao, S. P. S. Chundawat, C. Krishnan, V. Balan, and B. E. Dale, “Mixture optimization of six core glycosyl hydrolases for maximizing saccharification of ammonia fiber expansion (AFEX) pretreated corn stover,” *Bioresour. Technol.*, vol. 101, no. 8, pp. 2770–2781, Apr. 2010.

[51] G. L. Miller, “Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar,” *Anal. Chem.*, vol. 31, no. 3, pp. 426–428, Mar. 1959.

[52] D. L. Zechel, S. P. Reid, D. Stoll, O. Nashiru, R. A. J. Warren, and S. G. Withers, “Mechanism, Mutagenesis, and Chemical Rescue of a β-Mannosidase from Cellulomonas fimi †,” *Biochemistry*, vol. 42, no. 23, pp. 7195–7204, Jun. 2003.

[53] K. Horisawa, “Specific and quantitative labeling of biomolecules using click chemistry,” *Front. Physiol.*, vol. 5, Nov. 2014.

[54] X. Zhang and Y. Zhang, “Applications of Azide-Based Bioorthogonal Click Chemistry in Glycobiology,” *Molecules*, vol. 18, no. 6, pp. 7145–7159, Jun. 2013.

[55] M. Bjerknes, H. Cheng, C. D. McNitt, and V. V. Popik, “Facile Quenching and Spatial Patterning of Cylooctynes via Strain-Promoted Alkyne-Azide Cycloaddition of Inorganic Azides,” *Bioconjugate Chemistry*, vol. 28, no. 5. pp. 1560–1565, 2017.

[56] M. Bjerknes, H. Cheng, C. D. McNitt, and V. V. Popik, “Facile Quenching and Spatial Patterning of Cylooctynes via Strain-Promoted Alkyne-Azide Cycloaddition of Inorganic Azides,” *Bioconjug. Chem.*, vol. 28, no. 5, pp. 1560–1565, 2017.

[57] T. W. J. Gadella, “FRET microscopy : from principle to routine technology in cell,” vol. 241, no. May 2010, pp. 111–118, 2011.

[58] S. Planning, “Random Mutagenesis by PCR,” pp. 1–9, 2000.

[59] T. H. Jukes, “Transitions , Transversions , and the Molecular Evolutionary Clock,” pp. 2–3, 1987.