MIE 678: Integrated Methods in Protein Biochemistry

High-throughput screening of glycosynthases using azido sugars for oligosaccharides synthesis

Mohit Kumar[a,†], Chandra Kanth Bandi[a,†], Shishir P. S. Chundawat[a,*]

[a]Department of Chemical and Biochemical Engineering, Rutgers-The State University of New Jersey, Piscataway, New Jersey 08854, USA

*Corresponding author: shishir.chundawat@rutgers.edu

[†]Both authors contributed equally

Abstract: Glycosynthases are mutant glycosyl hydrolases that can synthesize glycosidic bonds between acceptor glycone/aglycone groups and activated donor sugars with suitable leaving groups (e.g., azido, fluoro). However, it has been challenging to rapidly detect glycosynthase reaction products involving azido sugars as donor sugars. This has limited our ability to apply rational engineering and directed evolution methods to rapidly screen for improved glycosynthases that are capable of synthesizing bespoke glycans**.** Here, we outline our recently developed screening methodologies for rapidly detecting glycosynthase activity using a model fucosynthase enzyme engineered to be active on fucosyl azide donor sugar. We created a diverse library of fucosynthase mutants using semi-random and random error prone mutagenesis and then identified improved fucosynthase mutants with desired activity using two distinct screening methods developed by our group to detect glycosynthase activity (i.e., by detecting azide formed upon completion of fucosynthase reaction); a) pCyn-GFP regulon method, and b) Click chemistry method. Finally, we provide some proof-of-concept results illustrating the utility of both these screening methods to rapidly detect products of glycosynthase reactions involving azido sugars as donor groups in general.

Keywords: Carbohydrate-Active Enzymes, Glycosyl Hydrolase, Glycosynthase, Glycans, Glycoconjugates, Human Milk Oligosaccharides, Directed Evolution, Rational Protein Engineering, High-Throughput Screening, Enzymology, Carbohydrates

1. Introduction

Proteins, nucleic acids, lipids, and glycans are the four major biomolecular components that make up all cellular systems (Varki & Gagneux, 2015). Among these, glycans are the most abundant biomolecules on the planet. Glycans are formed when monosaccharides form glycosidic bond linkages with other simple glycone groups (e.g., polysaccharides), or with other aglycone groups like proteins (e.g., glycoproteins), lipids (e.g., glycolipids), or nucleic acids (e.g., glycoRNAs) to form complex glycoconjugates. Glycans serve diverse metabolic, structural, and functional roles in cellular systems such as host cell-pathogen interactions, cell-cell communication, and cellular metabolism (**Figure 1**). In living systems, importance of glycans spans a range of complex events

in the overall cellular/organismal life cycle, starting from growth, development, functioning, survival, and finally death (Flynn et al., 2021a; Varki et al., 2022). Additionally, almost all cell surfaces display complex and diverse glycan profiles, which are recognized by most infectious agents and further mediate interactions between hosts and pathogens (Fan et al., 2022; Kyselova et al., 2008). Patterns of cellular glycosylation via modification of proteins and lipids has been shown to be directly correlated with many diseases (Reily et al., 2019). Recent advancements in the field of glycosciences including the discovery of glycosylated RNAs provides glimpses into the structural heterogeneity and functional complexity of glycans in living systems (Flynn et al., 2021b). There has been growing interest in drugs that target specific glycoproteins & glycan-active enzymes involved in host cell recognition, infection, and replication (Adamczyk et al., 2012; Dube & Bertozzi, 2005; Y. Li et al., 2021).

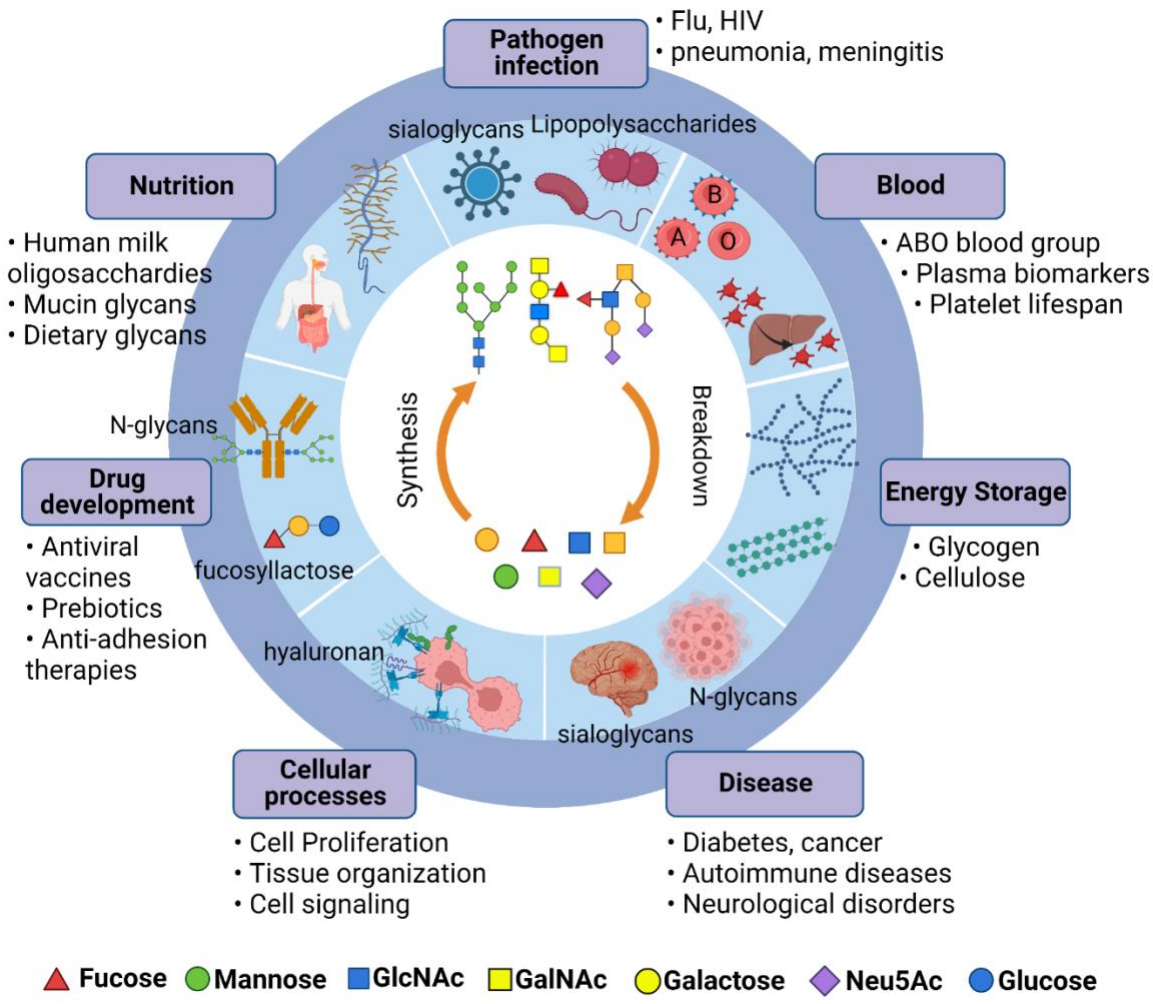

Figure 1. Importance of glycans spans from cellular process regulation to drug development. Here we illustrate broader role of glycans in diverse designated categories where glycans are either synthesized, broken down, and/or recognized by different enzymes in living systems.

Albeit the importance of glycans has been widely understood, their study has lagged behind compared to protein and nucleic acid focused research primarily because synthesis and modifications of glycans are often not practical in most research laboratories (Council, 2012; Lauc et al., 2014). Chemical synthesis has been successful, to an extent, in synthesizing defined oligosaccharides but at high operating costs and with poor product specificity (Plante et al., 2001;

Wang et al., 2013). Alternatively, enzymatic synthesis methods offer an alternative economical and highly selective approach for glycans synthesis (Nilsson, 1988). The enzymes that are directly involved in the synthesis, modification, and/or breakdown of glycans and glycoconjugates are classified in the Carbohydrate Active enZyme (CAZy) database (Cantarel et al., 2009; Park et al., 2010). As of June 30, 2022, 173 Glycoside Hydrolase (GH) families, 115 Glycosyltransferase (GT) families, 42 Polysaccharide Lyase (PL) families, 20 Carbohydrate Esterase (CE) families, and 17 Auxiliary Activity families were listed on the CAZy database.

Glycosyltransferases (GTs) are enzymes that natively catalyze the synthesis of glycans and glycoconjugates in biological systems using nucleotide sugars as activated donor groups. But the poor economic viability of using expensive nucleotide substrates for *in vitro* reaction, expression challenges associated with producing GTs in *E. coli*, and limited stability of membrane-associated GTs are often the common concerns that have limited use of GTs for glycans synthesis (Boltje et al., 2009; Weijers et al., 2008). Advancement towards improvement of protein expression and incorporation of glycan synthesis reaction networks into cellular metabolic pathways is under progress and can address some of these challenges (McArthur & Chen, 2016). Glycosyl hydrolases (GHs), on the other hand, are widely used, are easy to express in a variety of microbial expressions systems like *E. coli* and have been well characterized in the literature and CAZy database (Sathya & Khan, 2014; Shaikh & Withers, 2008). GHs primarily hydrolyze glycosidic bonds in glycans and glycoconjugates using either a retention or inversion mechanism based on the anomeric configuration of substrates reacted and products formed. While most GH enzymes hydrolyze their substrates, their innate reaction mechanism sometimes allows for synthesis of glycosidic

bonds under certain conditions via the transglycosylation reaction mechanism (Saumonneau et al., 2016). Various sequence and structural comparisons between native glycosyl hydrolases and transglycosidases (TGs) have revealed minor changes in the enzyme active site that often facilitates efficient synthesis of glycosidic linkages (Zeuner et al., 2018). These observations have guided researchers to systematically engineer GHs into TGs. Among the various engineering strategies applied, mutating the GH active site nucleophile site to facilitate transglycosylation reaction using activated donor sugars with suitable leaving groups, often mimicking the glycosyl-enzyme intermediate structure, has given rise to a new class of mutant GH enzymes called glycosynthases (Mackenzie et al., 1998).

Glycosynthases (GSs) are mutant glycosyl hydrolases that catalyze the synthesis of glycosidic bonds between an activated donor sugar (with a suitable leaving group like fluoride or azide. On the anomeric carbon) and an acceptor sugar or aglycone group. The catalytic nucleophile residue is mutated to a smaller amino acid such as alanine, glycine, or serine which makes the mutant hydrolytically inactive on glycosidic linkages (Ducros et al., 2003). When some of these mutant GHs are reacted with an activated donor sugar, the donor substrate docks within the enzyme active site and mimics the intermediate enzyme-substrate or Michaelis complex of the native GH. In the next step, an acceptor molecule docks within the active site facilitating the synthesis of a glycosidic bond between the donor and acceptor groups upon removal of the leaving group (Hayes & Pietruszka, 2017). In the absence of the nucleophile residue, the GS is unable to then hydrolyze the formed glycan product, and hence overcoming a major limitation of using GHs or TGs for glycans synthesis. This strategy has been successfully applied to engineer numerous

glycosyl hydrolases to their respective glycosynthases. Along with mutating the nucleophile site, making additional active site mutations has yielded engineered enzymes with improved specific activity and diverse acceptor group specificity (Cobucci-Ponzano et al., 2011). The ease of handling glycosyl hydrolases, established database of identified sequences and structures, and availability of structure-guided rational protein engineering strategies has made glycosynthase engineering a very promising approach for glycans synthesis (Zeuner et al., 2018). However, one of the major challenges has been the lack of available high-throughput screening methods that can rapidly monitor *in vitro* or *in vivo* glycosynthase activity to enable implementation of directed evolution strategies for engineering improved GSs (Ben-David et al., 2008a). To address this issue, limited number of screening techniques have been developed to screen glycosynthase in two decades specially for fluoride-based donor sugars (Andrés et al., 2014; Ben-David et al., 2008b; Hayes et al., 2017). Ayushi and co-workers from our lab recently developed fluorescence activated cell sorting (FACS) based high throughput screening method for mutant glycosynthases active on azido sugars and used directed evolution to evolve a model GS (Agrawal et al., 2021a). Furthermore, Chandra and co-workers from our lab have recently also developed an azide biosensor that can be implemented to detect GS activity on azido sugars (Bandi et al., 2021a).

Here, we demonstrate application of these two screening methods for azido sugar active GS enzymes and focus on demonstrating how a model GH29 family glycosyl hydrolase can be engineered and evolved into an active glycosynthase. GH29 enzymes are often α-fucosidases that use fucosylated oligosaccharides/polysaccharides as substrates for hydrolysis into constituent monomers using a retention mechanism. An α-L-fucosidase from *Thermotoga maritima (TmAfc),*

belonging to the GH29 family has been engineered to a transfucosidase (Klontz et al., 2020) and fucosynthase (Burgin & Mayes, 2019). However, the fucosynthase reaction efficiency is low, with limited acceptor sugar specificity, and therefore has room for further improvement. The structure of this enzyme has been solved and a detailed QM/MM simulation has been completed to understand the fucosynthase reaction mechanism. In its native hexamer form, this enzyme displays two domain folds, consisting of a $(\beta/\alpha)_8$ TIM-like domain and a C-terminal β-sandwich domain. The catalytic nucleophile of TmAfc is Aspartic acid (D224) and the bronsted acid/base residue is Glutamic acid (E266). The catalytic nucleophile mutant, TmAfc_D224G was shown to be hydrolytically inactive but was shown to function as a glycosynthase (i.e., fucosynthase) in the presence of β-fucosyl-azide (as donor azido sugar) and pNP-Xylose (as acceptor sugar). An azide ion is released as a by-product of the fucosynthase reaction and therefore quantifying the amount of azide ion released is indicative of the extent of fucosynthase reaction. Two distinct *in-vivo* azide quantification methods were developed and applied to quantify the amount of azide released for any universal glycosynthase reaction involving azido sugars (Agrawal et al., 2021b) (Bandi et al., 2021a). Our first method was an operon-based biosensor where the azide ion specifically induced Green Fluorescent Protein (GFP) expression in *E. coli* cells. While our second method was based on a click chemistry approach where the azide ion was reacted with a fluorescent cycloalkyne to quench the probe fluorescence intensity thus facilitating selective azide monitoring.

The purpose of this chapter is to describe how researchers can apply either the click chemistry and/or operon biosensor approaches for engineering *Thermotoga maritima* α-L-Fucosidase

(TmAfc) into an efficient fucosynthase. All procedures have been delineated, starting from DNA mutagenesis to enzyme activity assays, so that researchers interested can easily replicate these workflows. Specifically, we use two error prone PCR (epPCR) mutagenesis strategies to create semi-random and randomly targeted mutations in the *TmAfc_D224G* gene, which is the baseline GS mutant identified from earlier site directed mutagenesis studies. The mutant GS constructs were then transformed into *E. coli* cells and mutant protein expression was induced followed by the chemical rescue and glycosynthase based activity assays. The glycosynthase assays were measured using both of our reported azide detection methods. The constructs that showed altered activity can be then sequenced and corresponding mutations identified to gain a fundamental understanding of the structure-function relationships driving glycosynthase activity.

2. Screening using cyn regulon based azide detection method

2.1 Summary

Our regulon based azide biosensor is an engineered *cyn* operon that can induce protein expression in *E. coli* in the presence of azide ions. The native *cyn* operon in *E. coli* is inducible by cyanate ion. Since cyanate and azide ions are structurally homologous, the *cyn* operon was engineered to be selectively induced by azide. This novel azide based promoter system was cloned upstream of GFP gene to generate the plasmid called pCyn_v2_GFP. The resultant plasmid was recently used by our group to demonstrate the tunable expression of a model green fluorescence protein (GFP) (Bandi et al., 2021b). Azide ion is released as a by-product of the fucosynthase reaction between β-fucosyl azide and pNP-Xylose using *TmAfc_D224G* mutant.

Therefore, fucosynthase activity will be evaluated by the monitoring amount of azide released as estimated using *E. coli* cells containing the plasmid *pCyn_v2_GFP* as shown in **Figure 2**.

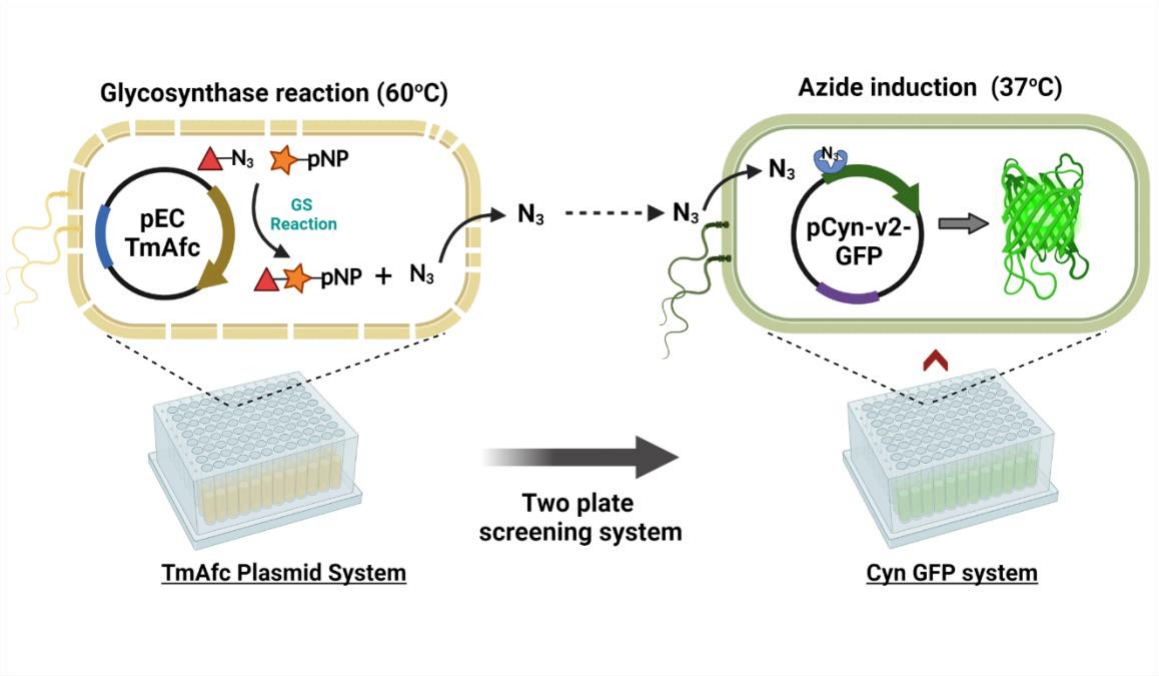

Figure 2. Schematic for medium-throughput screening of glycosynthase enzymes using azide biosensor. Cells containing *TmAfc* mutant genes are expressed and glycosynthase reactions are performed in a 96 well plate at 60°C. Reaction mixtures from *TmAfc* plates are then transferred to plate with exponentially growing pCyn_v2_GFP cells. Azide released from the GS reaction induces GFP expression in pCyn_v2_GFP transformed cells at 37°C. (Red triangle and orange star correspond to fucose and xylose sugars, respectively)

2.2 Preparation of error-prone PCR library

Random mutagenesis is a powerful technique implemented by protein engineers to generate novel mutants and the resultant mutants are then screened via suitable high-throughput selection methods to ultimately allow for directed evolution of proteins with desired functional properties. There are different methods to create such genetic diversity in the case of random mutagenesis; chemical mutagenesis, through mutator strains and by error prone PCR

mutagenesis (Labrou, 2010). The most commonly used epPCR method takes advantage of intrinsically low fidelity of Taq DNA polymerase, which gets further decreased by addition of Mn^{2+}, increasing the concentration of Mg^{2+} concentration, and using unequal dNTP concentration. This method is preferred choice when a certain segment of DNA, which is too long to be chemically synthesized as degenerate sequences (Wilson & Keefe, 2001). Creating high quality library of random sequences is an important step in the process of directed evolution and these libraries are then screened for individual molecules with desired phenotypes (McCullum et al., 2010). This process of generating mutants is a tradeoff between maximum diverse mutational spectrum of variants and the loss of critical function of gene. In order to increase the beneficial mutations, casting epPCR (cepPCR) was done where our gene of interest, *TmAfc_D224G* was divided into three fragments and each fragment was subjected to epPCR followed by ligation of fragment with the vector followed by transformation into cells (Yang et al., 2017). This mutant library was then screened using our in-house microplate based pCynGFP sensor medium-throughput screening method as described below.

2.2.1 Materials and Equipment

- 2X Phusion Master Mix (Thermo Fisher; Catalog Number: F531S)
- Taq DNA polymerase (NEB; Catalog Number: M0273L)
- Gibson Assembly Master Mix (NEB; Catalog Number: E2611S)
- Gel Extraction & PCR purification kit (IBI Scientific; Catalog Number: IB47010)
- dNTPs (NEB; Catalog Number: N0447L)
- Dpn1 enzyme (NEB; Catalog Number: R0176L)

- $MnCl_2$ (Thermo Fisher; Catalog Number: AC193451000)
- $MgCl_2$ (ACS Organics; Catalog Number: 223211000)
- Forward and Reverse Primers (IDT DNA)
- Template TmAfc-D224G DNA (Prepared in-house, see (Bandi et al., 2021b) for details)
- E.cloni 10g competent cells (NEB; Catalog Number: C2987H)
- SOC media (NEB; Catalog Number: B9035)
- PCR Thermocycler (Eppendorf)
- Electroporator (Eppendorf)
- GelDoc EZ gel imager (Bio-Rad)

2.2.2 Procedure:

1. The three fragments of gene (241 bp, 291 bp, 276 bp) was chosen in this case based on a rational engineering approach where the residues chosen in each gene fragment was within 7.5 Å distance from active site in addition to specific recommendations from our collaborators who performed extensive computational studies including QM/MM simulations (Burgin & Mayes, 2019).

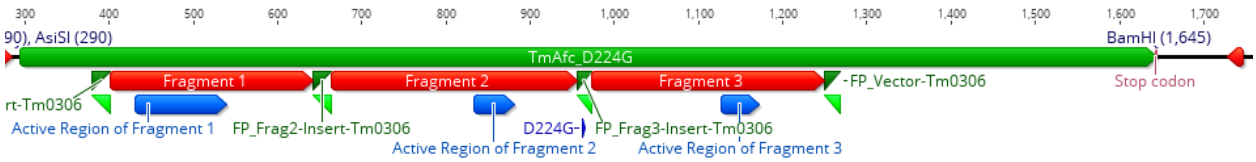

Figure 3: Semi-random mutagenesis of three gene fragments (red color) that are part of gene of interest TmAfc_D224G (green color). Corresponding forward and reverse primers for three fragments and the vector for molecular cloning are shown here as well.

2. The three gene fragments were then subjected to random mutagenesis in Insert PCR and cloned into an expression vector for Gibson assembly as per the standard protocol (Gibson et al., 2009). Run the Vector PCR and Insert PCR as directed in **Table 7**. (Note: *Consider the GC content or secondary structure, for high GC content or secondary structure, use higher denaturing temperatures (94–98°C) or cosolvents such as DMSO).*

Table 1. Vector and Insert fragment PCR reaction mixture and conditions

Vector PCR reaction mixture	**Vector PCR conditions**
• 1X Phusion Master Mix • 0.5 µM Forward and Reverse Primer • 20 ng TmAfc-D224G plasmid DNA **Total volume = 50 µL**	• Initial denaturation: 98°C/30 sec • [c]Denaturation: 98°C/10 sec • [c]Annealing: 60°C/10 sec • [c]Extension: 72°C/3 min • Final Extension: 72°C/5 min • Hold: 4°C c - Number of cycles: 30
Insert PCR reaction mixture (All fragments)	**Insert PCR conditions**
• 1.25 U Taq Master Mix • 1x Taq buffer • 0.5 µM Forward and Reverse Primer • 20 ng TmAfc-D224G plasmid DNA • 0.2 mM dATP and dGTP • 1 mM dCTP and dTTP • 1 mM $MnCl_2$ **Total volume = 50 µL for each insert**	• Initial denaturation: 95°C/60 sec • [c]Denaturation: 95°C/30 sec • [c]Annealing: 60°C/30 sec • [c]Extension: 68°C/3 min • Final Extension: 68°C/5 min • Hold: 4°C c - Number of cycles: 35

3. DNA gel electrophoresis after PCR Amplification: After PCR reaction is complete, use 5 µL of PCR product and mix with 1 µL of DNA loading dye. Add the resultant mixture to 1% agarose gel for insert and 0.7% agarose gel for vector along with appropriate DNA ladder for insert and vector DNA analysis. Run gel electrophoresis for 120 V for 45 mins and image the gel using an imager and verify the desired fragments in the gel based on sizes

for all three fragments of insert and vector. *(Note: At this step, if there are significant impurities in PCR mixture, whole PCR mixture could be loaded in fresh agarose gel and further extracted through gel extraction kit.)*

4. Dpn1 digestion, PCR product purification, and DNA concentration measurements:
    a. Perform Dpn1 digestion to cleave template of methylated DNA with remaining 45 µL PCR products using 1 µL of Dpn1, 2 µL of CutSmart buffer, 2 µL of PCR water at 37°C for 2 hr. (*Note: The preparation of DNA to be cleaved should be free of contaminants such as phenol, chloroform, alcohol, EDTA, detergents, or excessive salts, all of which can interfere with restriction enzyme activity*)
    b. After Dpn1 digestion, PCR product is purified using a standard cleanup kit from IBI scientific based on the manufacturers protocol to remove enzymes, salts, and other impurities. PCR purification kit from any vendor can be used. *(Note: Eluting bound DNA in pure water increases cloning efficiency).*
    c. To get an accurate measurement of extracted PCR product concentration, prepare and run the following mixture: 2 µL of purified PCR product; 3 µL of PCR water, and 1 µL of DNA loading dye alongside a 5 µL DNA ladder and run gel electrophoresis at 120 V for 55 mins in 0.7% agarose gel.
    d. Image the gel in GelDoc EZ imager and estimate the concentration of purified PCR products using the band intensity with respect to the standards in DNA ladder.
5. Gibson assembly and cell transformation:
    a. Prepare the reaction mixture for Gibson assembly step as per NEB protocol (https://nebiocalculator.neb.com/#!/ligation), using 100 ng of vector DNA; 1:10 molar

ratio of insert DNA; 10 µL of Gibson master mix along with the positive control supplied by vendor. The total reaction volume can vary based on DNA concentration. (<u>Note</u>*: Make sure to prepare this composition mixture from given manufacturer tool online and perform variation in insert: vector ratio to get optimum yield*).

b. Incubate samples in a thermocycler at 50°C for 15 minutes when 2 or 3 fragments are being assembled/60 minutes or when 4-6 fragments are being assembled (we did reaction for 60 mins since we have four fragments). Following incubation, store samples on ice or at -20°C for subsequent transformation if needed. (<u>Note</u>*: Count vector as a fragment when counting total fragments and incubating for higher time tends to give better results in general*).

c. Transform the entire Gibson reaction mixture into E. cloni 10g chemically competent cells from NEB. (<u>Note</u>*: Before going for Gibson with all insert and vectors DNA available, one can also run small scale Gibson reaction and send transformed colonies for sequencing to check on diversity of library. Once confirmed, one can run larger scale Gibson assembly reactions and collect all possible DNA mutants in library after plasmid extraction step*).

d. Add recovery media (SOC media) to transformed cells and incubate at 37°C for 2 hrs. Plate 50 µL-100 µL of mixture on large LB agar plate with 50 µg/mL Kanamycin and incubate at 37°C for 16 hours. The colonies obtained here can be used for evaluating the diversity and mutation rate of the epPCR library.

e. Plate the remaining mixture in multiple large LB agar plates to get a mat of colonies. All colonies can be scrapped using a cell scrapper and pooled into one tube. From this,

glycerol stocks could be prepared for future use, alternatively one can do mini prep with all cells and pool all plasmid DNA for storage after measuring concentration. This DNA library can be used anytime in the future for cell transformation to generate mutant library as per requirements. In our case, transformation was done from this stored library for the next step of the medium-throughput screening method.

6. epPCR test library sequencing:
   a. Pick 20 individual colonies from step 5d and grow them in 7 ml LB media with 50 µg/mL Kanamycin for 37°C for 16 hrs.
   b. Perform plasmid extraction and send the samples to DNA sequencing facility (Genewiz, NJ in our case) for DNA sequencing using primers specific to the gene of interest.
   c. The results from DNA sequencing can be analyzed and tabulated. The table below highlights all the point mutations obtained in our epPCR test library.

Table 2. Matrix of point mutations identified in TmAfc_D224G after epPCR

		Mutation to			
		0.5 mM Mn^{+2}			
Mutation from		**T**	**C**	**A**	**G**
	T	0	28	15	4
	C	6	0	5	0
	A	25	2	0	32
	G	8	1	23	0
Total mutations		149			

d. Using the matrix table data, calculate the overall mutation rate to estimate biases in mutation spectrum. Detailed information about analysis of the mutation spectrum and biases can be found elsewhere (Hanson-Manful & Patrick, 2013).

Table 3. Mutational spectrum of the TmAfc_D224G casting error-prone PCR library

	0.5 mM Mn^{+2}	
Types of mutation	**Frequency**	**Total Prop**
Transitions		
A → G, T → C	60	40.27%
G → A, C → T	29	19.46%
Transversions		
A → T, T → A	40	26.84%
A → C, T → G	6	4.02%
G → C, C → G	1	0.67%
G → T, C → A	13	8.72%
Summary of bias		
Transitions/transversions		
AT → GC/GC → AT	2.07	
A → N, T → N	106	71.15%
G → N, C → N	43	28.85%
Mutation rate		
Mutations per kb	4.79	
Mutations per TmAfc gene	6.47	

e. Input the information from the table above into PEDEL-AA server (http://guinevere.otago.ac.nz/stats.html), to obtain information about the epPCR

library diversity and average number of amino acid mutations per construct. Detailed information about PEDEL-AA analysis for library diversity can be found elsewhere (Hanson-Manful & Patrick, 2013).

Table 4. Summary of epPCR library characteristics predicted by PEDEL-AA analysis

Summary of library characteristics	0.5mM Mn^{+2}
Total library size	1.00E+06
Number of variants with no indels or stop codons	7.74E+05
Mean number of amino acid substitutions per variant	4.5
Non-mutated (wild type) sequences (% of library; PCR est.)	1.28%
Number of distinct full-length proteins in the library	7.15E+05
Number of distinct full-length proteins in the library	7.01E+05

2.2.3 Notes

1. The mutational frequency is the product of DNA polymerase error rate and number of duplications during PCR, a reasonable high rate could be achieved by varying the concentration of $MnCl_2$ and $MgCl_2$, and low amount of DNA template (i.e., as more times the target gets replicated, greater the error accumulated).
2. Generally, commercially available chemically competent cells give higher yield compared to in-house prepared cells, and if electroporation is being done, it should be done in chilled cuvettes for better transformation efficiency.
3. The primers designed for error prone PCR should be of similar melting range and it should not be very high as it will reduce false priming and therefore also ensure complete denaturation of extended primers (Daugherty et al., 2000).

2.3 Procedure for medium-throughput screening of mutants

Once the diversity and mutational frequency in the epPCR library is estimated, the next step is to screen the library to identify novel mutants with increased or decreased GS activity. With azide operon biosensor, we will screen for fucosynthase activity to identify potential targets of interest from the epPCR library using a plate-based medium-throughput screening approach.

2.3.1 Materials and Equipment

- β-fucosyl azide (Synthose; Catalog Number: FF757)
- pNP-Xylose (Carbosynth; Catalog Number: EN03230)
- pNP-Fucose (Carbosynth; Catalog Number: EN06434)
- 2 mL Deep 96 well plates (USA Scientific; Catalog Number: 1896-2000)
- Transparent 96 well plates (USA Scientific; Catalog Number: 5665-5101)
- Clear bottom opaque 96 well plates (USA Scientific; Catalog Number: 5665-5096)
- 96 well PCR Plates (Ultracruz; Catalog Number: 205891)
- 96 well U-bottom microplate (Bio-one; Catalog Number: 650201)
- SDS-PAGE protein gels (Genscript; Catalog Number: M42012)
- Mini Protean Tetra system (Bio-Rad)
- TLC plates (Analtech; Catalog Number: 21521) and TLC glass chamber
- B-PER Bacterial Protein Extraction Reagent (Thermo Fisher; Catalog Number: 78243)
- Multichannel Pipettes (Eppendorf) or Liquid Handling Pipetting System (Beckman)

- UV-Spectrophotometer (Molecular Devices, SpectraMax M5e)
- Cell culture incubator/shakers (Eppendorf, Innova)

2.3.2 Procedure

1. Azide-GFP calibration curve development: To quantify the amount of azide released (which is proportional to the extent of GS reaction), a calibration curve between azide concentration and GFP expression in *E. coli* cells is necessary. The curve can also determine the linear operating range of the *Cyn GFP* biosensor for azide detection using GFP fluorescence.
    a. Inoculate *E. coli* cells containing the pCyn_v2_GFP plasmid in 5 mL LB media with carbenicillin (100 µg/mL) in culture tubes and incubate at 37°C for 14-16 hr.
    b. Subculture 2% v/v of overnight grown culture in 50 mL LB media with carbenicillin (100 µg/mL) in 250 mL flasks and incubate at 37°C for 1.5-3 hr till the OD_{600} of the culture reaches 0.3-0.4 (*Note: Typically, it takes 2-2.5 hours after inoculation.*)
    c. Meanwhile, prepare the stock solutions of sodium azide (NaN_3) in DI water at varying concentrations ranging from 0.1 mM to 10 mM. (*Note: In this study we chose the following concentrations: 0.1 mM, 0.5 mM, 1 mM, 1.5 mM, 5 mM, 7.5 mM, and 10 mM. It is necessary to choose at least 5 concentrations for building a robust calibration curve that spans a broad range of azide concentrations relevant to the GS reaction.*)
    d. Once the OD_{600} of the culture reaches the desired value, transfer 450 µL of cells to 24 wells in a 2 mL deep well plate. Add 50 µL of the stock azide standards to the cells in triplicates. The remaining three wells serve as no inducer controls (0 mM azide or

buffer control added). (*Note: For measuring the glycosynthase reaction activity, 50 µL of glycosynthase reaction mixture will be added to cells instead of azide standards.*)

e. Cover the deep well plate with breathable film and incubate for 4 hours at 37 °C to allow for GFP expression. (*Note: The amount of time to incubate can be varied to increase or decrease the sensitivity of calibration curve*).

f. Sample out 200 µL of cells after 4 hours and transfer to a clear bottom black opaque microplates to measure the GFP fluorescence and OD_{600}.

g. Measure the fluorescence of cells at 488 nm excitation and 525 nm emission with a cutoff wavelength at 515 nm. Additionally, measure the OD_{600} of the samples.

h. Normalize the GFP fluorescence with the OD_{600} values of the cells and plot the normalized fluorescence against the effective azide concentration in the wells (***Figure 4***). Fit a four parameter logistic model to the plotted data to estimate the unknown parameters and obtain the calibration curve equation. The equation will be used for estimating the amount of azide present in unknown sample to evaluate the extent of glycosynthase reaction (O'Connell et al., 1993) (*Note: Here, a non-linear regression model was fitted to the data*. *The four parameter logistic model is a more suitable model for biological assays such as ELISA and dose response. If a linear fit is desired, lower induction concentrations must be used which decreases the operating range for the biosensors or introduces additional dilution steps to maintain lower concentrations.*)

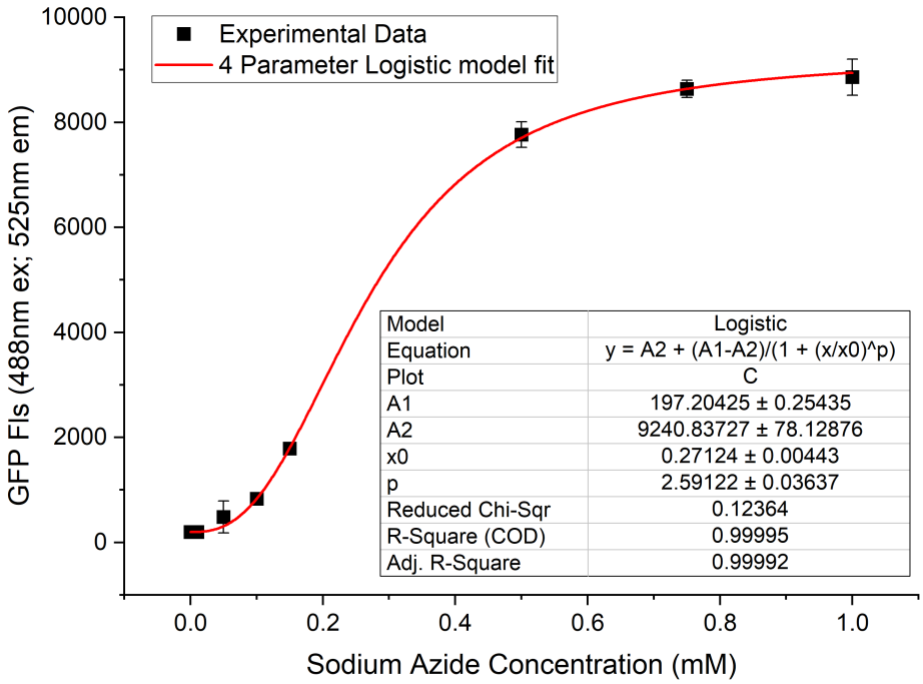

Figure 4. Calibration curve for determining unknown azide concentration in reaction wells and corresponding GFP fluorescence of cells induced in presence of added azide. Here, the sodium azide concentration is the equivalent concentration of azide in the GS reaction deep well plate. The GFP fluorescence values are normalized with the respective OD_{600} of the samples to account for cell death due to azide toxicity (particularly at higher azide concentrations). Data is collected for three biological replicates.

2. **Chemical rescue and glycosynthase reaction measurement for mutants**

   a. Pick and inoculate 84 random colonies after transformation of epPCR library along with four controls in 1 mL LB media with Kanamycin (50µg/mL) in 2 mL 96 deep well plate. Incubate the deep well plate overnight for 14-16 hrs at 37°C. The four controls are *E. coli* cells without any plasmid, *E. coli* with TmAfc_Wt plasmid (negative control as wild-type GH), *E. coli* with TmAfc_D224G plasmid (positive control as D224G GS mutant), and LB media + Antibiotic media background. The controls are present in triplicates in Row A of the plate and 84 mutants are present are single replicates in Rows B to H. (*Note: The E. coli cells with TmAfc_Wt are negative control; E. coli cells*

with TmAfc_D224G are positive glycosynthase control; E. coli cells without plasmid are negative GFP fluorescence control.)

b. Centrifuge the deep well plate at 3900 rpm for 15 min and discard the supernatant by inverting the plate.

c. Resuspend the cell pellet in the deep well plate in 200 µL of autoclaved DI water. (*Note: The resuspension volume of DI water can be adjusted based on the total amount of cells needed for glycosynthase reaction and chemical rescue reactions*)

d. To test the expression of TmAfc proteins in the resuspended cells, we run protein gel electrophoresis using the control samples. Run the lysed cell extracts of the control *E. coli* cells without plasmid, *E. coli* with TmAfc_Wt, *E. coli* with TmAfc_D224G through the SDS-PAGE gels and image the gel using Coomassie blue staining (**Figure 5A**).

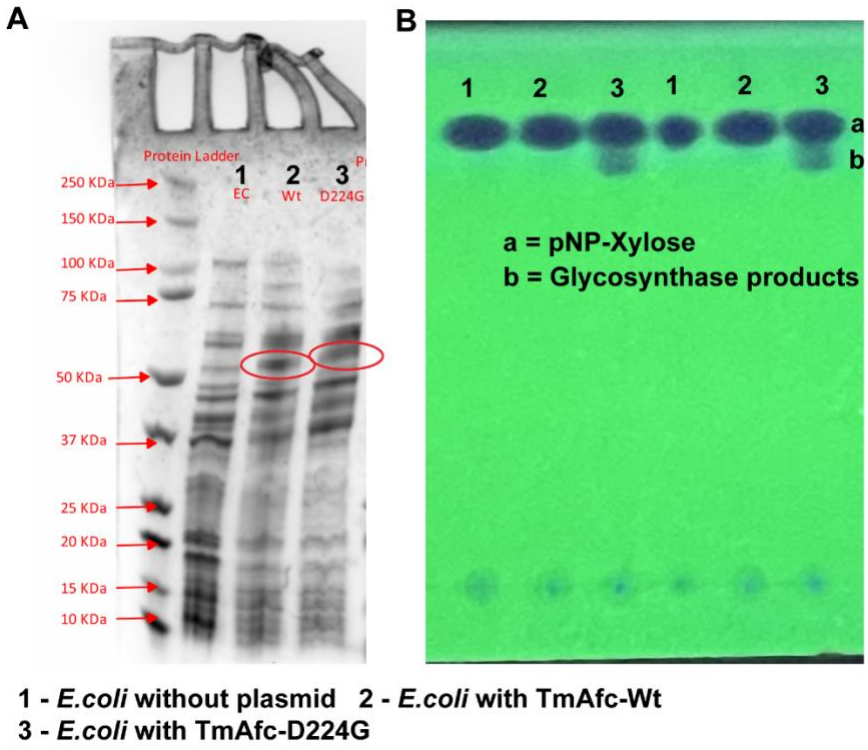

Figure 5. A) SDS-PAGE of control cell lysates (10 µL extract mixed with 10 µL 2x laemmli buffer: beta-mercaptoethanol (95:5; v/v) and resultant 10 µL was loaded in gel to run at 200 V for 30 mins in Bio-Rad Mini-Protean tetra system. **B) TLC plate for GS reaction product detection:** 2 µL of GS products were loaded in TLC plate (Analtech) and run with mobile phase composition- Ethyl acetate : Methanol: Water : Acetic Acid (4 : 2 : 1 : 0.1) in TLC chamber and was visualized under UV light after drying.

e. Prepare three plates each for chemical rescue (CR) reaction and glycosynthase reaction using the **Table 5** template in a 96 well PCR plate and 96 well U-bottom microplates respectively. (*Note: The resuspended cells from one plate are split into three CR plates and three GS plates for replicates*)

Table 5: Chemical rescue and glycosynthase reaction plate setup.

Chemical Rescue Reaction plate		**Glycosynthase Reaction plate**	
Reagents (Stock Concentration)	**Amount used for hydrolysis in µL (Well Concentration)**	**Reagents (Stock Concentration)**	**Amount used for hydrolysis in µL (Well Concentration)**

pNP-Fucose (25 mM)	6 (2 mM)	pNP-Xylose (100 mM)	50 (50 mM)
Sodium Azide (4M)	37.5 (1 M)	Beta-Fucosyl-Azide (100 mM)	10 (10 mM)
pEC-TmAfc cells	12.5	pEC-TmAfc cells	25
MES Buffer (1 M, pH - 6.5)	3.75 (50 mM)	MES Buffer (1 M, pH - 6.5)	5 (50 mM)
DI water	15.25	DI water	10
Total Volume	75	Total Volume	100

f. Seal the plates with aluminum film and incubate the CR reaction plate and GS reaction plate at 60°C for 2 hours and 24 hours, respectively.

g. Recover the CR plates after 2 hours of incubation and transfer 20 µL of CR reaction mixture to 180 µL of 0.1M NaOH in a 96 well clear bottom microplate. Measure the absorbance of the resultant mixture at 410 nm in a spectrophotometer. (*Note: Adding NaOH increases the pH of the reaction mixture thereby stopping the CR reaction and converting the released 4-nitrophenol to 4-nitrophenoxide anion*).

h. Recover the GS reaction plate after 24 hours of incubation and add 50 µL of GS reaction mixture to *E. coli* cells containing pCyn_v2_GFP plasmid grown until log phase as indicated in **Section 2.3.2 steps 1a-1d**. (*Note: 50 µL of GS reaction mixture is added instead of 50 µL of azide standards in the step 1d of section 2.3.2*).

i. Similar to Section 2.3.2 steps 1e-1g, incubate the cells at 37°C for 4 hours, pipette 200 µL of cells and measure the GFP fluorescence at 488 nm excitation and 525 nm emission with a cutoff wavelength at 515 nm. Additionally, measure the OD_{600} of the

cells. (*Note: GS reaction products can be analyzed using thin layer chromatography as shown in* ***Figure 5B***)

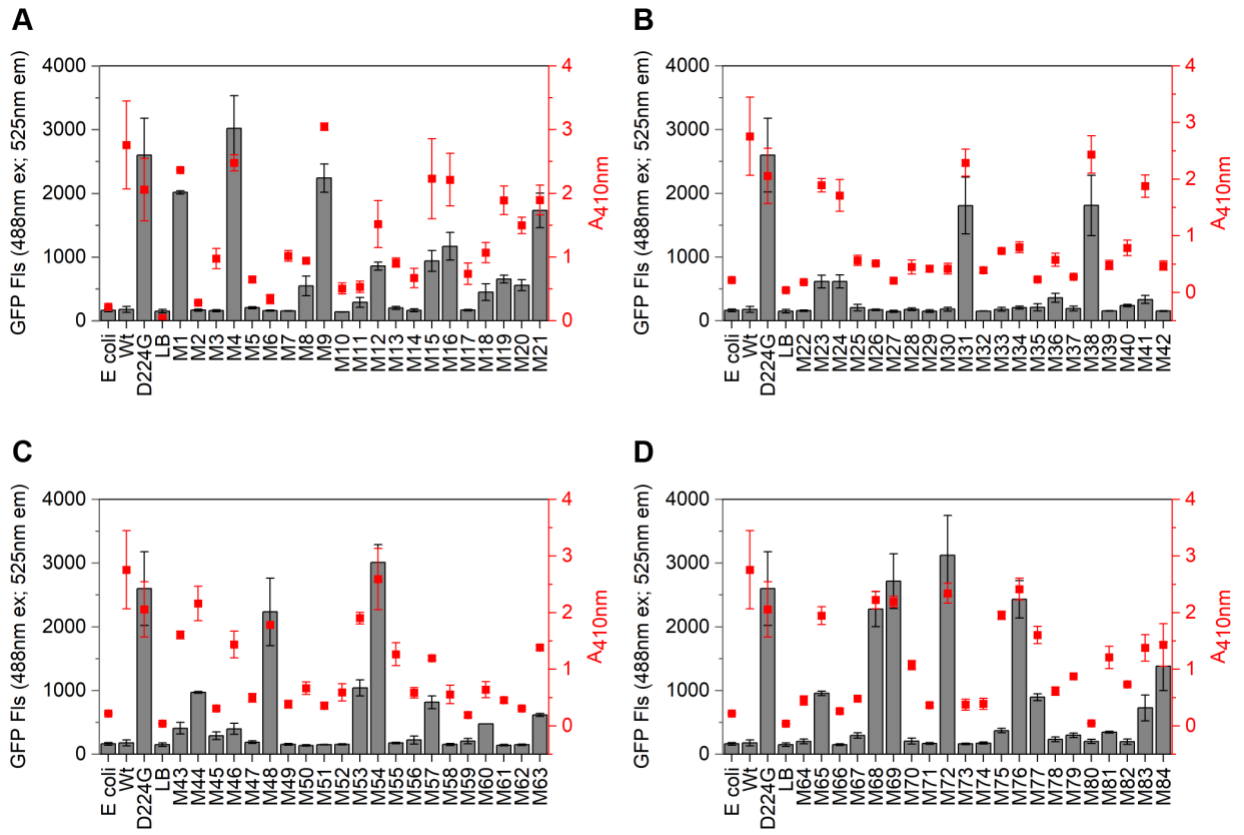

Figure 6. Glycosynthase reaction and Chemical rescue data for random epPCR library mutants. Here GS reaction equivalent azide release based GFP fluorescence values on left y axis (black color) and chemical rescue based pNP absorbance values on right y axis (red color) for four standards (E cloni, TmAfc_Wt, TmAfc_D224G, LB + Kan) and eighty four mutants along with their standard deviations (each reaction was done in triplicate manner).

j. Plot the A_{410nm} from chemical rescue assay results and GFP fluorescence from glycosynthase assay results in single graph as shown in ***Figure 6***. Comparing the chemical rescue and glycosynthase reaction data, the mutants were next classified into four different categories as indicated in **Table 6**. (*Note: The GFP fluorescence from GS reaction can be used to measure the unknown azide concentration using the*

*calibration curve equation in **Figure 4** and used to plot the graph instead of GFP fluorescence. The amount of azide can also be used to predict the amount of substrate converted into GS reaction product*.)

Table 6. Classification of mutants into four categories based on chemical rescue and glycosynthase activity seen in mutant library.

Number	Type	Mutants
Category 1	High GS (>2000), High CR (>1.5)	D224G, M1, M4, M9, M48, M54, M68, M69, M72, M76
Category 2	Low GS (<1000), High CR (>1.5)	M12, M15, M19, M20, M23, M24, M41, M43, M44, M46, M75, M77
Category 3	Low GS (<1000), Low CR (<1)	M3, M5, M7, M8, M13, M14, M17, M18, M25, M26, M28, M29, M30, M32, M33, M34, M35, M36, M37, M39, M40, M42, M49, M50, M51, M52, M56, M58, M60, M61, M62, M64, M67, M70, M73, M74, M78, M78, M79, M82
Category 4	No GS, No CR (Totally inactive)	M2, M5, M6, M10, M11, M22, M27, M35, M37, M45, M59, M66, M80

k. Select the most interesting mutants from each category, isolate the plasmids, and perform sanger sequencing to identify the mutations in each variant.

l. Several of these novel mutants could be used to gain a mechanistic understanding of GS reaction, and ultimately provide data for establishing a machine learning model that can identify a relationship between mutant sequences and their corresponding activity levels for machine learning assisted enzyme directed evolution (Wu et al., 2019).

2.3.3 Notes

1. **Section 2.3.2 step 1b:** The amount of time reaction for the *E. coli* cells to reach an OD_{600} of 0.3-0.4 depends on the inoculum percentage and type of *E. coli* cells used. It is recommended to measure the OD_{600} of the culture at regular intervals to avoid overgrowth.
2. **Section 2.3.2 step 1c:** The upper limit of azide standards concentration was chosen based on the concentration of β-Fucosyl azide substrate added to the glycosynthase reaction mixture. The concentration range can be varied according to the substrate concentration.
3. **Section 2.3.2 step 1d:** The ratio of amount of azide standards or GS reaction mixture to the *E. coli* cells was optimized for detecting 10 mM effective azide concentration which corresponds to 100% substrate conversion.
4. **Section 2.3.2 step 1h**: A non-linear regression model of four parameter logistic model was fit to obtain the calibration curve equation between azide concentration and GFP fluorescence. If a linear fit is desired, lower azide induction concentrations must be used which decreases the operating range of the biosensor or introduces additional dilution steps to maintain lower concentrations.
5. **Section 2.3.2 step 2i:** The GFP fluorescence from GS reaction can be used to measure the unknown azide concentration using the calibration curve equation in **Figure 4** and used to plot the graph instead of GFP fluorescence. The amount of azide can also be used to predict the amount of substrate converted into GS reaction products.

3. Screening using click chemistry based azide detection method

3.1 Summary

When the library size becomes very large (>10^6 mutants), identifying interesting mutants using a 96-well microplate-based assay would be cumbersome and pose logistical challenge. Therefore, it is imperative to have a high-throughput screening technique that can selectively identify mutants in a mixed population of wild type enzyme, and mutants with higher or lower activity than starting template. Here we employ a click chemistry-based fluorescence approach to screen for glycosynthase mutants (Agrawal et al., 2021b). Click chemistry is a reaction between an azide group and a fluorescent cycloalkyne group to form a triazole group (Moses & Moorhouse, 2007). We observed differences in fluorescence intensities of triazole-containing fluorophore products formed upon reaction with released azide ion versus glycosyl azides (Meldal & Tornøe, 2008). The intensity of triazole formed by released azide was significantly lower (or quenched) as compared to triazole formed by β-fucosyl azide. An azide ion is released as a by-product of fucosynthase reaction using *TmAfc_D24G* fucosynthase with β-fucosyl azide and pNP-Xylose substrates. By using the well-known click chemistry technique, we can thus selectively sort cells with varying fluorescence using FACS and screen for improved glycosynthase mutants (with lower fluorescence intensities compared to control) that can synthesize desired fucosylated oligosaccharides (Agrawal et al., 2021b). A greater decrease in fluorescence intensity would indicated a higher concentration of azide in the mixture and this principle serves as the basis of sorting cells using FACS.

3.2 Preparation of error prone PCR library

Random mutagenesis approach can be used to generate a diverse library of mutants based on only the sequence information of the gene of interest. This is particularly useful for cases where structural information is not available and rational engineering is not possible. Also, often random mutagenesis based directed evolution reveals novel mutations that are difficult to predict through rational structure-guided or computational analysis. Hence, random mutagenesis plays an important role in diverse mutant library generation for directed evolution-based protein engineering to advance our understanding of enzyme structure-function relationships. Over the years many approaches have been used to create random mutations and the most commonly used technique is epPCR (McCullum et al., 2010).

Basically, epPCR is a PCR performed using Taq DNA polymerase as enzyme for replicating the DNA. This technique uses the low fidelity characteristic of Taq polymerase in the presence of excess divalent cations to create random errors in the amplified region. Unlike the T7 polymerases, Taq DNA polymerase lacks a proofreading mechanism, and the error rate is increased in the presence of divalent metal cations (Cirino et al., 2003). Here, we randomly mutagenize the *TmAfc_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. We use Sequence and Ligation Independent Cloning (SLIC) for cloning amplified genes under error prone conditions into an E*. coli* vector for protein expression (M. Z. Li & Elledge, 2012).

3.2.1 Materials and Equipment

- 2X Phusion Master Mix (Thermo Fisher; Catalog Number: F531S)

- Gel/PCR Fragments Extraction kit (IBI Scientific; Catalog Number: IB47082)
- Taq DNA polymerase (NEB; Catalog Number: M0273L)
- T4 DNA polymerase (NE;, Catalog Number: M0203L)
- dNTPs (NEB; Catalog Number: N0447L)
- Dpn1 enzyme (NEB; Catalog Number: R0176L)
- $MnCl_2$ (Thermo Fisher; Catalog Number: AC193451000)
- $MgCl_2$ (ACS Organics; Catalog Number: 223211000)
- Forward and Reverse Primers (IDT DNA)
- Template *TmAfc-D224G* DNA (prepared in-house)
- E.cloni 10g competent cells (prepared in-house)
- PCR Thermocycler (Eppendorf)

3.2.2 Procedure:

1. The gene of interest was subjected to random mutagenesis in Insert PCR and cloned into an expression vector using sequence and ligation independent cloning (Stevenson et al., 2013). Run the Vector PCR and Insert PCR as directed in **Table 7**.

Table 7. Vector and Insert PCR reaction mixture and conditions

Vector PCR reaction mixture	**Vector PCR conditions**
• 1X Phusion Master Mix • 0.5 µM Forward and Reverse Primer • 20 ng *TmAfc-D224G* plasmid DNA	• Initial denaturation: 98°C/30 sec • [c]Denaturation: 98°C/10 sec • [c]Annealing: 60°C/10 sec • [c]Extension: 72°C/3 min • Final Extension: 72°C/5 min

	• Hold: 4°C c - Number of cycles: 30
Insert PCR reaction mixture • 1.25 U Taq Master Mix • 1x Taq buffer • 0.5 µM Forward and Reverse Primer • 20 ng *TmAfc-D224G* plasmid DNA • 0.2 mM dATP and dGTP • 1 mM dCTP and dTTP • 0.5 mM $MnCl_2$	**Insert PCR conditions** • Initial denaturation: 95°C/60 sec • [c]Denaturation: 98°C/30 sec • [c]Annealing: 60°C/30 sec • [c]Extension: 68°C/3 min • Final Extension: 68°C/5 min • Hold: 4°C c - Number of cycles: 20

2. After the PCR reaction is complete, use 2 µL of PCR product to mix with 3 µL PCR water and 1 µL of DNA loading dye. Add the resultant mixture to 1% agarose gel. Run gel electrophoresis for 120 V for 40 mins and image the gel using a DNA gel imager.
3. Perform Dpn1 digestion of the remaining PCR mixture using commercial kit and remove the unreacted products using PCR product cleanup kit and measure the DNA concentration of the eluted product.
4. Sequence and Ligation Independent Cloning:
    a. Prepare the reaction mixture for SLIC step as follows: 100 ng of vector DNA; 1:5 pmole ratio of insert DNA; 1.5 U of T4 DNA polymerase in 1x NEB Buffer 2.1. The total reaction volume can vary based on the concentrations of the DNA. (*Note: Make sure to prepare this mixture on ice as T4 DNA polymerase is active even at room temperature*).
    b. Incubate the reaction mixture at 25°C for 5 minutes and immediately place the mixture on ice.

c. Transform the entire SLIC reaction mixture into E. cloni 10g chemically competent cells using heat shock. (*Note: Electrocompetent cells can also be used. When using electrocompetent cells, make sure to clean up the SLIC mixture using PCR purification kit to remove buffer as salts in the buffer reduce electroporation efficiency*)

d. Add recovery media (SOC media) to transformed cells and incubate at 37°C for 2 hrs. Plate 50 µL-100 µL of mixture on an LB agar plate with 50 µg/mL Kanamycin and incubate at 37°C for 16 hours. The colonies obtained here will be used for evaluating the diversity and mutation rate of the epPCR library.

e. Add the remaining mixture to 5 mL LB media with 50 µg/mL Kanamycin and incubate at 37°C for 16 hours. Make multiple 1 mL glycerol stocks using the overnight culture to use for screening after library confirmation. (*Note: Alternatively, the remaining mixture can be plated in multiple large agar plates to get a mat of colonies. All the colonies can be scrapped using a cell scrapper and pooled into a single tube. The glycerol stocks of these pooled cells can be prepared and stored for screening*)

5. epPCR test library sequencing:

a. Pick 30 individual colonies from step 4d and grow them in 7 ml LB media with 50 µg/mL Kanamycin for 37°C for 16 hrs.

b. Perform plasmid extraction and send the samples to DNA sequencing facility for DNA sequencing using primers specific to the gene of interest.

c. The results from DNA sequencing can be analyzed and tabulated. The table below highlights all the point mutations obtained in our epPCR test library. More information about the results shown in the table can be found elsewhere (Agrawal et al., 2021b).

Table 8. Matrix of point mutations identified in TmAfc_D224G after epPCR

		Mutation to			
		0.5 mM Mn^{+2}			
Mutation from		T	C	A	G
	T	0	6	6	1
	C	8	0	2	1
	A	10	0	0	16
	G	1	0	9	0
Total mutations		60			

d. Using the matrix table data, calculate the overall mutation rate to estimate biases in mutational spectrum. Detailed information about analysis of the mutational spectrum and biases can be found elsewhere (Hanson-Manful & Patrick, 2013). More information about the results shown in the table can be found elsewhere (Agrawal et al., 2021b).

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

	0.5 mM Mn^{+2}	
Types of mutation	Frequency	Total Prop
Transitions		
A → G, T → C	22	36.67%
G → A, C → T	17	28.33%
Transversions		
A → T, T → A	10	16.67%
A → C, T → G	1	1.67%
G → C, C → G	1	1.67%
G → T, C → A	3	5.00%

Summary of bias		
Transitions/transversions		
AT → GC/GC → AT	1.29	
A → N, T → N	39	65.00%
G → N, C → N	21	35.00%
Mutation rate		
Mutations per kb	2.78	
Mutations per Tm0306 gene	3.75	

e. Input the information from the table above into PEDEL-AA server (http://guinevere.otago.ac.nz/stats.html), to obtain information about the epPCR library diversity and average number of amino acid mutations per construct. Detailed information about PEDEL-AA analysis for library diversity can be found elsewhere (Hanson-Manful & Patrick, 2013). More information about the results shown in the table can be found elsewhere (Agrawal et al., 2021b).

Table 10. Summary of epPCR library characteristics

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

3.2.3 Notes

1. Obtaining a large library diversity with the required number of average mutations is one of the critical steps for successful screening. A poorly characterized library can lead to a large number of false positives or false negatives during the screening step.
2. The amount of template DNA in PCR can be decreased to even 1–5 ng to reduce the percentage of non-mutated sequence identified in the final library during screening.
3. Commercial electrocompetent or chemically competent cells can be used to increase the transformation efficiency during the library transformation.
4. To get accurate library characteristics, pooled cells from the screening library in step 4e can be grown and plasmids can be extracted and sent for Next-Generation sequencing.

3.3 Procedure for high-throughput screening of mutants

3.3.1 Materials and Equipment

1. DBCO-PEG_4-Fluor545 (Sigma-Aldrich; Catalog Number: 760773)
2. β-fucosyl azide (Synthose; Catalog Number: FF757)
3. pNP-Xylose (Carbosynth; Catalog Number: EN03230)
4. pNP-Fucose (Carbosynth; Catalog Number: EN06434)
5. 40 µm nylon cell strainer (BD Falcon; Catalog Number: 352340)
6. Fluorescence activated cell sorter (Instrument details in (Agrawal et al., 2021b))

3.3.2 Procedure

1. **Cell culturing and protein expression:**

a. Take the glycerol stock pooled epPCR transformants from section 3.2.2 step 4e and grow in 5 mL LB media with 50 µg/mL Kanamycin at 37°C for 16 hrs.

b. Transfer 1 mL overnight culture to 20 mL LB media with 50 µg/mL Kanamycin and incubate at 37°C for 2-3 hrs until and OD_{600} of 0.4.

c. Induce protein expression by adding 1mM IPTG to the cultures and incubate 37°C for 1 hr. (*Note: Measure the OD_{600} after protein induction*)

d. Centrifuge 1 mL of induced culture in a microcentrifuge tube and discard the supernatant. Wash the cells twice with sterile 1X PBS (pH 7.4) and resuspend in 60 µL of 1X PBS.

2. **Glycosynthase reaction**

a. Add 10 mM β-fucosyl azide and 25 mM pNP-Xylose to the resuspended cells to a total volume of 150 µL. Incubate the mixture at 37°C for 2 hrs. The fucosynthase reaction between the substrates will occur in this step inside the cells.

b. After the reaction, centrifuge, and discard the supernatant and resuspend the cells in 150 µL 1X PBS and add 50 µM DBCO-PEG4_Fluor545 dye to cells.

c. Incubate the cells with the dye at 37°C for 30 mins. (*Note: The click chemistry reaction is nearly instantaneous. Shorter times can be used if necessary to decrease the incubation times. The optimum click chemistry conditions should be tested in vitro first before using them in vivo.*)

d. Centrifuge and resuspend cells into 1 mL 1x PBS buffer. Measure the OD_{600} using 100 µL of resuspended mixture in a 96 well microplate and multiply by a conversion factor of 2.5. (*Note: The conversion factor of 2.5 is to account for the smaller path length of microplate. More information about estimating this conversion factor is available in notes section 3.4*).

3. **FACS based cell sorting**

a. Once the OD_{600} is measured, dilute the cells in 1x PBS to get a final cell concentration in the order of 10^6. (*Note: This final concentration is optimal to run on FACS instrument to avoid clogging the capillary needle*).

b. Filter the samples using 40 µm filter and run the samples on FACS sorting machine with 560 nm excitation laser and 590 nm emission filter. (*Note: Alternatively, if 560nm excitation laser is unable, as is the case with many FACS machine, one can use 488 nm laser. However, the fluorescence signal is lower as compared to 560 nm laser).*

c. Obtained a sample fluorescence analysis of cell population in the mixture. Based on the positive (TmAfc_D224G) and negative (TmAfc_Wt) glycosynthase controls and additional sodium azide control, draw appropriate gates for sorting. (*Note: First gate can be drawn for fluorescence between positive and negative control. The second gate can be drawn for fluorescence less than positive control. More information about the gates is found in notes section 3.4*).

d. Sort the cells for which intensities falls in each gate into corresponding tubes or microplates containing 1 mL or 100 µL LB media with 50 µg/mL Kanamycin

respectively. (*Note: The container and volumes in which the sorting is done depends on how the sorted samples will be used in future. More information about this is found in the notes section 3.4*).

e. Grow the sorted cell cultures at 37°C for 16 hours and prepare glycerol stocks for storage.

3.4 Notes

1. This conversion factor should be calculated before the experiment in each lab. This can be done by measuring the OD_{600} of a grown culture in cuvettes and 96 well plate simultaneously at multiple dilutions. A slope of linear plot between the OD_{600} of cuvette and OD_{600} of microplate will give the conversion factor.

2. The gates drawn in a FACS sorting run are subjective to each project. For example, the first gate can be drawn for fluorescence between positive (TmAfc_D224G) and negative (TmAfc_Wt) control. These cells in gate indicate having decreased glycosynthase activity because less decrease in fluorescence intensity equals to less azide released. The second gate can be drawn for fluorescence less than positive control. The cells that fall in this gate indicate having increased glycosynthase activity because larger decrease in fluorescence is equivalent to increased azide released during GS reaction.

3. Sorted sample processing

   a. If the sorted cells are grown in tubes and no subsequent sorting is required, the overnight culture can be used for plasmid extraction. The extracted plasmids can be

sent for Next-Generation DNA Sequencing to identify the sequence of mutants present in each sorted gate.

b. If the sorted cells are grown in tubes and further sorting is desired to improve the activity of protein, multiple rounds of sorting or directed evolution of the targeted constructs can be performed by repeating all steps from section 3.3.

c. If single cells are sorted into 96 well microplates, the cells can sub-cultured, induced for protein expression and tested for glycosynthase activity. Sequence information can be obtained for the verified constructs to get a structure-function relationship.

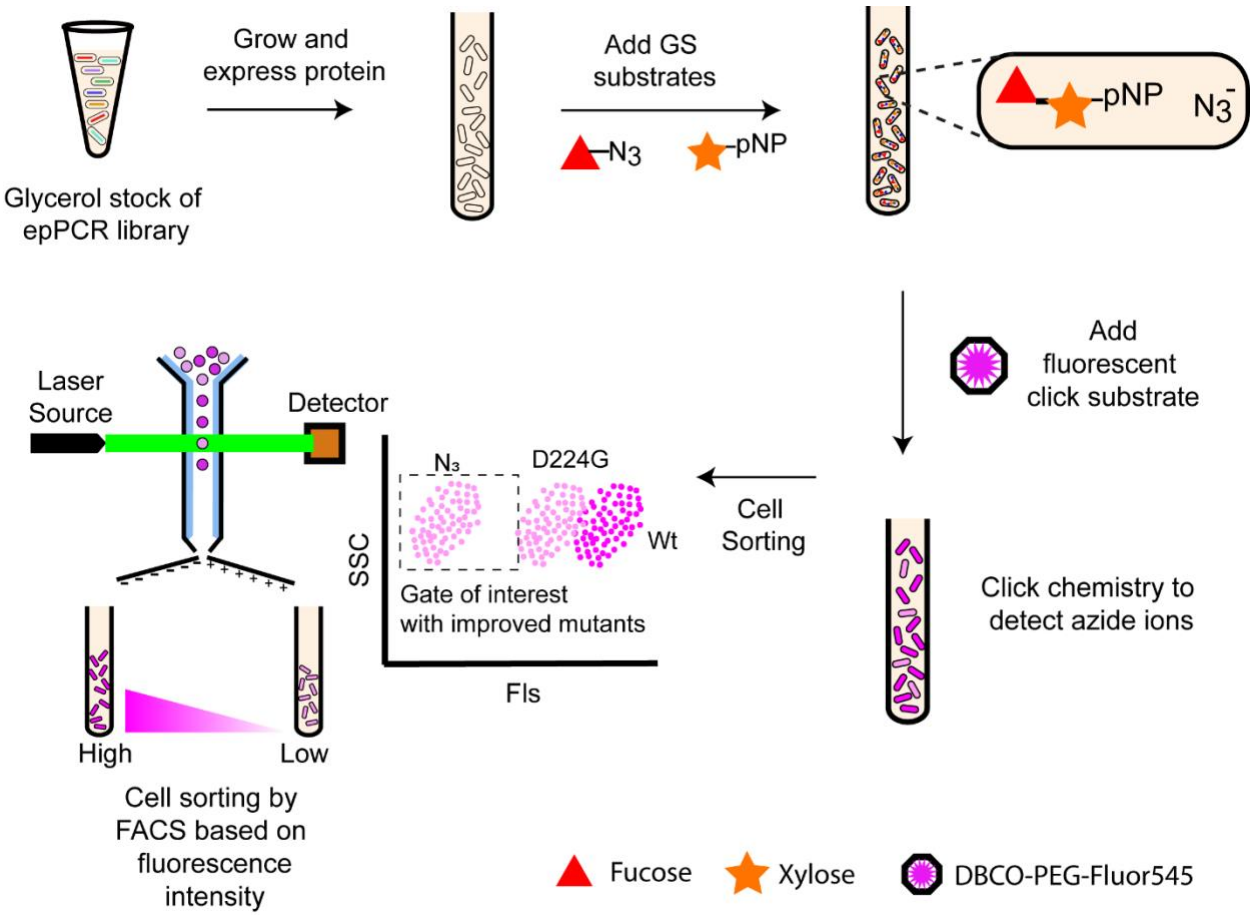

Figure 7: Screening of improved glycosynthase activity mutants using click chemistry screening method: The mutant cell library is first exposed to substrates for performing the glycosynthase reaction *in vivo* followed by adding the click reagent (DBCO-TAMRA or DBCO-PEF-Fluor545) for

click-labeling and detection of cells containing lower concentration of azide or azido sugars. Cells with lower fluorescence intensity (FI) are then sorted using FACS prior to isolating novel glycosynthase mutants or repeating another round of HTS using a directed evolution approach to enrich mutants sorted in previous rounds of screening.

4. Summary and Conclusions

Despite click-chemistry compatible azido sugars being used to study glycosylation pathways in living organisms to understand their function, localization, and metabolism in last two decades, oligosaccharides have not yet been synthesized extensively using azido sugars as donor groups. *In vivo* use of azido-based drugs or molecular imaging probes requires monitoring of organic azide stability and there are limited *in vivo* options available to rapidly detect and monitor azide in comparison to other leaving groups (such as pNP). Azido sugars as glycosyl donors are therefore currently of limited use in high-throughput screening during enzyme mutagenesis studies.

For our pCynGFP screening system, we used our in-house novel azide inducible promoter system engineered for *E. coli* bacteria whereby varying the inducer (i.e., azide ion) concentration the azide promoter system allows tunable expression for heterologous reporter GFP expression. This screening system allowed us to engineer CAZymes such as glycosyl hydrolases (or glycosynthases) that use azido-sugars as donor sugar substrates and autonomously monitor azide release during hydrolysis or transglycosylation. The development of an azide-specific promoter system for diverse cell types would also benefit various chemical biology-based research applications including azido-based drug molecules development and as molecular probes for real-time multicolor fluorescence imaging.

Probiotic gut microbes are selectively recognized and metabolically regulated by fucosylated glycans like human milk oligosaccharides (HMOs). Based on the Lewis blood group antigens,

HMOs contain a tetrasaccharide backbone that is selectively fucosylated to N-acetylglucosamine and/or galactose. GS activity between fucosyl azide and various acceptor sugars for the *TmAfc_D224G* mutant is low in terms of activity as well as specificity. However, this thermophilic enzyme may be engineered to enable bespoke HMO synthesis using diverse acceptor sugars. Using our optimized microplate based azide dectection method, we are currently identifying novel GS that are selectively engineered and evolved to synthesize specific HMOs. Using this approach, we will be able to identify GS mutants that are capable of efficiently synthesizing complex HMOs or Lewis blood group antigens.

Using our click chemistry-based screening system, the azido moiety can be selectively conjugated to alkyne-based fluorophores through copper-free click chemistry under suitable conditions compatible with the *in vivo* environment. We can further optimize these interactions to identify cells intracellularly expressing GS mutants with significantly higher activity using our proposed click chemistry screening method. To reduce false positive detection using similar ultra-high-throughput screening (uHTS) assays, future work must investigate the mechanism for the selective reduction of fluorescence of non-glycosylated triazole products. The choice of SPAAC reagent may also affect the uptake or diffusion of the formed triazole products in/out of the cells. As a result, mutant selection in uHTS would be less efficient. From the flow cytometry validation assays, there does not appear to be a significant amount of triazole product diffusing out from the cells. It is therefore necessary to perform additional research in order to gain a deeper understanding of the active and/or passive transmembrane transport properties of the specific SPAAC dye and its triazole products.

Finally, we a remark on using specific screening methods for azido sugars on GS activity. The casting epPCR (semi-random) library was used with pCyn_v2_GFP and click chemistry with the epPCR library (random) for screening, however any screening system can be combined with any mutagenesis method. Here we focused our work on providing a general framework for protein engineers seeking to alter protein activity, understand reaction mechanisms, or trying to investigate structure-function relationship of CAZymes. Since there are not enough screening systems for azido sugars, the mutagenesis coupled with screening strategies for azido sugars will give a benchmark for directed evolution studies in this field, specifically the Cyn regulon sensing system was tested for fluorescence with different azide concentrations and used for screening glycosynthase assay for the first time in case of CAZyme mutagenesis studies.

Acknowledgements:

Funding was provided by The National Science Foundation (NSF CHE award 1904890) and Rutgers School of Engineering. We would like to thank several current/past members of the Chundawat lab for their continued support in performing our day-to-day experiments. Special thanks to Dr. Tucker Burgin and Dr. Heather Mayes for providing recommendations for fucosynthase mutagenesis regions based on unpublished computational simulations.

References:

Adamczyk, B., Tharmalingam, T., & Rudd, P. M. (2012). Glycans as cancer biomarkers. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1820*(9), 1347–1353. https://doi.org/10.1016/j.bbagen.2011.12.001

Agrawal, A., Bandi, C. K., Burgin, T., Woo, Y., Mayes, H. B., & Chundawat, S. P. S. (2021a). Click-Chemistry-Based Free Azide versus Azido Sugar Detection Enables Rapid in Vivo Screening

of Glycosynthase Activity. *ACS Chemical Biology*, *16*(11), 2490–2501. https://doi.org/10.1021/acschembio.1c00585

Agrawal, A., Bandi, C. K., Burgin, T., Woo, Y., Mayes, H. B., & Chundawat, S. P. S. (2021b). Click-Chemistry-Based Free Azide versus Azido Sugar Detection Enables Rapid in Vivo Screening of Glycosynthase Activity. *ACS Chemical Biology*, *16*(11), 2490–2501. https://doi.org/10.1021/ACSCHEMBIO.1C00585

Andrés, E., Aragunde, H., & Planas, A. (2014). Screening glycosynthase libraries with a fluoride chemosensor assay independently of enzyme specificity: Identification of a transitional hydrolase to synthase mutant. *Biochemical Journal*, *458*(2), 355–363. https://doi.org/10.1042/BJ20131057

Bandi, C. K., Skalenko, K. S., Agrawal, A., Sivaneri, N., Thiry, M., & Chundawat, S. P. S. (2021a). Engineered Regulon to Enable Autonomous Azide Ion Biosensing, Recombinant Protein Production, and in Vivo Glycoengineering. *ACS Synthetic Biology*, *10*(4), 682–689. https://doi.org/10.1021/ACSSYNBIO.0C00449

Bandi, C. K., Skalenko, K. S., Agrawal, A., Sivaneri, N., Thiry, M., & Chundawat, S. P. S. (2021b). Engineered Regulon to Enable Autonomous Azide Ion Biosensing, Recombinant Protein Production, and in Vivo Glycoengineering. *ACS Synthetic Biology*, *10*(4), 682–689. https://doi.org/10.1021/acssynbio.0c00449

Ben-David, A., Shoham, G., & Shoham, Y. (2008a). A Universal Screening Assay for Glycosynthases: Directed Evolution of Glycosynthase XynB2(E335G) Suggests a General Path to Enhance Activity. *Chemistry & Biology*, *15*(6), 546–551. https://doi.org/10.1016/j.chembiol.2008.05.005

Ben-David, A., Shoham, G., & Shoham, Y. (2008b). A Universal Screening Assay for Glycosynthases: Directed Evolution of Glycosynthase XynB2(E335G) Suggests a General Path to Enhance Activity. *Chemistry and Biology*, *15*(6), 546–551. https://doi.org/10.1016/j.chembiol.2008.05.005

Boltje, T. J., Buskas, T., & Boons, G. J. (2009). Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research. *Nature Chemistry 2009 1:8*, *1*(8), 611–622. https://doi.org/10.1038/nchem.399

Burgin, T., & Mayes, H. B. (2019). Mechanism of oligosaccharide synthesis: Via a mutant GH29 fucosidase. *Reaction Chemistry and Engineering*, *4*(2), 402–409. https://doi.org/10.1039/c8re00240a

Cantarel, B. I., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37*(Database issue). https://doi.org/10.1093/NAR/GKN663

Cirino, P. C., Mayer, K. M., & Umeno, D. (2003). Generating Mutant Libraries Using Error-Prone PCR. In *Directed Evolution Library Creation* (pp. 3–10). Humana Press. https://doi.org/10.1385/1-59259-395-X:3

Cobucci-Ponzano, B., Strazzulli, A., Rossi, M., & Moracci, M. (2011). Glycosynthases in Biocatalysis. *Advanced Synthesis & Catalysis*, *353*(13), 2284–2300. https://doi.org/10.1002/ADSC.201100461

Council, N. R. (2012). *Transforming glycoscience: a roadmap for the future*.

Daugherty, P. S., Chen, G., Iverson, B. L., & Georgiou, G. (2000). Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proceedings of the National Academy of Sciences*, *97*(5), 2029–2034. https://doi.org/10.1073/pnas.030527597

Dube, D. H., & Bertozzi, C. R. (2005). Glycans in cancer and inflammation — potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery*, *4*(6), 477–488. https://doi.org/10.1038/nrd1751

Ducros, V. M.-A., Tarling, C. A., Zechel, D. L., Brzozowski, A. M., Frandsen, T. P., von Ossowski, I., Schülein, M., Withers, S. G., & Davies, G. J. (2003). Anatomy of Glycosynthesis. *Chemistry & Biology*, *10*(7), 619–628. https://doi.org/10.1016/S1074-5521(03)00143-1

Fan, X., Song, Q., Sun, D. en, Hao, Y., Wang, J., Wang, C., & Chen, X. (2022). Cell-type-specific labeling and profiling of glycans in living mice. *Nature Chemical Biology 2022 18:6*, *18*(6), 625–633. https://doi.org/10.1038/S41589-022-01016-4

Flynn, R. A., Pedram, K., Malaker, S. A., Batista, P. J., Smith, B. A. H., Johnson, A. G., George, B. M., Majzoub, K., Villalta, P. W., Carette, J. E., & Bertozzi, C. R. (2021a). Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell*, *184*(12), 3109-3124.e22. https://doi.org/10.1016/j.cell.2021.04.023

Flynn, R. A., Pedram, K., Malaker, S. A., Batista, P. J., Smith, B. A. H., Johnson, A. G., George, B. M., Majzoub, K., Villalta, P. W., Carette, J. E., & Bertozzi, C. R. (2021b). Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell*, *184*(12), 3109-3124.e22. https://doi.org/10.1016/J.CELL.2021.04.023/ATTACHMENT/3B503342-FA9B-4309-87D5-732F8708C5A5/MMC5.XLSX

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Hanson-Manful, P., & Patrick, W. M. (2013). *Construction and Analysis of Randomized Protein-Encoding Libraries Using Error-Prone PCR* (pp. 251–267). https://doi.org/10.1007/978-1-62703-354-1_15

Hayes, M. R., Bochinsky, K. A., Seibt, L. S., Elling, L., & Pietruszka, J. (2017). Development of a colourimetric assay for glycosynthases. *Journal of Biotechnology*, *257*, 162–170. https://doi.org/10.1016/j.jbiotec.2017.02.005

Hayes, M. R., & Pietruszka, J. (2017). Synthesis of glycosides by glycosynthases. In *Molecules* (Vol. 22, Issue 9). MDPI AG. https://doi.org/10.3390/molecules22091434

Klontz, E. H., Li, C., Kihn, K., Fields, J. K., Beckett, D., Snyder, G. A., Wintrode, P. L., Deredge, D., Wang, L. X., & Sundberg, E. J. (2020). Structure and dynamics of an α-fucosidase reveal a mechanism for highly efficient IgG transfucosylation. *Nature Communications 2020 11:1*, *11*(1), 1–14. https://doi.org/10.1038/s41467-020-20044-z

Kyselova, Z., Mechref, Y., Kang, P., Goetz, J. A., Dobrolecki, L. E., Sledge, G. W., Schnaper, L., Hickey, R. J., Malkas, L. H., & Novotny, M. v. (2008). Breast Cancer Diagnosis and Prognosis through Quantitative Measurements of Serum Glycan Profiles. *Clinical Chemistry*, *54*(7), 1166–1175. https://doi.org/10.1373/clinchem.2007.087148

Labrou, N. (2010). Random Mutagenesis Methods for In Vitro Directed Enzyme Evolution. *Current Protein & Peptide Science*, *11*(1), 91–100. https://doi.org/10.2174/138920310790274617

Lauc, G., Krištić, J., & Zoldoš, V. (2014). Glycans - the third revolution in evolution. *Frontiers in Genetics*, *5*(MAY), 145. https://doi.org/10.3389/FGENE.2014.00145/BIBTEX

Li, M. Z., & Elledge, S. J. (2012). *SLIC: A Method for Sequence- and Ligation-Independent Cloning* (pp. 51–59). https://doi.org/10.1007/978-1-61779-564-0_5

Li, Y., Liu, D., Wang, Y., Su, W., Liu, G., & Dong, W. (2021). The Importance of Glycans of Viral and Host Proteins in Enveloped Virus Infection. *Frontiers in Immunology*, *12*, 1544. https://doi.org/10.3389/FIMMU.2021.638573/BIBTEX

Mackenzie, L. F., Wang, Q., Warren, R. A. J., & Withers, S. G. (1998). Glycosynthases: Mutant glycosidases for oligosaccharide synthesis. *Journal of the American Chemical Society*, *120*(22), 5583–5584. https://doi.org/10.1021/JA980833D/ASSET/IMAGES/MEDIUM/JA980833DU00002A.GIF

McArthur, J. B., & Chen, X. (2016). Glycosyltransferase engineering for carbohydrate synthesis. *Biochemical Society Transactions*, *44*(1), 129–142. https://doi.org/10.1042/BST20150200

McCullum, E. O., Williams, B. A. R., Zhang, J., & Chaput, J. C. (2010). *Random Mutagenesis by Error-Prone PCR* (pp. 103–109). https://doi.org/10.1007/978-1-60761-652-8_7

Meldal, M., & Tornøe, C. W. (2008). Cu-Catalyzed Azide−Alkyne Cycloaddition. *Chemical Reviews*, *108*(8), 2952–3015. https://doi.org/10.1021/cr0783479

Moses, J. E., & Moorhouse, A. D. (2007). The growing applications of click chemistry. *Chem. Soc. Rev.*, *36*(8), 1249–1262. https://doi.org/10.1039/B613014N

Nilsson, K. G. I. (1988). Enzymatic synthesis of oligosaccharides. *Trends in Biotechnology*, *6*(10), 256–264. https://doi.org/10.1016/0167-7799(88)90058-3

O'Connell, M. A., Belanger, B. A., & Haaland, P. D. (1993). Calibration and assay development using the four-parameter logistic model. *Chemometrics and Intelligent Laboratory Systems*, *20*(2), 97–114. https://doi.org/10.1016/0169-7439(93)80008-6

Park, B. H., Karpinets, T. v, Syed, M. H., Leuze, M. R., & Uberbacher, E. C. (2010). CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology*, *20*(12), 1574–1584. https://doi.org/10.1093/glycob/cwq106

Plante, O. J., Palmacci, E. R., & Seeberger, P. H. (2001). Automated Solid-Phase Synthesis of Oligosaccharides. *Science*, *291*(5508), 1523–1527. https://doi.org/10.1126/science.1057324

Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. (2019). Glycosylation in health and disease. *Nature Reviews Nephrology 2019 15:6*, *15*(6), 346–366. https://doi.org/10.1038/s41581-019-0129-4

Sathya, T. A., & Khan, M. (2014). Diversity of glycosyl hydrolase enzymes from metagenome and their application in food industry. *Journal of Food Science*, *79*(11), R2149–R2156. https://doi.org/10.1111/1750-3841.12677

Saumonneau, A., Champion, E., Peltier-Pain, P., Molnar-Gabor, D., Hendrickx, J., Tran, V., Hederos, M., Dekany, G., & Tellier, C. (2016). Design of an α-L-transfucosidase for the synthesis of fucosylated HMOs. *Glycobiology*, *26*(3), 261–269. https://doi.org/10.1093/GLYCOB/CWV099

Shaikh, F. A., & Withers, S. G. (2008). Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesisThis paper is one of a selection of papers published in this Special Issue, entitled CSBMCB — Systems and Chemical Biology, and has undergone the Journal's usual peer review process. *Biochemistry and Cell Biology*, *86*(2), 169–177. https://doi.org/10.1139/O07-149

Stevenson, J., Krycer, J. R., Phan, L., & Brown, A. J. (2013). A Practical Comparison of Ligation-Independent Cloning Techniques. *PLoS ONE*, *8*(12), e83888. https://doi.org/10.1371/journal.pone.0083888

Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., & Seeberger, P. H. (2022). Essentials of Glycobiology. *Essentials of Glycobiology*. https://doi.org/10.1101/9781621824213

Varki, A., & Gagneux, P. (2015). *Biological Functions of Glycans* (3rd ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY). http://europepmc.org/books/NBK453034

Wang, Z., Chinoy, Z. S., Ambre, S. G., Peng, W., McBride, R., de Vries, R. P., Glushka, J., Paulson, J. C., & Boons, G.-J. (2013). A General Strategy for the Chemoenzymatic Synthesis of Asymmetrically Branched *N* -Glycans. *Science*, *341*(6144), 379–383. https://doi.org/10.1126/science.1236231

Weijers, C. A. G. M., Franssen, M. C. R., & Visser, G. M. (2008). Glycosyltransferase-catalyzed synthesis of bioactive oligosaccharides. *Biotechnology Advances*, *26*(5), 436–456. https://doi.org/10.1016/j.biotechadv.2008.05.001

Wilson, D. S., & Keefe, A. D. (2001). Random Mutagenesis by PCR. In *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471142727.mb0803s51

Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J., & Arnold, F. H. (2019). Machine learning-assisted directed protein evolution with combinatorial libraries. *Proceedings of the National Academy of Sciences*, *116*(18), 8852–8858. https://doi.org/10.1073/pnas.1901979116

Yang, J., Ruff, A. J., Arlt, M., & Schwaneberg, U. (2017). Casting epPCR (cepPCR): A simple random mutagenesis method to generate high quality mutant libraries. *Biotechnology and Bioengineering*, *114*(9), 1921–1927. https://doi.org/10.1002/bit.26327

Zeuner, B., Muschiol, J., Holck, J., Lezyk, M., Gedde, M. R., Jers, C., Mikkelsen, J. D., & Meyer, A. S. (2018). Substrate specificity and transfucosylation activity of GH29 α-L-fucosidases for enzymatic production of human milk oligosaccharides. *New Biotechnology*, *41*, 34–45. https://doi.org/10.1016/j.nbt.2017.12.002