SELECTIVE GROWTH OF TARGETED BACTERIA USING ENZYMATICALLY SYNTHESIZED OLIGOSACCHARIDES

By

NAMRATHA SUBHASH

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, 2018

ABSTRACT OF THE THESIS

Selective growth of targeted bacteria using enzymatically synthesized oligosaccharides

By NAMRATHA SUBHASH

Thesis Director:

Dr. Shishir P. S. Chundawat

Designer oligosaccharides can be used as prebiotics for selective bacterial growth and are therefore relevant to understanding how complex biological systems function in a nutrient-limited environment. Enzymatic synthesis offers an alternative route to producing designer oligosaccharides with higher reaction specificity, product purity, and lower production costs compared to standard chemical synthesis routes. Here, we focus on glycosyl hydrolase (GH) enzymes from GH family 29 that were reverse engineered to function as glycosynthase enzymes for synthesis of fucosylated oligosaccharides. First, we used two distinct model fucosylated oligosaccharides as a nutrient carbon source to highlight how targeted growth of bacteria is feasible for bacterial species expressing the wild-type GH 29 enzymes over bacterial species not expressing the relevant gene of interest. This proof-of-concept experiment helped clearly establish how designer oligosaccharides could be used to modulate the growth of specific bacteria from a complex microbial milieu. Next, detailed bioinformatics analyses and mechanistic assays of two GH family 29 enzymes were carried out to study the impact of certain active-site nucleophilic site mutations on glycosynthase activity using model substrates. Here, we specifically focused on GH29 enzymes with known distinct substrate preferences belonging to *Bifidobacterium longum* subspecies

infantis or *Blon_2336* (ACJ53394.1) and *Thermotoga maritima* or *Tm_0306* (AAD35394.1). Additional molecular docking simulations were carried out to provide a clear rationale for why certain GH29 family of enzymes are able to selectively hydrolyze certain isomers of fucosyllactose to facilitate selective bacterial growth. Lastly, molecular docking simulations were conducted to identify other novel mutation sites to the native GH29 enzyme that would be necessary to improve fucosylated oligosaccharide synthesis efficiency using engineered glycosynthases. In summary, this study provides a clear rationale for how distinct carbohydrate-based oligosaccharides, can be synthesized chemoenzymatically using a rationale structure-guided GH enzyme engineering approach, and how these distinct carbon sources can be used to selectively target growth of certain bacterial species carrying the relevant GH genes of interest.

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. Shishir Chundawat, for his academic guidance, understanding and support throughout my graduate studies. I would like to thank Dr. Charles Roth and Dr. Haoran Zhang for being part of my thesis committee. I would like to acknowledge Chandra Kanth Bandi for being an excellent mentor, for all his technical support and his patience in teaching me all the relevant lab techniques. I would like to thank all my laboratory colleagues for being amazing lab partners. I would like to thank Dr. Andrzei Joachimiak from Argonne National Lab (Illinois) for kindly gifting the *Bifidobacterium longum* subsp. *infantis* gene. I would like to thank everyone in the Chemical and Biochemical Engineering Department (CBE). I would like to thank my friends and roommates for keeping me motivated during my work.

Lastly, I am grateful to my mother, father, and brother whose support as always, plays a major role in my achievements. This would not have been possible without you.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

Chapter 1 : Introduction

1.1 Background

In human milk there are over 200 structurally distinct oligosaccharides or glycans that have been identified to date ([2], [3], [4], [5]). Of these, fucosylated oligosaccharides are one of the most abundant class of glycans found in human milk. The human milk oligosaccharides (HMOs) are known to play a critical role in establishing a stable gut microbiome in infants. These oligosaccharides can passage via the infant's acidic stomach pH conditions as well as escape digestion by pancreatic enzymes to facilitate growth of beneficial gut bacteria. Several *in vitro* and *in vivo* studies conducted on HMOs have helped us understand the physiological benefits of human milk on infant health [6]. In this specific *in vitro* study, the focus was on understanding how neutral HMOs, acidic HMOs and partially metabolized lactose HMO fractions are beneficial to the intestinal health of a breast-fed infant by stimulating growth of beneficial microbiota (e.g., some *Bifidobacterium* species) and helping prevent infectious growth of pathogenic microbiota. Glucose, galactose, N-acetylglucosamine, fucose, and sialic acid (like N-acetylneuraminic acid) are the five predominant monosaccharides that makeup most HMOs [7]. Most classes of HMOs are built around a lactose core [8] that is located at the reducing end and the fucosylated HMOs have a fucose moiety at the non-reducing end. Here, we focus on model fucosylated lactose, namely 2'-FL and 3'-FL. 2'-FL is one of the most abundant oligosaccharide found in human milk and it makes up the core structure of more than 30% of all HMOs [9].

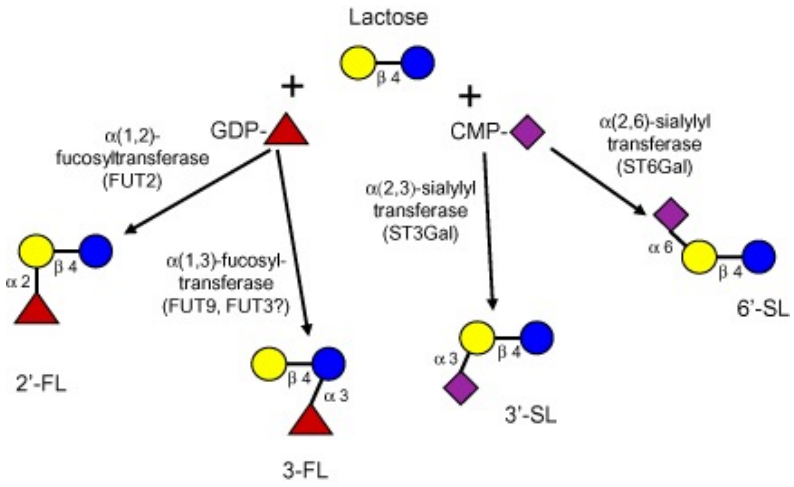


Figure 1-1: Synthesis routes of core fucosylated HMOs like 2'-FL, 3-FL, 3'SL and 6-SL using mammalian fucosyltransferases in humans are shown here. The yellow circles are galactose, blue circles are glucose, red triangles are fucose and purple diamonds are sialic acid (figure taken from Castanys-Muñoz E et al. [10])

In human cells, 2'-FL is produced by a specific α-1,2-fucosyltransferase (FUT2) enzyme that transfers fucose to the 2nd position of galactose (β-Gal) residue, whereas for 3-FL a variety of fucosyltransferases are thought to also work [10]. Production of 3'-sialyllactose (3'-SL) and 6'-sialyllactose are also possible using specific sialylyltransferases. The glycosyltransferase (GT) enzymes necessary for synthesis of diverse HMOs are not uniformly distributed in all humans so they may or may not be present in mother's milk based on her specific genes [11]. For infants who cannot be breast fed due to certain complications, such HMOs must be fed through a formula-based milk and thus, there is an unprecedented increase in demand for alternative sources of these HMOs. Development of feasible synthetic bulk preparation methods of these complex glycans and understanding their biological roles will be of great value to the society. Various attempts have been made to extract these glycans from human and cow milk [12] or produce these chemically using traditional organic synthesis methods. But it was found that due to numerous protection and de-protection steps involved, the chemical synthesis of exact HMO stereoisomers can become very expensive [13]. This has led to the utilization of enzymatic approaches that

utilize various GTs and glycosidases/glycosyl hydrolases (GH) (EC 3.2.1.) as biocatalysts instead. GTs are natural enzymes which help synthesize a glycosyl bond whereas GHs are ideally meant to break these glycosyl bonds. Simple oligosaccharides like 2'-FL can be produced in *E. coli* using FUT2 GTs as shown in Figure 1-1 but the challenge is to produce more complex HMOs [14-18] and with high-level protein expression and purification of GTs using *E. coli* [19] Two specific fucosidase enzyme genes from GH family 29 were chosen for this study to explore their potential to synthesize HMOs, namely *Bifidobacterium longum* subsp. *infantis, Blon_2336* (ACJ53394.1) and *Thermotoga maritima, Tm_0306* (AAD35394.1). The former gene codes for BiAfc (subfamily B) enzyme, which is an α-1,3/4-fucosidase; whereas the latter encodes for TmαFuc (subfamily A), which is an α-1,2-fucosidase. GH 29 family of enzymes hydrolyzes glycan oligosaccharides into simple products using a two-step retaining mechanism as illustrated in the Figure 1-2 below. For example, consider the case of GH29 reaction in the presence of an activated sugar donor like p-nitrophenyl-fucose (pNP-fucose). When this substrate enters the catalytic pocket of the enzyme (say TmαFuc), the catalytic nucleophile residue attacks the anomeric carbon-1 (C-1) position attached to para-nitrophenyl (pNP) moiety located at the α-position. The transition state complex formation stabilization helps the enzyme drive forward the hydrolysis of the glycosidic bond further using hydrogen extracted from the carboxylic acid of the acid/base residue. This results in an inverted configuration enzyme-substrate complex. At this juncture, the reaction can either proceed forward as a hydrolysis reaction in the presence of water, which is usually the case for GH enzymes, or it could be a transglycosylation reaction in the presence of another sugar molecule. A native GH enzyme can behave sometimes like transglycosidases (TG) under

specific reaction conditions or sometimes some GH enzyme families might display predominantly TG activity. In the presence of even small quantities of water, the reaction favors hydrolysis of the enzyme-substrate complex. However, if an acceptor group such as lactose is properly oriented and supplied to within the enzyme-substrate complex, then it may result in formation of TG products like fucosyllactose (FL). The disadvantage of a TG reaction using native GH enzymes is the highly plausible secondary hydrolysis of the TG products. Secondary hydrolysis further leads to degradation of TG products into hydrolysis products. In such scenarios, to avoid secondary hydrolysis reactions mutant glycosynthases (GS) enzymes can be used instead.

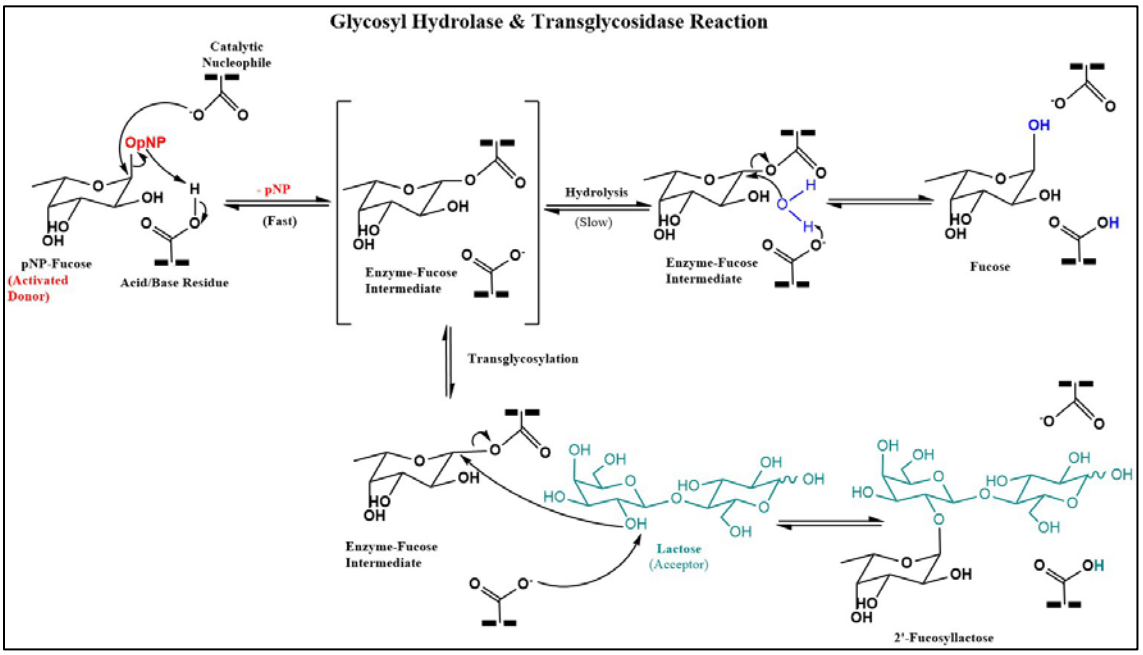


Figure 1-2: GH29 driven hydrolysis and TG reaction mechanism using pNP-fucose as the activated donor and lactose as the acceptor in the wild type enzyme active site (drawn using ChemDraw)

A GS enzyme is a GH enzyme mutated at its catalytic nucleophile residue to a residue that cannot participate as a nucleophile to attack the C1 carbon of the donor sugar group. Residues like Alanine, Serine and Glycine are commonly chosen since they lack a

carboxylic group to help reduce secondary hydrolysis reactions of the TG products. However, the enzyme needs an external nucleophile to form the enzyme-substrate complex as it is lacks its own, or an activated sugar molecule should be a functional substitute for the missing nucleophilic residue. Small molecules with similar electronic or steric features that complement the enzyme active-site could function as external nucleophiles. External nucleophilic or electron-donating compounds like sodium azide or sodium formate are often used to form β-fucosyl intermediate compounds. Such reactions are termed as 'chemical rescue', as the name itself suggests, these molecules help rescue the hydrolysis activity of mutant enzyme. In GS reactions, the mutated enzymes are fed with substrates (fucose based) in the β-conformation like β-fucosyl azide or β-fucosyl formate to mimic the enzyme-substrate transition state complex. In the presence of excess of acceptors sugar like lactose or xylose, the reaction slowly shifts towards formation of the TG products as discussed earlier. The advantage of using GS enzymes is the absence of the native nucleophilic group that prevents secondary hydrolysis of synthesized TG products.

However, we currently still lack insight on how to effectively engineer GH into GS using a structure guided enzyme engineering approach.

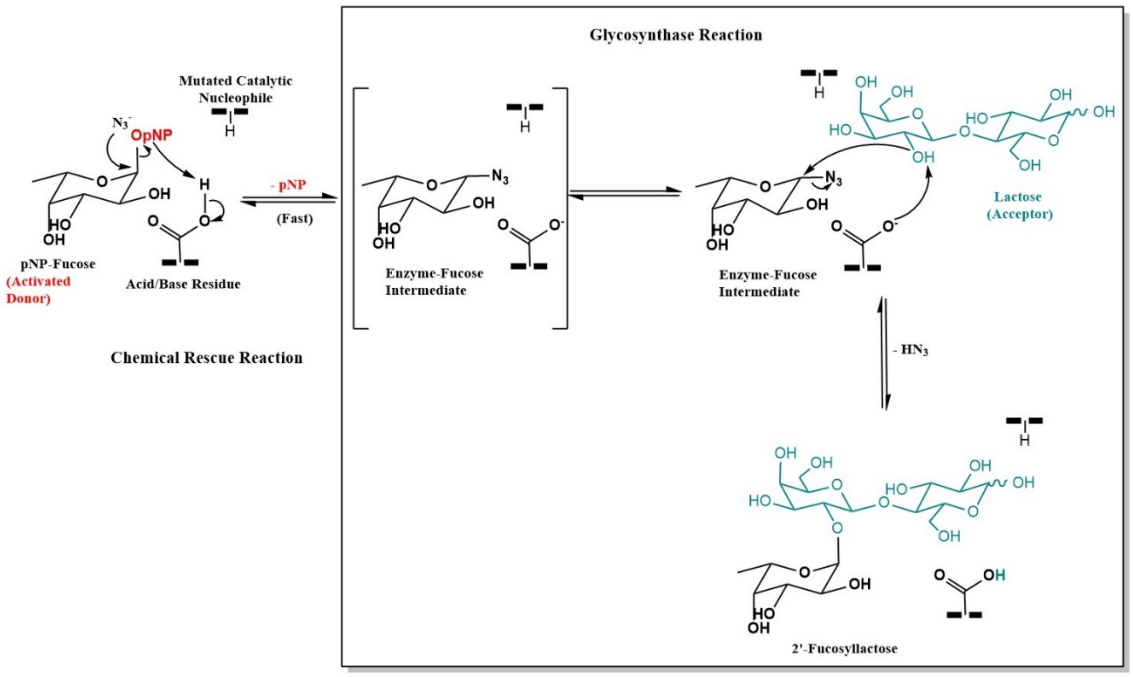


Figure 1-3: Mechanism of chemical rescue and hypothesized GS mechanism reaction using pNP-fucose as substrate for chemical rescue and β-fucosyl azide (donor) with lactose (acceptor) for GS reaction (drawn using ChemDraw), respectively. The β-conformation of β-fucosyl azide mimics the enzyme-substrate transition state complex of a GH reaction. In the presence of excess of acceptors sugar like lactose, the reaction slowly shifts towards formation of the fucosyllactose products.

1.2 Objectives of thesis

In this research, we have studied in detail how the GS reaction (prepared using two GH29 enzymes) can be used to create simple fucosylated glycans to ultimately synthesize more complex HMO-like glycans. It is important to first identify the sugar acceptor-binding site of these enzymes since we intend to reverse the enzyme's hydrolysis activity. The acceptor in a hydrolysis reaction does not need to bind to the product site, in fact the hydrolysis products are cleaved from the enzyme active site as soon as they are formed. In a GS reaction, the acceptor group needs to bind to the enzyme to remain available for reaction with an activated fucose moiety. This is also the reason behind adding stoichiometrically

excess of acceptor sugars since water is omnipresent and can be easily accessible compared to the acceptor sugars. The acceptor sugars are also too large in terms of size or are not able to bind properly to the desired product site due to hydrophobic interactions with certain amino acids near the active site. Therefore, rationally driven mutations in the catalytic pocket aiming to create a void for large acceptor sugars or making the pocket less hydrophobic provides a solution to creating more efficient GS capable of synthesizing HMOs. Therefore, for the present thesis, there are two specific goals that were pursued:

1) Develop a rational-structure guided approach to synthesize designer HMOs by reverse engineering the active site residues of a native family 29 glycosyl hydrolases (GH) to make highly effective glycosynthases (GS)

2) Demonstrate the selective prebiotic potential of designer HMOs for targeting and modulating the growth of selectively engineered bacteria carrying relevant GH genes.

In the section 1.1, the reaction mechanism that would be applicable for completion of the first goal of this thesis was briefly discussed. For the GS reaction, β-fucosyl azide (as activated donor) and lactose or pNP-lactose (as acceptors) were used. Since, lactose is a large sugar group for the catalytic pocket of these enzymes, the requirement of further structure-guided mutations was analyzed using various modeling tools. Necessary sites for mutation were picked based on multiple sequence alignments, phylogenetic tree analysis, and molecular docking simulations (Chapter 5).

For completion of the second goal, we intend to use these designer glycans (that are theoretically synthesized using the engineered enzymes using an approach highlighted under aim 1, to selectively grow specific bacteria carrying specific GH genes that facilitate utilization of glycans as a carbon source. These bacteria are engineered *E. coli* cells that were transformed with plasmid DNA vectors carrying either *Blon_2336* or *Tm_0306* wild-type GH29 genes under T7/ T5- type promoter expression control. These transformed bacteria upon induction of T7/ T5 promoters can produce specifically BiAfc_WT or TmαFuc_WT GH29 enzymes, respectively. We hypothesized that without expression of this catabolic GH machinery, the engineered *E. coli* bacterial cells cannot grow on fucosylated HMOs. Research conducted by Yu Z-T *et al.* [1] (Figure 1-4) also provides support to this hypothesis as shown using native Bifidobacterium based bacterial systems. In particular, they grew four individual representative strains of infant gut microbes *B. longum* ATCC15697, *B. longum* JCM7007, *E. coli* K12 and *C. perfringens* to mimic the infant microbiota, under in vitro conditions for 48 hours fermentation time, in the presence of key HMOs like purified 2′-FL, 3-FL, LDFT (lactodifucotetraose), FOS (fructooligosaccharides) or a complete HMOs mixture. Both the bifidobacterium strains, B. *longum* ATCC15697 and *B. longum* JCM7007, exhibited strong growth in the presence of 5 g/L of HMOs and slightly above average growth when supplemented with 2'FL, 3-FL and FOS. In contrast, *C. perfringens* and *E. coli* K12 species showed suppression of growth

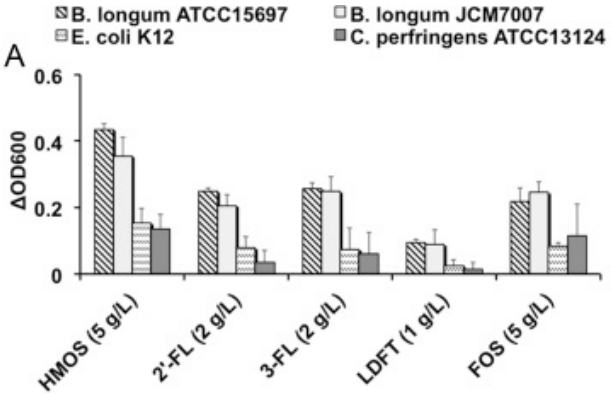


Figure 1-4: Growth of B. longum ATCC15697, B. longum JCM7007, E. coli and C. perfringens when measured at 600 nm after 48 hours incubation with purified 2′-FL, 3-FL, LDFT, FOS or a complete HMOS mixture [1].

in nearly all cases. The key observation here was that the native *E. coli* K12 did not show growth in the presence of any of these HMOs. However, to-date controlled growth experiments for well-defined and engineered bacterial systems (capable of expressing only select GH genes) on model fucosylated HMOs have not been completed to carefully study the syntrophic growth of microbiota in a complex system.

Sela *et al.* [20] have also shown that the wild-type protein expressed from gene *Blon_2336* can hydrolyze 3'-FL as it is predicted to be an α-1,3/4-fucosidase. To validate this experimentally, they tested various GH29 and GH95 fucosidase genes from *Bifidobacterium longum* subsp. *infantis* ATCC 15697 namely *Blon_0248, Blon_0426, Blon_0346, Blon_2335 and Blon_2336* on various HMO-like substrates (e.g., 2'-FL, 3'-FL, and Fucα-1,2-Gal). The genes *Blon_0248, Blon_0426, Blon_2336* code for α-L-fucosidases belonging to GH29 family, *Blon_2335* codes for a GH95 α-fucosidase while *Blon_0346* is an additional gene that has been annotated as a β-galactosidase that also possesses a α-fucosidase domain. They found that *Blon_2335* enzyme, which is an α-1,2-fucosidase, exhibited stronger cleavage activity of the two α-1,2 linked substrates and it partially cleaved α-1,3 linkages of 3'-FL as well. *Blon_2336* enzyme cleaved the 3'-FL bond and was identified as an α-1,3/4-fucosidase. A thin later chromatography (TLC) based analysis (Table 1) was conducted to find the fucosidase linkage specificity of each expressed gene of interest. This

α-Fucosidase	Linkage specificity to:		
	2'-FL	3'-FL	Fucα-1,2-Gal
Blon_0248	–	–	–
Blon_0426	–	–	–
Blon_0346	–	–	+
Blon_2335	+++	++	+++
Blon_2336	–	+++	–

Table 1: Hydrolysis of 2'-FL, 3'FL and Fucα-1,2-Gal type HMOs using Bifidobacterium longum subsp. *infantis fucosidase genes of GH 29 and 95; TLC based product bands were scored qualitatively for in vitro enzyme activity, where; –, no signal; +, weak; ++, moderate; +++, strong [20].*

research clearly proved that *Blon_2336* gene specifically hydrolyses 3'-FL but not 2'-FL. Our current research aims at engineering *E. coli* to be incorporated with the *Blon_2336* and *Tm_0306* genes that express GH29 enzymes that target hydrolysis of distinct fucosylated lactose. Specifically, the genetically modified cells are hypothesized to show selective growth in the presence of specific glycans like 2'-FL and 3'-FL based on the expressed enzyme's substrate specificity.

Chapter 2 : Proof-of-concept engineered cells growth on fucosylated HMOs

2.1 Bacterial culture growth studies using cells with and without control plasmids

Bacteria can utilize the various chemicals present in its surrounding environment for growth as nutrients. Many bacteria like *E. coli* BL21 can be grown in common lab culture media that are designed to provide all the essential nutrients in solution. The major elements needed for bacterial growth are carbon, nitrogen, oxygen and hydrogen. Luria-Bertani (LB) media is a nutrient rich culture medium that is most commonly used to grow various bacterial cells and it is composed primarily of yeast extract, tryptone and NaCl. The advantage of growing *E. coli* as a model system among other gut-microbiome specific bacteria is their rapid growth rate under aerobic conditions and basic nutritional requirements, which can also be met with a minimal medium instead of LB medium [21]. To study the growth of bacteria in the presence of select carbon sources, a nutrient limited minimal media was first chosen. Minimal media is made of M9 minimal salts that are basically Na_2HPO_4, KH_2PO_4, NaCl, NH_4Cl salts with $MgSO_4$ and $CaCl_2$. The engineered *E. coli* expressing suitable GH29 enzymes, as discussed in section 1.2, can hydrolyze specific FL into fucose and lactose. Lactose, can then be broken down to fermentable sugars by the *lacI* gene product present as part of plasmid DNA. The *lacI* codes for a β-galactosidase enzyme that can hydrolyze lactose into glucose and galactose. Glucose and galactose sugars can be both readily metabolized by *E. coli* as a carbon source driving for cell growth. Our goal was set up proof-of-concept experiments that confirm that only engineered *E. coli* cells with the ability to produce both fucosidase and *lacI* gene products would be able to grow in the presence of FL. But first we had to determine the minimum

concentration of lactose necessary for the cultures to show considerable amount of cell growth that could be readily monitored via OD_{600} absorbance measurements. Thus, a systematic study was performed to monitor *E. coli* cell density in the presence of a carbon source enriched in varying concentrations of fucose, lactose and both added together. Here, the *E. coli* cells were transformed with a suitable control plasmid carrying the necessary *lacI* genes. This study would help us verify if cells devoid of the control plasmid, which carried a *lacI* DNA fragment, would show selective growth in the presence of any of these carbon sources.

2.1.1 Materials and methods

To determine the minimum lactose concentration required to achieve measurable changes in cell density, the *E. coli* BL21 (DE3) expression strain was chosen. Experiments were conducted by growing control BL21 cells and BL21 cells transformed with pEC_GFP_CBM3a plasmid. The pEC plasmid consists of a *lacI* gene, as discussed earlier, and a gene under T5-promotor control to express green fluorescent protein or GFP-fused CBM3a as highlighted in another study published by Chundawat lab [22]. This control plasmid was transformed into cells, to produce BL21_pEC_GFP_CBM3a or BL21_GFP strains that were used first to optimize cell using lactose as a carbon source. Also, it would be easier to visualize if the green fluorescent protein was being expressed by the cells in the presence of lactose (which is provided as a carbon source here), which also behaves as an inducer of the *lac* operon under T5-promoter control [23],[24]. These cells were already available in the lab as glycerol stock stored at -80°C and were used directly as described below. Original glycerol stocks were first used to inoculate a starter culture of LB media

(10 ml) which was incubated overnight (~16 hours) at 37°C in an incubator shaker from New Brunswick Scientific Excella E24 series at 200 rpm. Next, for BL21 control only and BL21_pEC_GFP_CBM3a cells, the suitable antibiotics (e.g., chloramphenicol and kanamycin, respectively) were included in the respective media to avoid contamination and facilitate targeted cell line growth. Then, 5% of these culture volumes were then used as the starting inoculum for a larger batch of minimal media (120 ml), which had a minimum of 2% glucose (w/v) concentration with the respective antibiotics. The cells grown to the exponential growth phase (monitored via OD_{600} measurements) where they are expected to show high growth rates for future experiments. The cells were recovered from the exponential growth phase after gently centrifuging using the Eppendorf Centrifuge 5810 R at 3900 rpm for 15 minutes at 4°C to obtain the cell pellets. The aim was to transfer the recovered cells from their exponential growth phase into a medium that is devoid of the regular carbon source such as glucose. To remove trace amounts of glucose present from the cell pellet due to residual minimal media from earlier, the cell pellets were gently resuspended at 4°C into a minimal media without any glucose and were centrifuged again to pelletize the cells. This procedure was repeated twice after which the washed cells pellets were resuspended in a minimal media without glucose and kept on ice to avoid further cell growth. Each of the respective cell cultures were added to microplate wells at defined starting cell ODs with varying concentrations of either lactose, fucose or both sugars together. The 2 ml deep well microtiter plates with total culture volume of 1.5 ml shaken at 37°C, 200 rpm. The growth curve of all cells cultures was studied at defined durations by measuring their optical density at 600 nm wavelength absorbance using Spectramax

M5e plate reader (Molecular Devices) at regular intervals for a total fermentation time of 50 hours and at 37°C, 200 rpm.

2.1.2 Results and discussion

The BL21 controls (no plasmid) and BL21_ GFP (with BL21_pEC_GFP_CBM3a plasmid) cells were transferred from LB media to minimal media containing 2% glucose and were grown till they reached an OD_{600} in the range of 0.4-0.6. The cells were then washed and grown in minimal media in the presence of increasing concentrations of lactose alone (0-3 mM), fucose alone (0-6 mM), or both sugars together. Based on the results shown in Appendix, Figure (A-1 to A-3), it is clearly evident that the BL21 control cells

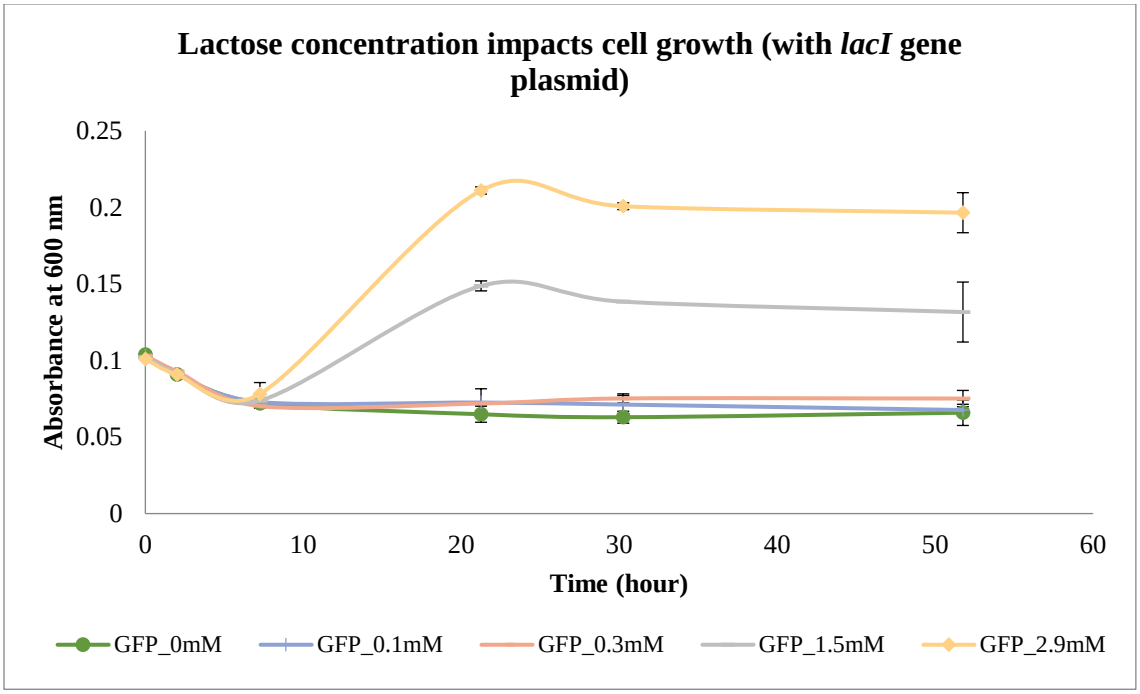


Figure 2-1: Growth curve of BL21_GFP (with plasmid) cells in the presence of varying concentrations of lactose is shown. Here, minimal 1.5 mM lactose concentrations showed clear indication of minimal cell growth. The experiment was performed using duplicates and the error bars depict for a single standard deviation from the mean value reported for biological replicates.

(without any plasmid) in the presence of varying concentrations of lactose, fucose, and both sugars did not show significant growth due to lack of the necessary catabolic funcions

genes encoded on the *lac* operon. However, the BL21_GFP cells (with the plasmid) displayed growth in the presence of lactose as well as in the presence of both lactose & fucose at a minimum concentration of 1.5 mM lactose. No significant growth was seen in the presence of fucose (Appendix, Figure A-4), further indicating that these cell lines lacked the suitable enzymes to metabolise this sugar.

Based on the % hydrolytic activity expected for the expressed fucosidase enzyme, which was found to be ~90% (based on other experiments described in the subsequent chapters), and based on simple mass balance calculations, we back calculated the minimum amount of FL (i.e., 1.67 mM) necessary to obtain hydrolyzed lactose based product concentration of 1.5 mM. These experiments helped define the minimum concentration of FL that would be needed to achieve a

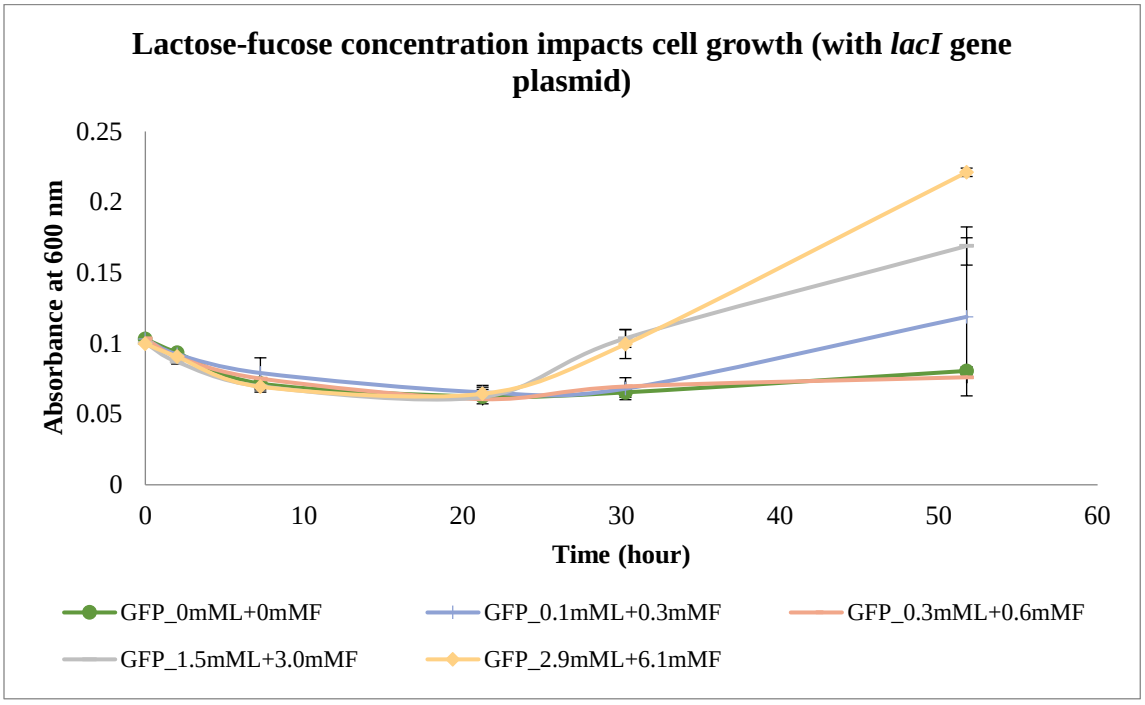


Figure 2-2: Growth curve of BL21_GFP (with plasmid) cells in the presence of varying concentrations of both lactose and fucose. Here also 1.5 mM lactose concentration allowed considerable cell growth to be confirmed as also seen in Figure 2-1 and was therefore chosen as concentration basis for further experiments. The experiment was performed using duplicates and the error bars depict for a single standard deviation from the mean value reported for biological replicates.

corresponding lactose concentration that is necessary to result in measurable changes in overall cell density/ growth in our assay setup.

$$Fucose - Lactose \overset{Enzyme}{\longleftrightarrow} Fucose + Lactose$$

Thus, 1 mole FL should yield upon hydrolysis 1 mole fucose + 1 mole lactose. Hence, further growth studies using BL21 control (no plasmids) and fucosidase/ *lacI* gene plasmid carrying strains (i.e., BL21_pEC_Tm0306_WT and BL21_pET26b+_Blon2336_WT) were based on using a minimum concentration of 2 mM 2'-FL and 3'-FL as substrates considering that the enzyme's activity is 90%.

2.2 Growth studies using two distinct fucosyllactose glycan isomers as substrates

The cells that can additionally express fucosidase enzymes can hydrolyze FL, to release lactose, which can be further broken down by the β-galactosidase produced by the *lacI* gene on the plasmid to facilitate overall cell growth on FL. The β-galactosidase enzymes can hydrolyze lactose into galactose and glucose as discussed earlier. Based on this understanding, BL21_pEC_Tm0306_WT and BL21_pET26b+_Blon2336_WT cell lines were generated that can selectively hydrolyze 2'-FL and 3'-FL, respectively, by expressing highly specific α-1,2-fucosidase and α-1,3/4-fucosidase, respectively.

A similar growth study as described earlier was conducted to show that 2'-FL and 3'-FL can selectively support the growth of cells that produce the necessary fucosidase enzymes and cells not expressing this specific enzyme would show no growth at all.

2.2.1 Materials and methods

Original glycerol stocks were inoculated into LB media BL21 control, BL21_pEC_Tm0306_WT and BL21_pET26b+_Blon2336_V374A_WT to prepare starter culture. The cultures of 10 ml volumes were grown for ~16 hours at 37°C in an incubator shaker at 200 rpm. For BL21 control, chloramphenicol was used while for the other two cultures kanamycin was used as the select antibiotic. The procedure for growing cell cultures was similar to the previous section but with a slight modification. As soon as the OD of the cells in minimal media with 2% (w/v) glucose reached the range of 0.4 to 0.6, the cultures were induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for protein expression at 37°C for 2 hours. To ensure that respective proteins were expressed when the cells were grown in minimal media; a SDS (Sodium Dodecyl Sulfate) PAGE analysis was conducted for samples taken from the cell culture. The samples for SDS-PAGE analysis were obtained from small scale protein expression of 5 ml culture volumes each. Culture volumes of 5ml each were centrifuged at 3900 rpm for 20 minutes to recover the cell pellet and lyzed using B-PER II reagent (Thermo-Fisher). Next, 10 μL samples of the original culture extract, lysed cytoplasmic soluble protein, and insoluble pelleted protein fractions obtained after reacting the cell pellets with B-PER II reagent were added into Eppendorf PCR tubes with 10 μL SDS-PAGE sample loading buffer (95% laemmli buffer, 5% β – mercaptoethanol) and denatured in an Eppendorf Mastercycler Nexus Gradient (Conditions: 95°C; 5 minutes, 10°C; 10 minutes). After loading the samples and running SDS-PAGE on a Genscript pre-cast gel at 200 V for 40 minutes, the gel was analyzed in a BioRad Gel Doc EZ-Imager after staining the gel with Coomassie dye.

Cell pellets recovered after protein expression were washed as before and resuspended in minimal media without any carbon source while keeping in an ice bath. Furthermore, prior to adding the cell cultures into 2 ml based 96 deep well plate prepared with 2'-FL and 3'-FL added in the respective wells, all the three cultures were diluted with minimal media (without any carbon source) to have the same starting OD_{600} cell density. Each well in the 96 deep well plate had a total volume of 1.5 ml which included the new carbon source of 2'-FL or 3'FL and aliquots of each cell culture in minimal media without any carbon source. These cells were grown at 37°C, 200 rpm in an incubator shaker. The growth curve profile of these cell cultures was then monitored for a total duration of ~5 days by measuring the change in OD_{600} density. Additionally, 100 µL samples taken from the cell culture for measuring OD600 were lysed using lysozyme and 2 µL of each was spot on a TLC plate at regular intervals. TLC analysis was performed on TLC Silica Gel 60 F254 Merck plates using a mobile phase of ethyl acetate: 2-propanol: acetic acid: water (3:2:1:1). Plates were sprayed with 0.1% orcinol and 10% H_2SO_4 mix followed by heating the plates for 15 minutes at 100°C to allow detection of product bands upon charring. The TLC plates were then visualized using the BioRad Gel Doc EZ-Imager.

2.2.2 Results and discussion

The growth study conducted on 2'-FL and 3'-FL as carbon sources for different cell lines for exactly 5 days and 17 hours showed very interesting results. These results provide a rationale for this thesis work and clearly show that when engineered cells provided with designer HMOs, only bacteria expressing the correct GH show increment in growth compared to other bacteria.

As seen below in Figure (2-3), 3’-FL is specifically a substrate for α-1,3/4-fucosidase enzymes, and therefore BL21_pET26b+_Blon2336_V374A_WT transformed cells showed consistently increased cell growth in minimal media (with FL only and without glucose as carbon source).

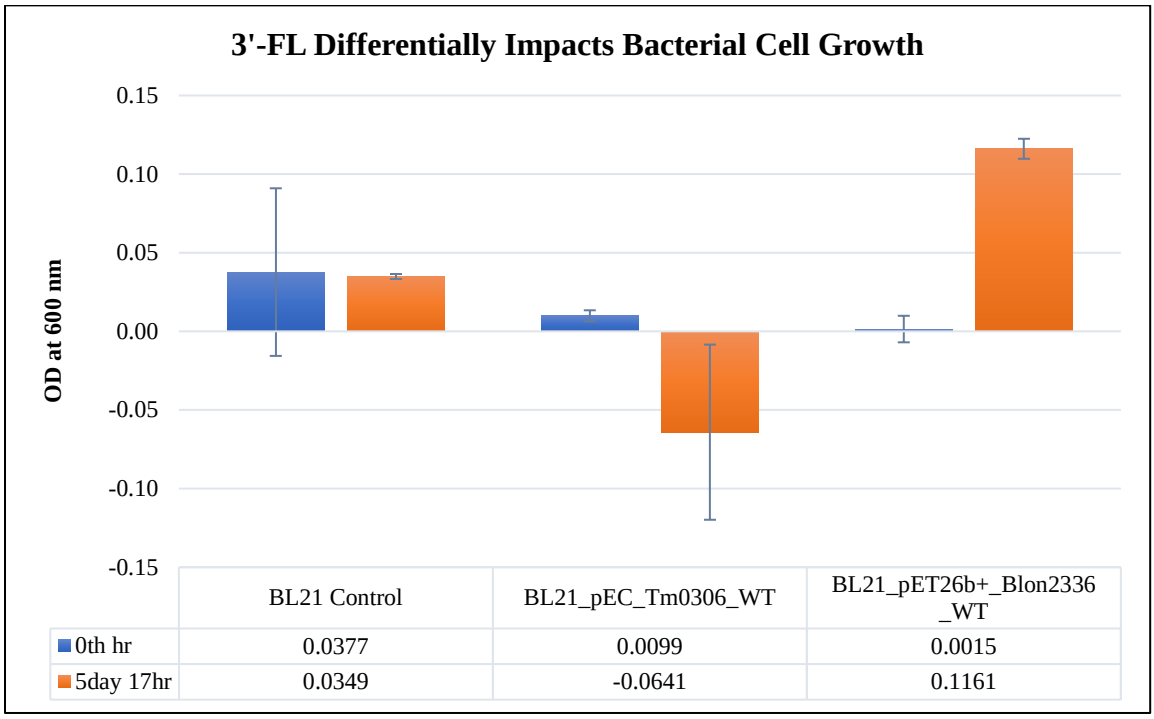


Figure 2-3: BL21 control, BL21_pEC_Tm0306_WT and BL21_pET26b+_Blon2336_WT cell growth in minimal media with 2mM 3’-FL. Note that BL21 control cultures showed no further cell growth whereas BL21_pEC_Tm0306_WT could not metabolize 3’-FL and hence displayed cell death phase indicated by drop in OD. Only BL21_pET26b+_Blon2336_WT cell cultures showed considerable growth and an increase in OD as they were able to metabolize 3’-FL due to the presence of BiAfc_WT hydrolytic enzymes in the cell cytoplasm. The experiment was performed using duplicates and the error bars depict for a single standard deviation from the mean value reported for biological replicates.

Further BL21_pEC_Tm0306_WT cells did not show any growth as these cells were unable to could not metabolize 3’-FL due to the specific nature of the expressed GH29 enzymes. As expected, BL21 control also displayed no change in growth since the control cell lines lacked any GH enzymes that could hydrolyze the oligosaccharide to facilitate lactose metabolism. In parallel we conducted another controlled cell growth study using 2’-FL as carbon source to target BL21_pEC_Tm0306_WT cell lines. However, these experiments

did not display favorable results as predicted. There are several possible reasons for this anomalous result. Firstly, the BL21_pEC_Tm0306_WT cells seem to have a problem with expression of the hydrolytic enzyme to begin with. These cells did not express TmαFuc_WT very well in its soluble form at high enough yields, so most of the protein may have remained trapped in inclusion bodies or as mostly insoluble protein [25] in cells as also seen clearly in the comassie blue dye stained SDS-PAGE gel imaging results image (Figure 2-4) for protein expression tests [26]. Therefore, while we were able to extract sufficient soluble protein from the cells for confirming *in vitro* activity, the total yield was poor compared to the Blon_2336 enzyme.

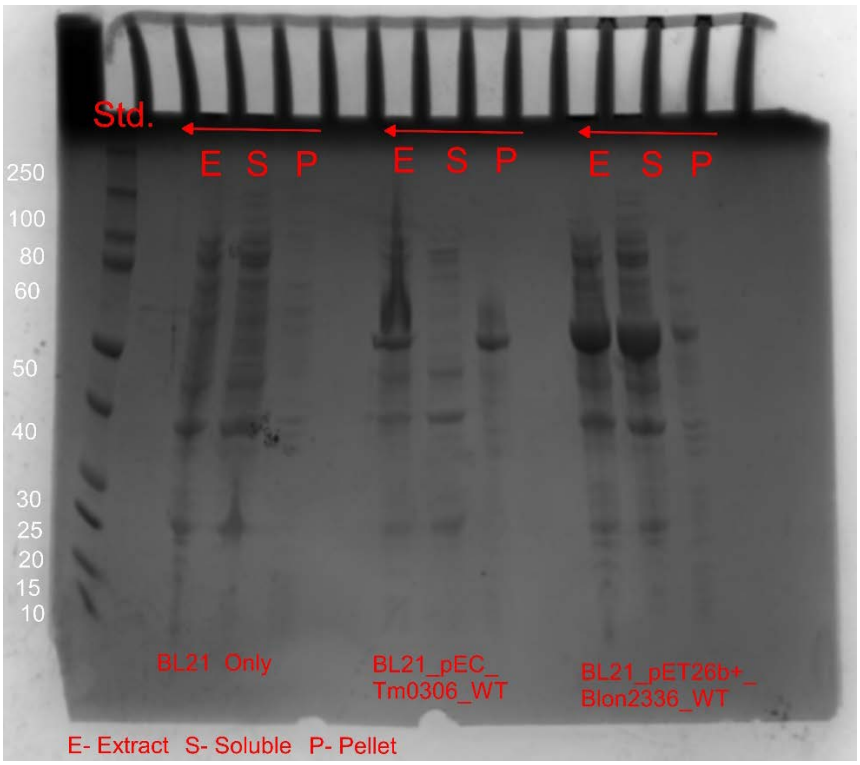


Figure 2-4: SDS Page of small scale protein expression of BL21 control, BL21_pEC_Tm0306_WT and BL21_pET26b+_Blon2336_WT cells after protein expression in minimal media with glucose using 0.1 mM IPTG for 2 hours at 37°C; TmαFuc was seen mostly in its pelleted form and BiAfc clearly shows a bold band between 50-60 kDa in the soluble form. Protein standards are shown on the left end in white (kDa). Lanes 'E' indicate the total extracted protein, lanes 'S' indicate the soluble protein and lanes 'P' indicate pelleted protein which is not available in the soluble form.

Secondly, the cells were grown at a temperature of 37°C to metabolize the FL, which is not the ideal temperature for activity of this thermophilic TmαFuc_WT enzyme. The breakdown of 2'-FL did indeed take place, but it was very slow and took more than 40 hours to show significant hydrolysis of the HMO glycan.Thus, the available cells at 0^{th} hour would have slowly died as no lactose was released due to poor activity of expressed protein at 37°C even after prolonged durations. Independent *in vitro* enzyme assays conducted with the purified TmαFuc_WT enzyme >60°C clearly showed that the expressed enzyme is indeed active but has poor activity at lower temperatures (data shown in next chapter). This confirmed that TmαFuc_WT is less active at temperatures lower than its

optimum activity temperature, as also reported previously for other *T. maritima* proteins [27].

TLC (TLC Silica Gel 60 F254 Merck plates) based analysis was further conducted to follow the metabolism of the different glycans provided to various cell cultures. TLC images (see Figure 2-5) clearly showed a distinct pattern of corresponding substrate/ product specific spots seen for TmαFuc_WT expressing cells grown on 2'-FL and BiAfc_WT expressing cells grown on 3'-FL. It was noticed that 2'-FL undergoes a slow hydrolysis to fucose and lactose due to the presence of TmαFuc_WT but only after a prolonged course of incubation (e.g., 137.5 hour). The hydrolyzed products lactose and fucose appear as faint spots on the TLC plate to support the marginal decrease in 2'-FL substrate spot density. It is also clear that the slowly- released lactose has not been broken down to galactose and glucose after such prolonged times, indicating that the cells are mostly unviable and hence show lack of growth. However, clear spots of lactose, fucose and galactose are observed in the case of breakdown of 3'-FL in the presence of BiAfc_WT for cell cultures within a short duration of only 8 hours. No spots corresponding to glucose were seen on the TLC plates, suggesting that these sugars got rapidly consumed leading to specific growth of these cell cultures.

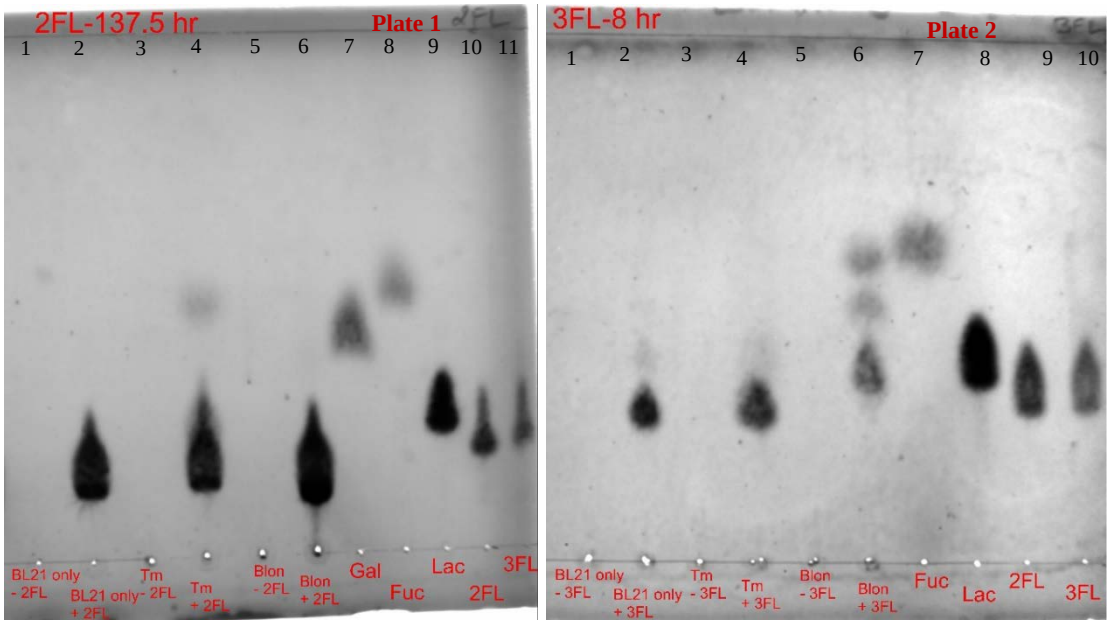


Figure 2-5: TLC plate images validate the selective consumption and breakdown mechanism of HMOs by BL21 control, BL21_pEC_Tm0306_WT and BL21_pET26b+_Blon2336_WT cell lines in the absence and presence of 2'-FL(see left plate 1)/3'-FL(see right plate 2) when cell cultures were grown at 37°C for a duration of ~137 hours and 8 hours, respectively. Plate 1 shows very minor hydrolysis of 2'-FL(lane 4) take place with BL21_pEC_Tm0306_WT cells at 37°C after prolonged growth times while Plate 2 shows complete hydrolysis of 3'-FL(lane 6) by BL21_pET26b+_Blon2336_WT into fucose and lactose. The lactose is further broken down into glucose and galactose by the β-galactosidase expressed in cells. Glucose is not visible here as a clear spot as it was rapidly consumed by the cells for growth.

2.2.3 Conclusion

Based on the growth studies performed here, it can be concluded that the selective growth of BL21_pET26b+_Blon2336_V374A_WT was clearly observed only with the appropriate substrate of 3'-FL (but not 2'-FL). We further validated that 3'-FL was hydrolyzed by expressed BiAfc_WT that facilitated released lactose metabolism by the cells. Meanwhile, in the case of BL21_pEC_Tm0306_WT, we hypothesized the observed poor cell growth on 2'-FL arises largely due to poor GH29 expression yields and low enzyme activity at mesophilic growth temperatures.

Chapter 3 : Engineering & expression of TmαFuc and BiAfc GH29 enzymes

In chapter 3 we discuss the various approaches undertaken to optimize the expression conditions for the respective TmαFuc and BiAfc proteins in *E. coli*, as well as engineering these enzymes to generate various GH29 active site nucleophile mutants for follow up activity assay in chapter 4.

3.1 *Bifidobacterium longum* subsp. *infantis* GH29 gene

The *Bifidobacterium longum* subsp. *infantis* gene (as plasmid p15TV-LIC_Blon2336_V374A) was provided as a gift by Dr. Andrzei Joachimiak from Argonne National Lab (Illinois). Details about this clone and solved enzyme structure are provided in another study published by Dr. Joachimiak [20]. The mutation to residue V374, facilitates protein crystallization, and is not located near the catalytic pocket of this enzyme. Thus, it was assumed that this mutation does not significantly affect the native enzyme's catalytic activity. Hence, no alterations were made, and the gene was moved directly into our pET26b+ expression vector and fused with 6x-Histidine tag at N-terminus using PIPE/ SLIC cloning method (refer Appendix E for PIPE/ SLIC primers used). This pET26b+ expression plasmid was then transformed into *E.cloni* cells (Lucigen) and the inserted sequence was confirmed by sending the extracted plasmid DNA to Genscript for sequencing using T7 forward and T7 reverse primers (from IDTDNA). After confirming the inserted DNA sequence, the extracted plasmid DNA was transformed into BL21 (DE3) competent cells for protein expression [28], and other mutation studies. These cells grown in LB media were aliquoted as 15% glycerol stock solutions prior to flash freezing and storage in the -80°C freezer.

Protein Expression and Purification

To extract purified protein (BiAfc_WT) from the cells upon expression, a starter culture of BL21_pET26b+_Blon2336_V374_WT transformed cells were inoculated in LB media and transferred into a larger batch after overnight cell growth at 37°C, 200 rpm in an incubator shaker. A larger batch of cell culture was grown till the exponential growth phase of the cells was attained. The OD_{600} of the cells was measured in cuvettes using a Molecular Devices Spectramax M5E cuvette reader. The cells were then induced with 1 mM IPTG for protein induction/expression at 37°C for 4 hours. Cell pellets were recovered by spinning the media down at 8000 rpm, 4°C for 15 minutes using a centrifuge. To recover the expressed protein from these wet cell pellets, the pellets were first resuspended in a suitable lysis buffer (composed of 20 mM sodium phosphate, 500 mM NaCl, 20% (v/v) glycerol; pH 7.4) also containing lysozyme and protein inhibitor cocktail (PIC). The cell lysis buffer was chosen to ensure optimum pH for lysozyme action and expressed protein stability. Lysozyme (Sigma Aldrich, USA) facilitated cell wall breakages and allowed release of the expressed protein. While PIC (1 µM E-64 from Sigma #E3132, 0.5 mM benzamidine from Calbiochem #199001, and 1mM EDTA tetrasodium dihydride) ensured that the proteases released upon cell lysis do not digest the targeted protein. For every 3 grams of wet cell pellets, 15 ml of lysis buffer, 15 µL of lysozyme and 200 µL of PIC were added. The cell pellets were slowly suspended in a 50 ml falcon tube with the above-mentioned solutions using a spatula on ice always. The frothing of solution was avoided to prevent any damage to the cells and targeted protein present in this cell lysis mixture. The cell lysis mixture was then sonicated using Misonix Sonicator 3000 at an output level of 4.5, for a 10 second pulse on time and a 30 second pulse off time for a total period of 5

minutes. The solution temperatures were never allowed to exceed 7-8°C during sonication. The cell lysate was then centrifuged at 20,000 rpm for 1 hour at 4°C to recover the supernatant from the cell lysate pellet. The supernatant was collected and filtered through 0.2 µM filter (Fisherbrand) prior to further purification the same day.

The BiAfc_WT from the supernatant was purified and desalted using an automated Bio Rad NGC FPLC system. In the purification process, first IMAC (Immobilized Metal Affinity Chromatography) was utilized using Ni^{2+} based affinity column to capture Histidine-tagged proteins [29]. The 6X- Histidine tag present on the protein of interest makes it is easier to separate this protein from the rest of background *E. coli* proteins expressed in a BL21 cell [30]. A 5 mL HisTrap FF Crude column (GE Healthcare) loaded with Ni^{2+} IMAC affinity resin was used to capture the targeted protein. First, the column was washed using 100% IMAC B solution (100 mM MOPS, 500mM Imidazole, 500 mM NaCl; pH 7.4) and then equilibrated with an equal volume of 100% IMAC A (100 mM MOPS, 10 mM Imidazole, 500 mM NaCl; pH 7.4) mobile phase. Next, the filtered cell lysate supernatant was loaded onto the column and washed using 5%:95% IMAC B: IMAC A mobile phase to remove any non-specifically bound proteins. Next, the bound his-tagged proteins were eluted off the column using a 100% IMAC B solution as the eluent. The eluted concentrated protein was then desalted, and the protein was transferred into 10 mM 2-(N-morpholino)ethane sulfonic acid (MES) at pH 6.5 using a PD-10 gravity column and aliquoted prior to storage. Concentration of the protein was estimated using Molecular Devices Spectramax M5e microplate reader at UV 280 nm (using appropriate molecular weight and extinction coefficients for the various target proteins). SDS-PAGE analysis was also conducted using the purified protein at 200 V for 40 minutes followed by stain free or

coomassie blue staining of SDS-PAGE gel. The aliquoted protein was flash frozen and stored at -80°C.

3.2 Thermotoga maritima GH29 gene

The *Thermotoga maritima* gene (*Tm0306_WT*) was custom synthesized by Genscript in a pU57_Tm0306_WT vector. Restriction cloning was used to move the gene from pUC57 to pEC plasmid using the restriction sites, AsisI and BamH1. This gene was then fused with 8x-Histidine tags at the N-terminus using standard cloning methodss. Details on pEC vector can be found in paper published by Dr. Chundawat [31]. The resulting pEC_Tm0306_WT plasmid was then transformed into *E. cloni* cells to grow cells and this plasmid was then extracted from cells using an IBI Scientific MiniPrep plasmid DNA extraction kit. The resultant plasmid was sent for sequencing using NcoI forward and T7 reverse primers (from IDTDNA) to Genscript. After confirming the DNA sequences, the plasmid was again transformed into BL21 (DE3) competent cells for protein expression and purification.

The resultant BL21_pEC_Tm0306_WT expression strain was used for all protein expression and purification work as described in the previous section. However, upon running an SDS PAGE gel and measuring the target GH29 protein concentration, it was noted that the yield of protein (TmαFuc) was extremely low. The pEC vector has a high copy number i.e. the average number of plasmid copies per cell is high [32], and thus programs the cells to produce protein at high yields. We hypothesized that the protein yield could be improved by shifting the gene of interest into a low copy number plasmid like pET26b+, which was also available in the laboratory. The section 3.1.2.1 discusses details

about the same. Further experiments were then conducted to analyze small-scale protein expression using various strains of *E. coli* with the cloned pEC_Tm0306_WT plasmid to optimize protein expression yields.

3.2.1 PIPE and SLIC cloning to create pET26b+_Tm0306_WT

The technique of Polymerase Incomplete Primer Extension (PIPE) Cloning Method along with sequence- and ligation-independent cloning (SLIC) ([33, 34]) was used to transfer the gene of interest i.e. *Tm0306_WT* (insert) from pEC to pET26b+(vector). The plasmid pET26b+_Blon2336_WT was used to extract the vector product for cloning. The PIPE procedure involves extraction of the aforementioned DNA segments. To extract the respective vector and insert DNA products, a PCR reaction was set up in PCR tubes with i) ~20 ng of DNA, ii) 10 µM forward and reverse primers (Table 2), for the vector and insert and iii) 1xMM (MM stands for PCR reactions master mix containing the high-fidelity DNA polymerases etc. from F531S ThermoFisher) in each well. To ensure no degradation of plasmid DNA and the MM, all tubes were always kept on an ice bath. The reaction was conducted in a thermomixer and annealing temperatures (shown in Table 2) close to the T_m of each primer were used. These primers specifically amplified the required segments of DNA (i.e. the vector and insert regions).

PIPE/SLIC Primers		T_m (°C)
Insert Forward	5'-TATTTCCAGGGCCATATGATCAGCATGAAGCCG-3'	52.4
Insert Reverse	5'-CGCAAGCTTCGTCATTATTCCTCCACCGC-3'	47.8
Vector Forward	5'-GGTGGAGGAATAATGACGAAGCTTGCGG-3'	52.6
Vector Reverse	5'-CTTCATGCTGATCATATGGCCCTGGAAATACAA-3'	51.2

Table 2: PIPE/SLIC primers used to create pET26b+_Tm0306_WT using pET26b+_Blon2336_WT (vector) & pEC_Tm0306_WT (insert)

DNA gel electrophoresis was conducted with the resultant PCR mix at 120 V for 40 minutes to confirm PCR reactions. The amplified DNA segments separated out from the rest of the plasmid DNA during electrophoresis and were identified on a 0.7% agarose gel under UV light after staining with ethidium bromide. The gel containing the required DNA segments was cut out from the gel under UV light after ensuring that proper personnel protective gear was worn. Gel Extraction Kit was used to extract the DNA from the cut gel which involves a warming the gel segment to melt it and filter out the DNA after a few buffer washes. The resulting vector and insert DNA concentration was measured on Molecular Devices Spectramax M5E microplate at 260 nm.

The SLIC recombination functions using an exonuclease (T4 DNA polymerase in the absence of dNTPs) to generate single-stranded DNA overhangs in the insert and vector sequences during *in vitro* reaction. The length of the single-stranded DNA generated using T4 DNA polymerase (M0203S from New England BioLabs Inc.) is controlled by the time of this reaction. These single-stranded overhangs are ligated inside the *E. coli* used for transformation resulting in gap repair and generating recombinant DNA.

To perform the SLIC experiment, first, 1 µL of DpnI enzyme (R0176S from New England BioLabs Inc.) was added to 1:2 picomolar vector-to-insert ratio and incubated at 37°C for 1 hour to digest any remaining template DNA. The resultant product was placed immediately on ice and treated with 0.5 µL of T4 DNA polymerase in 1.1 µL NEBuffer 2.1 buffer in a 10 µL reaction at 25°C for 3 minutes. The ligation mix was then transformed into *E. cloni* chemically competent cells using heat shock at 42°C for 45 seconds [35] and

grown overnight on LB agar plates with Kanamycin resistance. A small culture volume of cells was next grown in LB media using the at least 5-10 positive colonies picked from the agar plate (after additional PCR based colony screening) from transformation. The resulting plasmid DNA was extracted and sent for sequencing to identify positive clones. After confirmation, the positive plasmid DNA clones were transformed into BL21 (DE3) cells. As described before, the expression cells were next grown in LB media till the exponential growth phase and then induced with 0.1 mM IPTG for protein expression at 37°C for 4 hours. The culture media was centrifuged at 3900 rpm for 20 minutes and cell pellets isolated were utilized for small scale protein expression testing (procedure as described under section 2.2.1). Three cultures namely BL21 (DE3) control only, BL21_pEC_Tm0306_WT and BL21_pET26b+_Tm0306_WT grown in LB media at 25°C and 37°C were tested.

3.2.2 Small scale protein expression using various *E. coli* strains

The TmαFuc protein yield was low when expressed in BL21(DE3) cells using pEC_Tm0306_WT plasmid so the same plasmid. Thus, this plasmid was also transformed into other strains of *E. coli* including T7Shuffle and Rossettagami (RG2) cells.

T7 Shuffle cell expression at 25°C

T7Shuffle competent cells were kindly provided by Dr. Zhang's lab. Both plasmids, pEC_Tm0306_WT and pET26b+_Tm0306_WT were transformed into T7 shuffle competent cells using the heat shock method discussed in section 3.2.1. The positive screened colonies obtained after transformation were grown in LB media at 37°C till the

exponential growth phase for the cells was attained. Protein expression was then induced using 0.1 mM IPTG for 20 hours at 25°C. T7 shuffle only cells were grown alongside as blank for this experiment. The media were centrifuged at 3900 rpm for 20 minutes and the obtained cell pellets were used for small-scale protein expression analysis.

RG2 cell expression in TB+G media

RG2 competent cells available in the laboratory were used here. The pET26b+_Tm0306_WT plasmid was transformed into RG2 cells using the same heat shock procedure described above. A starter culture of RG2 only (as blank) and RG2_ pET26b+_Tm0306_WT cells were inoculated in LB media overnight at 37°C at 200 rpm. Next, 5% of this culture volume was transferred into TB+G Studiers auto-induction media which is an auto-induction media with 1.2% tryptone, 2.4% yeast extract, 2.3% KH_2PO_4, 12.5% K_2HPO_4, 0.375% aspartate, 2 mM $MgSO_4$, 0.8% glycerol, 0.015% glucose and 0.5% α-lactose. As soon as the OD value in the shake flasks reached the exponential stage, the temperature in the incubator is turned down to 25°C and the cultures were grown for 20 additional hours. The media was centrifuged at 3900 rpm for 20 minutes and the obtained cell pellets were then used for small-scale protein expression analysis.

3.3 Results and discussion

3.3.1 BiAfc_WT

The protein BiAfc_WT was produced using BL21_pET26b+_Blon2336_WT cells grown till an OD of 0.4-0.6 as described in section 3.1. The final purified and desalted protein yielded a concentration of 10.38 mg/ml when measured against a blank of 10 mM MES

pH 6.5 on Spectra drop microvolume plate. Beer- Lamberts' law was used to convert the absorbance reading at 280 nm to concentration in terms of mg/ml.

$$C = \frac{A_{280} \times M.W.}{\varepsilon \times l}$$

Where

C= Concentration of protein (mg/ml)

A_{280}= Absorbance at 280 nm

M.W.= Molecular weight of protein obtained using its sequence, 55567.78 g/mol

E= Extinction coefficient, 100840

l= Pathlength (cm)

3.3.2 TmαFuc_WT

The pEC vector is known to have a high copy number and hence, the yield of protein when expressed in such vectors is expected to be high which can impact soluble protein expression for some challenging proteins. However, preliminary experiments performed did not yield high concentrations of soluble protein and we were able to yield only 0.7 grams purified protein from a 2L culture. Previous research [27] based on thermostable proteins from *Thermotoga maritima* has shown that these proteins are often trapped within bacterial inclusion bodies [36]. Bacterial inclusion bodies (IBs) are functional, non-toxic protein and cell pellets aggregates often found to occur during protein expression in recombinant bacteria like *E. coli*. The mostly insoluble pelleted form of the protein contains most of the protein and yields only a small fraction in the soluble form. Hence, in order to achieve greater yield of protein, a few small-scale protein expression experiments were carried out.

Firstly, the plasmid for expression was modified to a low copy number plasmid (pET26b+) using the same host of BL21 (DE3) cells assuming it would aid in better folding of the protein and prevent IB formation. The PIPE and SLIC cloning technique was used to transfer Tm0306_WT from pEC to pET26+ vector and transformed into BL21 (DE3) cells. A small-scale protein expression study was conducted using both the vectors in BL21 (DE3) with BL21 control as blank. The proteins expressed in BL21 control cells would help in identifying the native proteins of the host cell. Protein induction was tested at 25°C and 37°C. An SDS-PAGE (Figure 3-1) was conducted which indicated both vectors resulted in protein being trapped mostly inside inclusion bodies. The bold bands appearing on P2 and P3 lanes of the gel indicate mostly pelleted form of protein.

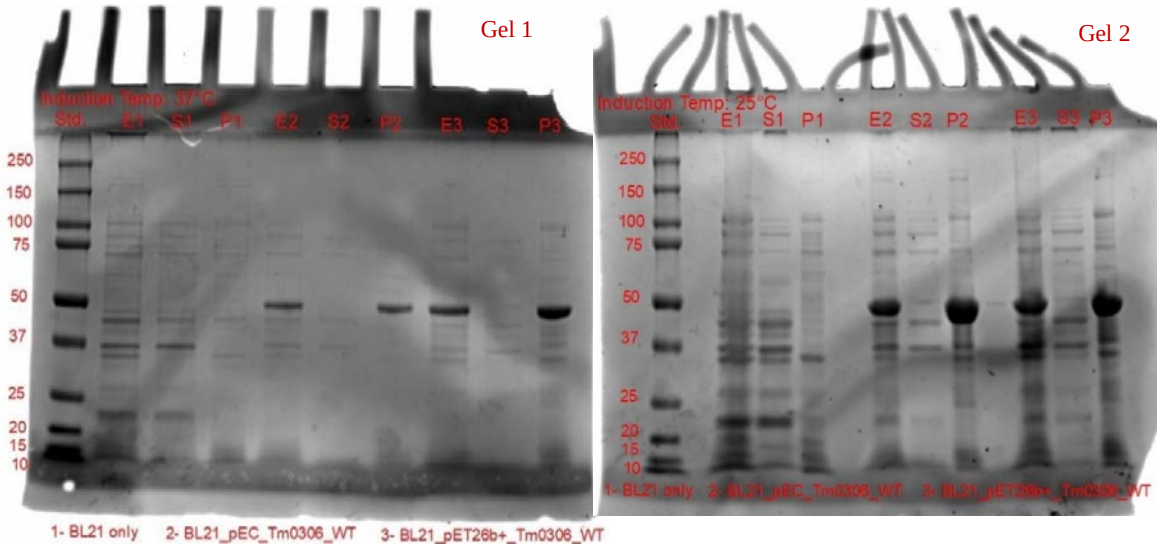


Figure 3-1: SDS-PAGE gels showing total cell extract, soluble and pelleted protein fractions of BL21 control, BL21_pEC_Tm0306_WT & BL21_pET26b+_Tm0306_WT cells from small scale protein expression studies at 25°C (see left gel 1) and 37°C (see right gel 2). In both the gels, lane 1 shows the protein ladder (units in kDa), lanes 2, 3 and 4 show the extract, soluble and pelleted proteins of BL21 control (marked on gel as E1, S1 and P1), lanes 5, 6 and 7 show the extract, soluble and pelleted proteins of BL21_pEC_Tm0306_WT (marked on gel as E2, S2 and P2)and lanes 8, 9 and 10 show the extract, soluble and pelleted proteins of BL21_pET26b+_Tm0306_WT(marked on gel as E3, S3 and P3).

Secondly, the pEC and pET26b+ created plasmids were transformed into other *E. coli* expression host cells like T7 Shuffle and RG2. In TmαFuc, C364 and C365 [37] form a

rare disulfide bridge between consecutive residues and it is known that the cysteine residues along with disulfide linkages are important for protein structure and folding [38]. Proper structural orientation and folding of the protein would conserve the protein's activity. In native *E. coli*, including BL21 (DE3) due to challenge to form disulfide bonds in the reducing environment of its cytoplasm, misfolded proteins containing disulfide bonds are likely expressed as insoluble inclusion bodies. T7 shuffle and RG2 strains of *E. coli* are engineered to be capable of correctly folding disulfide-bond containing proteins in its cytoplasm [39]. The plasmids were thus transformed into these strains of *E. coli* and expressed on a small scale. When these samples were run on an SDS-PAGE (Figure 3-2), it was found that most of the protein was still pelleted, as seen in the lanes indicating P2, P3 on the left gel and P2 on the right gel.

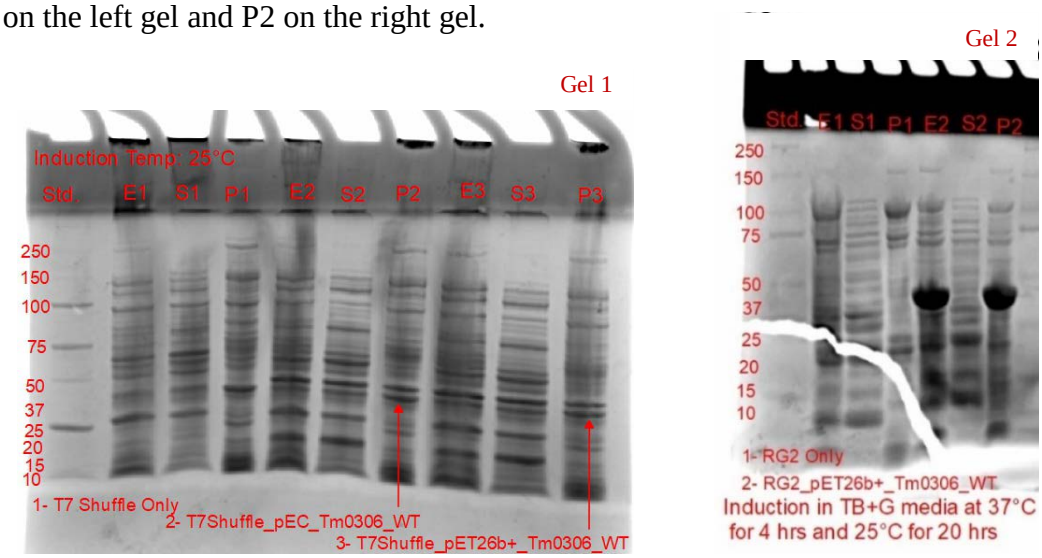


Figure 3-2: SDS-PAGE gel showing total cell extract, soluble and pelleted protein from T7 shuffle control, T7shuffe_pEC_Tm0306_WT & T7shuffle_pET26b+_Tm0306_WT cells (see left gel) and RG2 control & RG2_pET26b+_Tm0306_WT cells (see right gel) during small scale protein expression studies. In both the gels, lane 1 shows the protein ladder (units in kDa). On the left gel, lanes 2, 3 and 4 show the extract, soluble and pelleted proteins of T7 shuffle control (marked on gel as E1, S1 and P1), lanes 5, 6 and 7 show the extract, soluble and pelleted proteins of T7shuffle_pEC_Tm0306_WT (marked on gel as E2, S2 and P2) and lanes 8, 9 and 10 show the extract, soluble and pelleted proteins of T7shuffle_pET26b+_Tm0306_WT(marked on gel as E3, S3 and P3). On the right gel, lanes 2, 3 and 4 show the extract, soluble and pelleted proteins of RG2 control (marked on gel as E1, S1 and P1) and lanes 5, 6 and 7 show the extract, soluble and pelleted proteins of RG2_pET26b+_Tm0306_WT (marked on gel as E2, S2 and P2).

These experiments did not show any increase in the protein yield, so the earliest expression strain of BL21_pEC_Tm0306_WT was used to produce protein on a large-scale using the methodology discussed in section 3.1.

3.4 Site directed mutagenesis to create GH29 nucleophilic mutants

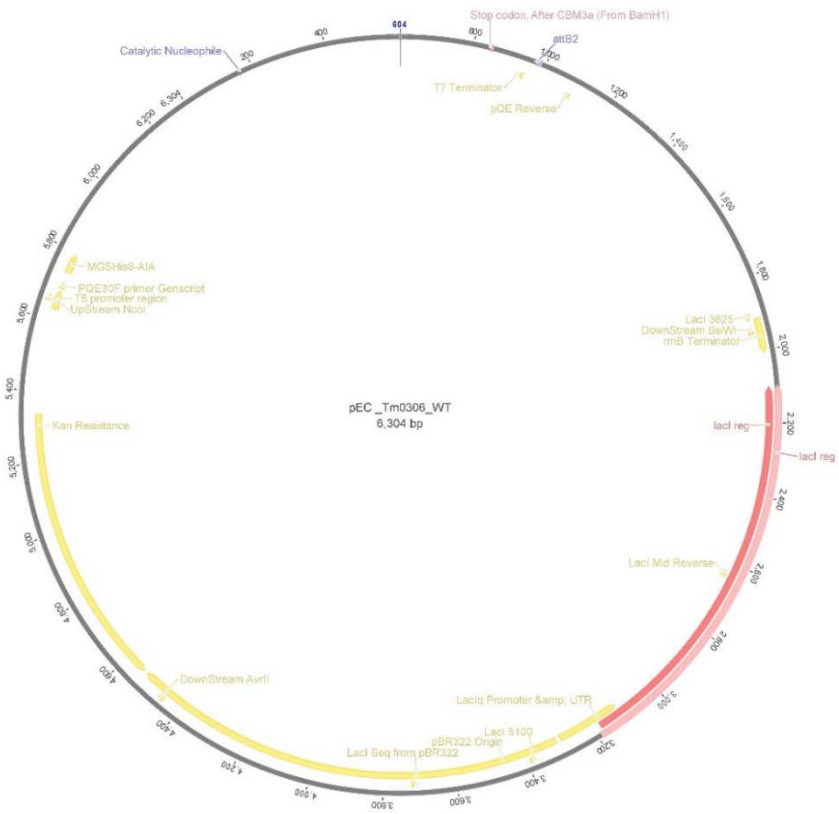


Figure 3-3: Circular plasmid map for pEC_Tm0306_WT shown using Geneious software. The catalytic nucleophile D224 (seen on the top left corner) was mutated using site directed mutagenesis to A/S/G.

Site directed mutagenesis was performed to introduce point mutations in the GH29 genes using designed primers (with the desired mutation) using a standard Quikchange type PCR based mutagenesis protocol that amplifies the entire double-stranded DNA plasmid template. A methylation-dependent endonuclease (i.e. DpnI enzyme) is used to eliminate the parent template plasmid and the resultant PCR products are next transformed into bacteria. Colonies of bacteria are then isolated from the resulting transformed cells grown on an agar plate and screened for the desired mutation using colony screening and DNA sequencing. The positive clones are sequenced to confirm the desired modification and the absence of undesired modifications. Under the current section,

the site directed mutagenesis resulted in catalytic nucleophilic mutants being created for pEC_Tm0306_WT and pET26b+_Blon2336_WT as shown in Figures (3-4, 3-5).

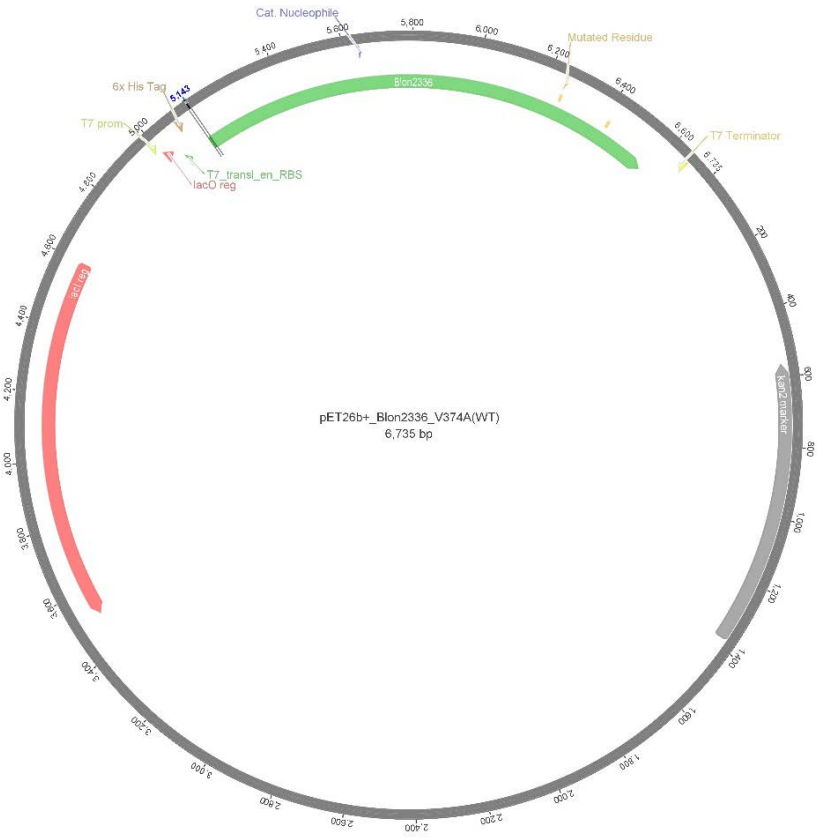


Figure 3-4: Circular plasmid map for pET26b+_Blon2336_V374A_WT shown using Geneious software. The catalytic nucleophile D172 (seen on the top) was mutated using site directed mutagenesis to A/S/G.

The Aspartic acid catalytic nucleophile in each parent GH29 i.e., namely D224 and D172, was mutated to alanine, serine and glycine using the primers shown below. These primers were designed using *Geneious 8.1.9* software by ensuring that the primers are atleast 20-35 base pairs (bp) long with melting temperatures (T_m) ranging from 55-60°C, have compositional percentage of guanine & cytosine (%GC) between 40 to 60%, the end terminals of each primer is either guanine or cytosine (which are triple bonded to each other for double stranded DNA), and lastly have a higher hairpin T_m than normal T_m. These primers sequences were sent to IDTDNA for preparation and then used after diluting from stock 100 µM to 10 µM.

	Primers used for site directed mutagenesis (SDM) of both GH29 parent enzyme constructs	
	Forward primer	Reverse primer
D224A	5'-GATGTTCTGTGGAACGCCATGGGTTGGCCGGAG-3'	5'-CTCCGGCCAACCCATGGCGTTCCACAGAACATC-3'
D224S	5'-GATGTTCTGTGGAACTCCATGGGTTGGCCGGAG-3'	5'-CTCCGGCCAACCCATGGAGTTCCACAGAACATC-3'
D224G	5'-GATGTTCTGTGGAACGGCATGGGTTGGCCGGAG-3'	5'-CTCCGGCCAACCCATGCCGTTCCACAGAACATC-3'
D172A	5'-GTCTGGCTTGCTGGCGCCAT-3'	5'-ATTGGCGCCAGCAAGCCAGAC-3'
D172S	5'-CGTCTGGCTTTCTGGCGCCAAT-3'	5'-ATTGGCGCCAGAAAGCCAGACG-3'
D172G	5'-CGTCTGGCTTGGTGGCGCCAAT-3'	5'-ATTGGCGCCACCAAGCCAGACG-3'

Table 3: Forward and reverse primers used for SDM of pEC_Tm0306_WT to D224A/S/G (rows 1, 2 & 3) and pET26b+_Blon2336_WT to D172A/S/G (rows 4, 5 & 6).

3.4.1 Materials and methods

The wild-type (WT) plasmid DNA stock was prepared after diluting the original stock obtained by extracting the plasmid from *E.cloni* cells using MiniPrep Plasmid extraction kit from IBI Scientific. The reaction volume for site directed mutagenesis includes 1xMaster Mix or 1xMM (which consists of the required DNA polymerase and nucleotides Phusion PCR Master Mix from F531S ThermoFisher), >20 ng/ µL of WT plasmid DNA, 0.5 µM of each forward and reverse primers and the rest PCR water (high purity water). The reaction was setup by adding each the aforementioned reagents in the reverse order while placing it on ice to ensure that the DNA doesn't get damaged and to prevent the polymerase from starting the reaction. 1xMM was added the last because it consists of the enzyme (polymerase), which would start reaction if the reaction mix were tranferred from

ice. The reaction mix was added to 200 µL PCR tubes, which were kept in a ThermoMixer from Eppendorf. The PCR amplification reactions conditions set were:

Procedure		Temp (°C)	Duration (sec)
Blon2336			
1	Initial denaturation	98	30
2	Denaturation	98	10
	Annealing	69 to 72	30
	Extension	72	210
3	Final extension	72	300
4	Hold	10	∞
Tm0306			
1	Initial denaturation	98	30
2	Denaturation	98	10
	Annealing	69 to 72	30
	Extension	72	190
3	Final extension	72	300
4	Hold	10	∞

Table 4: SDM reaction conditions to create Blon2336 and Tm0306 mutants

In the above Table 4, a duration of 210 seconds was allowed as the extension time keeping in view that it likely take ~30 seconds to build 1000 bp of DNA and in our case, pET26b+_Blon2336_WT plasmid DNA has 6735 bp. After completion of this reaction, the reaction mix has the native WT DNA as well as the newly constructed mutated DNA as a single-stranded chain (instead of the standard double stranded circular plasmid form). To ensure that the reaction mix does not have any of the WT DNA, DpnI digestion enzyme (R0176S from New England BioLabs Inc.) was added and the reaction was run at 37°C for 60 minutes. The DNA samples were then analyzed using a 0.7% agarose gel in 1XTAE buffer using ethidium bromide as a detecting agent. DNA gel electrophoresis was performed at 120V for about 30 minutes. Genscript confirmed the DNA sequences before

we transformed the plasmid DNA in BL21 (DE3) cells for expressing the mutant proteins for large scale protein purification, as described previously in the chapter.

3.4.2 Results and discussion

The aforementioned approach was used to create the pET26b+_Blon2336_D172A/S/G mutants using a multi-step approach. Firstly, it was confirmed by performing DNA gel electrophoresis which PCR reaction conditions clearly showed (Figure 3-6) bands in between the 6000-8000 bp range as the target DNA product is supposed to have 6735 bp even upon mutation. The DNA in the lanes, which showed proper bands, were next transformed into cloning cells to extract plasmid DNA then shipped to Genscript for DNA sequencing. Similarly, catalytic nucleophilic mutants for pEC_Tm0306_WT, i.e., D224A/S/G, were already created by in collaboration by two of the colleagues, Chandra Kanth Bandi and Ayushi Agrawal working in our laboratory using a similar procedure as described above (refer Table 4 for PCR reaction conditions).

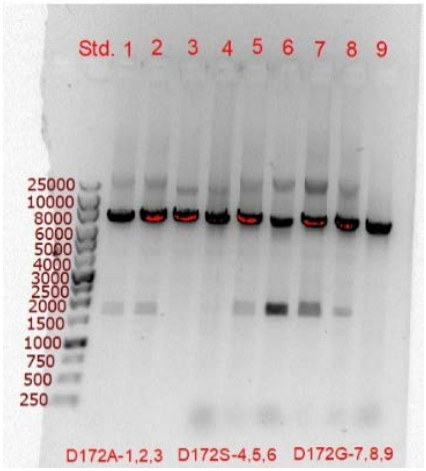


Figure 3-5: DNA Gel with Blon2336 mutants in the range of 6000-8000 bp indicating positive mutants.

Protein Concentrations (mg/ml)			
BiAfc		TmαFuc	
WT	10.38	WT	0.5
D172A	5.77	D224A	0.426
D172S	8.9	D224S	0.175
D172G	8.35	D224G	0.35

Table 5: Table showing protein concentration values obtained from large scale protein expression

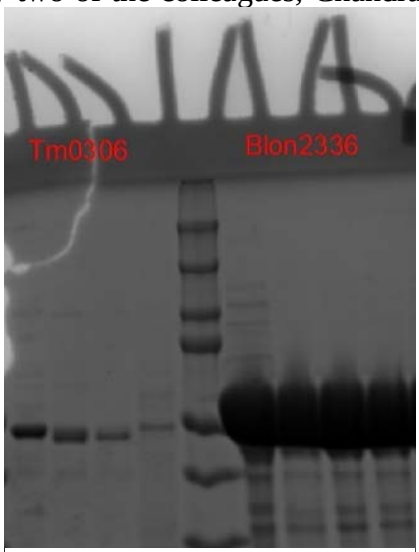


Figure 3-6: SDS-PAGE of TmαFuc_WT & BiAfc_WT and all of their mutants; BiAfc proteins clearly indicating higher concentration (right) and TmαFuc proteins showing lower soluble protein concentrations(left). The DNA ladder has been added in the center lane 5.

All sequence confirmed mutagenized plasmid DNA were stored at -80°C. The DNA was next transformed into BL21 (DE3) cells and a large volume of WT and mutant protein cultures of each respective gene were grown at 37°C till their exponential phase was attained. Then, the cells were induced for protein expression using 1 mM IPTG at 37°C for 4 hours. The expressed protein was then purified using the procedure discussed in section 3.1. The resultant protein concentrations purified from IMAC and finally desalted are listed in Table 5. Also, the SDS-PAGE image (Figure 3-7) for all purified proteins also distinctly affirms the concentration values.

3.5 Enzymatic hydrolysis

The enzymes of interest in this study are α-L-fucosidases that have the ability of hydrolyzing fucosylated oligosaccharides by cleaving off the L-fucosyl moiety at the non-reducing end with a high degree of stereospecificity. Thus, the hydrolysis and transglycosylation activity of these enzymes were further tested in the presence of various well-defined model substrates.

3.5.1 Materials and methods

Hydrolysis assays setup using 20 picomoles of each WT-original enzyme were conducted using various substrates (like pNP-fucose, 2Cl-4NP-fucose, 2'-FL and 3'-FL) at 2 mM concentrations. The experiments with a reaction volume of 100 µL were performed in a clear bottom microplate at 60°C, and 30°C, at 300 rpm for ~3.5 hour in 50 mM MES buffer pH 6 and 6.5, respectively for TmαFuc and BiAfc enzymes. TLC based analysis was conducted for samples collected at intermediate time intervals of 30 minutes, 1 hour and

3.5 hours using TLC Silica Gel 60 F254 Merck plates. The TLC mobile phase was ethyl acetate: 2-propanol: acetic acid: water (3:2:1:1). The TLC plates were visualized using the BioRad Gel Doc EZ Imager after spraying with 0.1% orcinol and 10% H_2SO_4 and heating the plates for 15 minutes at 100°C to char the substrate/ product spots for visualization.

3.5.2 Results and discussion

Based on the TLC results obtained after 3.5 hours, the terminal percentage hydrolysis yield for both the enzymes on each of the tested substrate was calculated. The absolute intensity of each spot was measured after deducting the intensity of a blank spot of the same diameter using the Bio-Rad Image Lab analysis software that was also used to image the TLC plate. The final terminal % hydrolysis of each substrate was calculated with respect to the amount of substrate hydrolyzed during the reaction compared to the substrate present at the start of the reaction.

$$\%\ \text{Hydrolysis} = \frac{(\text{Substrate}_0 - \text{Substrate}_t)}{\text{Substrate}_0} \times 100$$

% Terminal hydrolysis yield of substrates		
Total Reaction Time	3hour 36 min	3 hour 31 min
Substrate	TmαFuc_WT	BiAfc_WT
pNP-fucose	36±5	0
2Cl-4NP-fucose	30±3	59±4
2'-FL	34±3	0
3'-FL	0	100±1

Table 6: Terminal percent hydrolysis yield for various substrates in the presence of TmαFuc_WT & BiAfc_WT

These % hydrolysis values (Table 6) have been calculated from the TLC images shown in Figure 3-8 and 3-9 below. It can be inferred that TmαFuc_WT is active on the first three substrates and is inactive on 3'-FL due to the highly specific nature of this enzyme, which is an α-1,2-fucosidase. Only 2Cl-4NP-fucose and 3'-FL (as confirmed previously in [20] study) as substrates were hydrolyzed by BiAfc_WT. But this enzyme was inactive on both pNP-fucose and 2'-FL (as mentioned in literature [20]). The enzyme's inactivity on 2'-FL could be due to the active specificity for breaking α-1,3/4 fucosyl bonds.

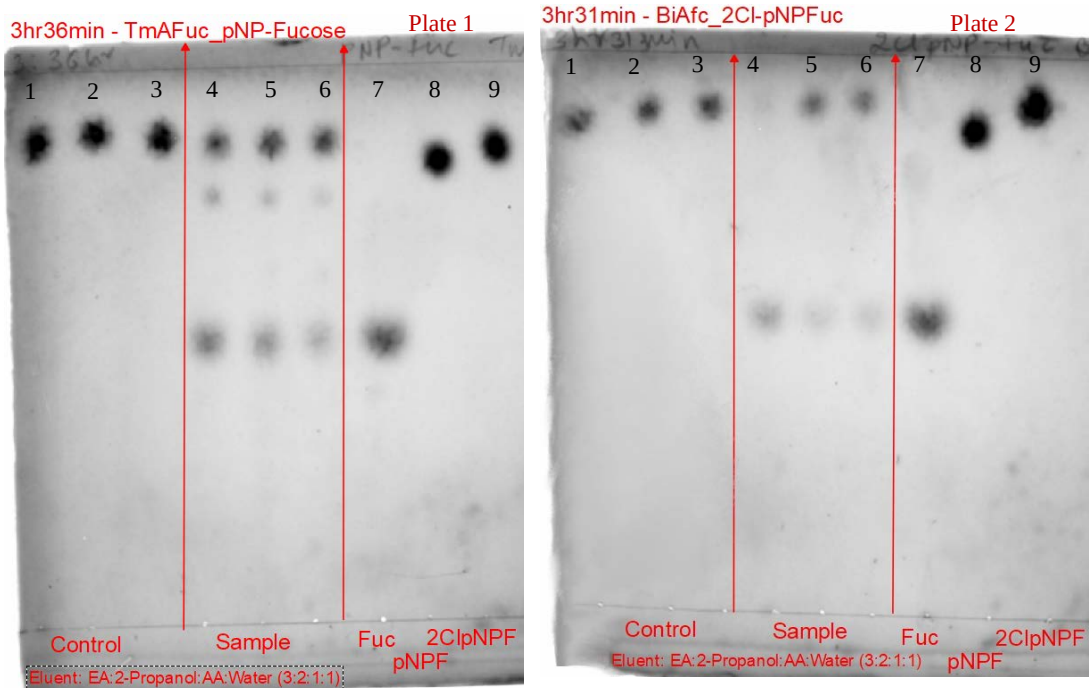


Figure 3-7: TLC images of hydrolysis reaction product profile for TmAFuc_WT with pNP-fucose (left plate 1) and BiAfc_WT with 2Cl-4NP-fucose (right plate 2) as substrates, respectively. Lane 1, 2, 3: Control (without enzyme); Lane 4, 5, 6: TmAFuc_WT or BiAfc_WT showing partial hydrolysis of either pNP-fucose or 2Cl-4NP-fucose in ~3.5 hours, respectively. Lane 7, 8, 9: fucose, pNP-fucose & 2Cl-4NP-fucose marker standards.

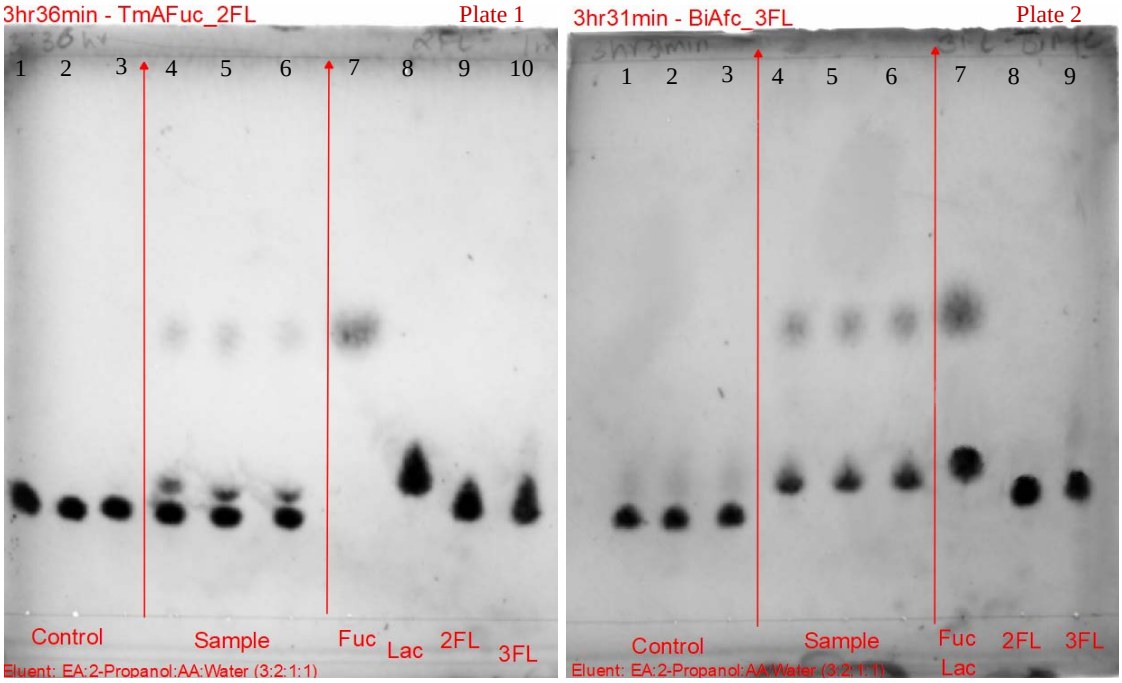


Figure 3-8: TLC images of hydrolysis reaction product profile for TmAFuc_WT with 2'-FL (left plate 1) and BiAfc_WT with 3'-FL (right plate 2) as substrates, respectively. Lane 1, 2, 3: Control (without enzyme), lane 4, 5, 6: TmAFuc_WT showing partial hydrolysis of 2'-FL and BiAfc_WT showing complete hydrolysis of 3'-FL in ~3.5 hours, respectively. Lane 7, 8, 9, 10: fucose, lactose, 2'-FL & 3'-FL marker standards.

Chapter 4 : Efforts towards creating fucosyllactose directly using TmαFuc and its various nucleophilic mutants

This section discusses the preliminary attempts made to synthesize a model FL as the secondary aim of this thesis. Based on promising hydrolysis assay results reported in section 3.5, TmαFuc was studied further to explore possible chemoenzymatic routes to synthesize either 2'-FL or 3'-FL directly using an engineered fucosynthase. As described in the section 3.4, BL21_pEC_TmαFuc_WT was mutated at its catalytic nucleophile residue to obtain BL21_pEC_TmαFuc_D224A/S/G. A chemical rescue study was conducted by Ayushi Agrawal in our lab on the same enzyme and its mutants (1 µg) in the presence of 2M sodium azide and sodium formate using 1 mM pNP-fucose as substrate for 1.5 hours total duration. In that study, it was found that the hydrolytic enzyme activity of the D224G mutant increased 8-fold (over the mutant containing no external nucleophiles to rescue the hydrolytic activity) in the presence of azide anions, whereas no recovery of activity was found in the presence of formate. This finding was not in line with the results reported in a previous study studying the crystal structure of the same *T. maritima* α-L-fucosidase. Interestingly, we found that the D224A and D224S mutants were both not able to regain their activity in the presence of either external nucleophile. Based on this finding, the initial experiments were performed using only D224G mutant. It was further verified by us that the D224A and D224S mutants also did not show any recovery of hydrolytic activity even while testing other nucleophiles for chemical rescue.

4.1 Materials and methods

4.1.1 Chemical rescue of TmαFuc and its mutants using varying concentrations of sodium azide

All enzyme assays of total volume 100 µL were performed using 2 mM pNP-Fucose as the substrate and 20 picomoles of protein (WT & D224G mutant) along with effective molar concentrations of sodium azide ranging from 0.1- 2 M in 50 mM MES (pH 6). The experiments were performed at 60°C at 300 rpm for 2 hours. After terminating the reaction at 2 hours, NaOH solution was added to monitor the released pNP groups spectrophotometrically. This was done to drive forward the formation of corresponding phenoxide anion at alkaline conditions from pNP that was released during the reaction. Briefly, 100 µL of 0.1 M NaOH was added to 20 µL of reaction mixture sample diluted along with 100 µL of water at 37 °C. The increase in pH of the reaction medium thereby increase the yellow colored pNP anion appearance in the wells. This facilitates detecting pNP in solution by reading well absorbances at 405 nm. For determining pNP concentration, varying concentrations of pNP standards were prepared by serial dilution and their absorbances were measured at 405 nm using a Molecular Devices Spectramax M5E microplate reader. The calibration curve generated for the standards was then used to calculate the pNP released for unknown samples.

4.1.2 Chemical rescue of TmαFuc and its mutants using various external nucleophiles

In search of better leaving groups, additional external nucleophiles were tested for chemical rescue activity of all the three catalytic nucleophilic mutants of TmαFuc_WT. These include nucleophiles that are commonly reported in the literature such as sodium azide,

sodium formate, imidazole, sodium acetate, sodium thiocyanate, sodium bromide and sodium cyanate. An enzyme assay using 4 mM pNP-Fucose as substrate and 20 picomoles of each enzyme was performed using 1 M fixed concentrations of individual external nucleophiles. The experiment was conducted using 50 mM MES buffer at pH 6 at 60°C at 350 rpm for a duration of 18.5 hours. Next, 0.1M NaOH solution was added as before to detect pNP released in the assay as described previously.

TLC plate-based analysis was also conducted after 18.5 hours using TLC Silica Gel 60 F254 Merck plates using a mobile phase of ethyl acetate: 2-propanol: water (70:20:10) for three of the nucleophiles namely; azide, imidazole, and cyanate. The TLC plates were then visualized using the BioRad Gel Doc Imager after spraying 0.1% orcinol and 10% H_2SO_4 to char the spots darker upon heating the plates for 15 minutes at 100°C.

A 3,5-dinitrosalicylic acid (DNS) based reducing sugars assay was also performed on the samples after 18.5 hours. DNS assay allows quantifying the released fucose reducing sugar that should be directly to the amount of pNP released pNP-fucose. The DNS reagent reacts with reducing sugars (in this case fucose) to indirectly form 3-amino-5-nitrosalicyclic acid which strongly absorbs light at 540 nm. This reaction also changes the DNS reagent color from yellow to dark red in the presence of fucose sugars upon heating. Thus, 60 µL of DNS stock reagent was added to each reaction well in a microplate along with 30 µL of sugar standard or fucose sample. The 96 well plate was sealed and heated using a thermocycler to 95°C for 5 minutes and then cooled to 10°C for 10 minutes. Next, 36 µL of the total reaction volume was transferred to a clear bottom 96 well plate diluted along with 160 µL

of DI water. The absorbance of each well was read using a Molecular Devices Spectramax M5E microplate reader at 540 nm. A standard curve for varying concentrations of known concentration fucose standards was also generated to estimate the amount of fucose released from pNP-fucose during reaction for unknown samples.

4.1.3 Glycosynthase reactions of catalytic nucleophiles of TmαFuc

Glycosynthase (GS) reactions were setup, intending to synthesize FL, using 2 mM β-fucosyl azide as donor sugar and 10 mM lactose or pNP-lactose as the acceptor sugar. The donor: acceptor molar ratio was fixed at 1:5 with 100 picomoles of TmαFuc_WT or D224G mutant enzyme added in each well of total volume 50 µL. The reaction was performed using 50 mM MES buffer at pH 6 at 60°C and 300 rpm for a total duration of 72 hours. Reactions were monitored by taking samples intermittently during the reaction (24, 48, 72 hours) for TLC analysis. We used TLC Silica Gel 60 F254 Merck plates and mobile phase of ethyl acetate: 2-propanol: acetic acid: water (3:2:1:1) at regular intervals of 0^{th} hour, 24 hours, 48 hours and 72 hours. These special TLC plates can be readily viewed under UV light to allow detection of fluorescent bands that either correspond to the pNP-lactose substrate or released pNP or other unknown pNP containing product. Following the UV based TLC plate analysis; the plates were then stained for detection of sugars as per standard orcinol/ acid detection procedures described in section 2.2.2.

4.2 Results and discussion

4.2.1 Chemical rescue of TmαFuc and its mutants using varying concentrations of sodium azide

The amount of pNP released in the unknown reaction samples was calculated based on the calibration curve that relates pNP concentration to the absorbances at 405 nm by a multiplication factor of 1.1096. Based on the NaOH assay performed after two hours of chemical rescue reactions of TmαFuc_WT and its D224G catalytic nucleophile mutant, it was observed that the D224G regains its hydrolytic activity up to 13% compared to 4% activity in the absence of any an external nucleophile (see Figure 4-1). The molar concentrations of azide ranging from 0.5-2 M gave nearly identical recovery of activity. As the 1 M sodium azide result showed the minimal error and was an intermediate value in the range tested, this concentration was chosen to perform further experiments. It was concluded that azide likely would behave as a good leaving group if it was to be used in a GS-type reaction with β-fucosyl azide as the donor sugar.

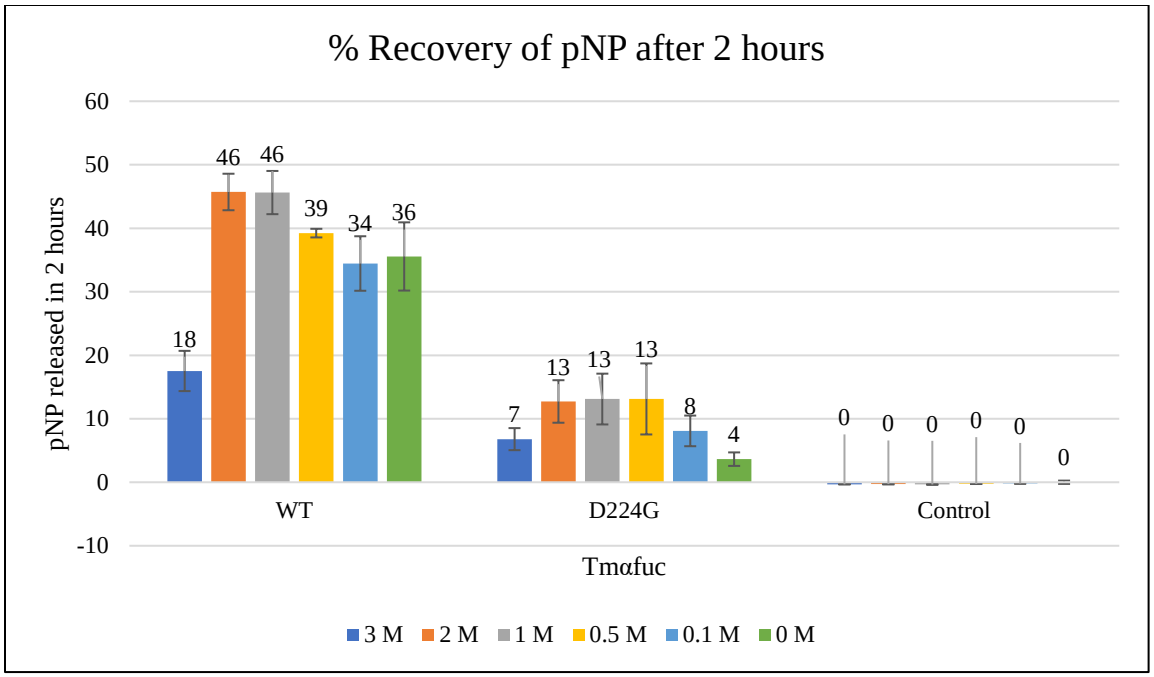


Figure 4-1: Chemical rescue of 4 mM pNP-fucose hydrolysis in the presence of varying concentrations of sodium azide (0-2M) shown here. Sodium azide allowed 13% recovery in hydrolytic activity on pNP-fucose for the D224G mutant. Based on studying the effect of nucleophile concentration on chemical rescue, 1M azide concentration was chosen for further experiments. Here, control has the respective substrates and nucleophiles without the enzyme. The experiment was performed using duplicates and the error bars depict for a single standard deviation from the mean value reported for biological replicates.

However, as the % recovery was not high enough and inconsistent with published work, this warranted further research on the impact of other external nucleophiles on recovery of the catalytic nucleophile mutant activity. This study was also performed to reaffirm the conclusions drawn from the experiments using azide as the nucleophile.

4.2.2 Chemical rescue of TmαFuc and its mutants using various external nucleophiles

Based on the results obtained from this experiment (Figure 4-2), it can be stated that the mutation D224G showed the best chemical rescue results among all other catalytic nucleophile mutants. This could be because glycine is a special amino acid (which has H as the side group) that makes it easier for the external azide group to attack the anomeric

carbon of pNP-fucose and rescue the mutant enzyme activity. This observation is not valid for the other two tested amino acids (Figure 4-3) likely due to steric hindrance. However, the underlying mechanism that can explain why certain mutations at the active site nucleophilic residue are amenable to chemical rescue is still lacking and would require additional work to be conducted (e.g., QM/MM simulations).

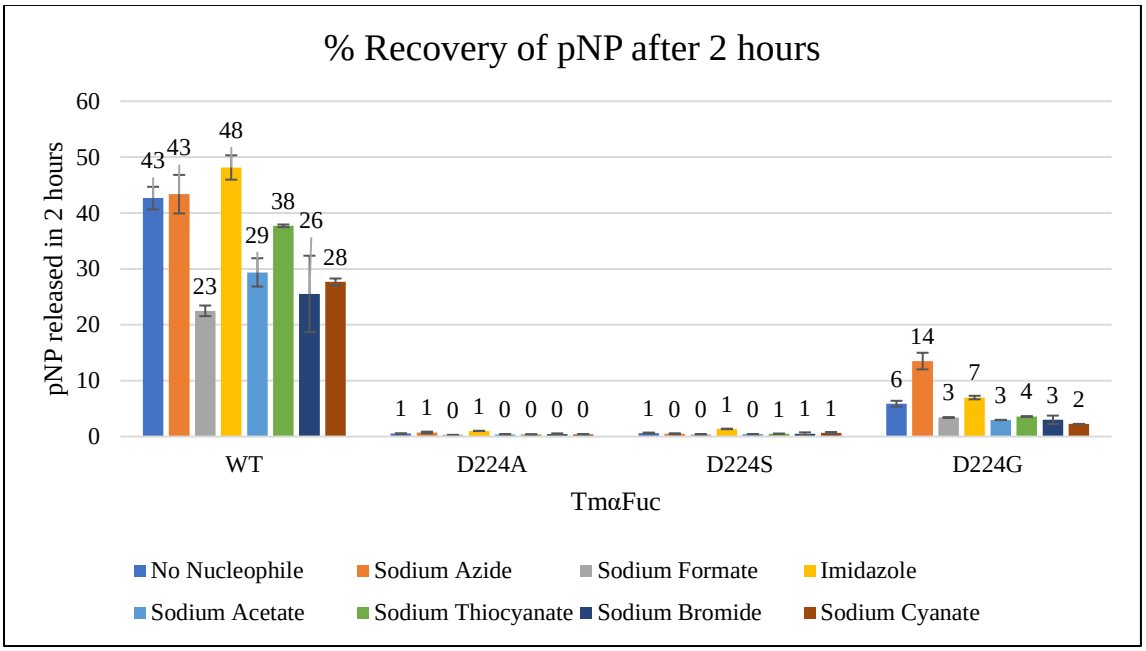


Figure 4-2: Chemical rescue of 4 mM pNP-fucose hydrolysis in the presence of various external nucleophiles (at fixed 1 M concentration) shown here. Prominently, azide and imidazole alone gave the highest (14% and 7%, respectively) activity recovery with D224G mutant. The experiment was performed using duplicates and the error bars depict for a single standard deviation from the mean value reported for biological replicates.

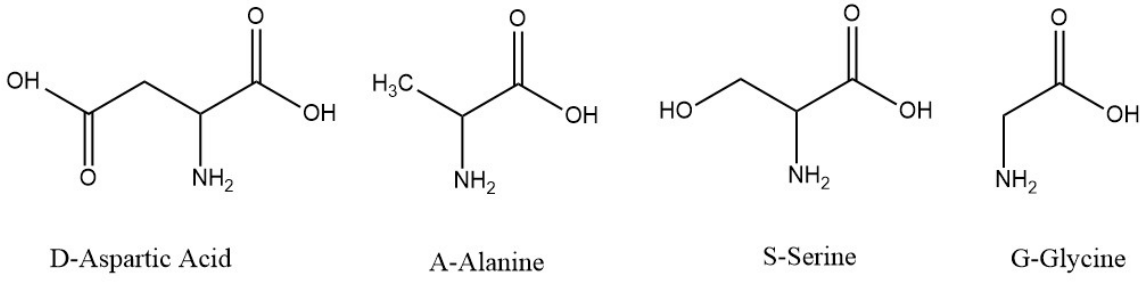


Figure 4-3: Relevant amino acid structures; D is the original catalytic nucleophile of TmαFuc at 224th position and it was mutated to A, S and G (drawn using ChemDraw).

While, it is possible that the various external nucleophiles might be effective for chemical rescue only within a certain concentration range, our preliminary analysis at a fixed 1 M concentration loading suggests that amongst all the nucleophiles tested, only azide and imidazolium ions seemed to revive the hydrolytic activity of the mutant enzyme to 14% and 7%, respectively (Figure 4-2). Azide acts as the better external nucleophile among the two candidates. Interestingly, both ions also seemed to increase the hydrolytic activity of the wild-type enzyme. Hence, this can explain why azide behaves as a good leaving group when present as β-fucosyl azide. This has already been tested using the same D224G mutant based GS that added fucose group to pNP-xylose acceptor sugar [40]. Imidazole on the other hand could also behave as a good leaving group [41, 42] besides azide due to the presence of the deprotonated imidazole group acting as a

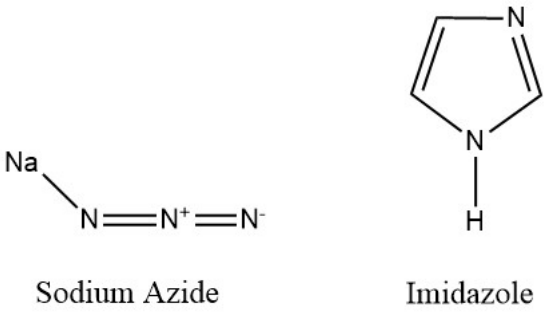


Figure 4-4: Chemical structure of effective external nucleophiles (drawn using ChemDraw).

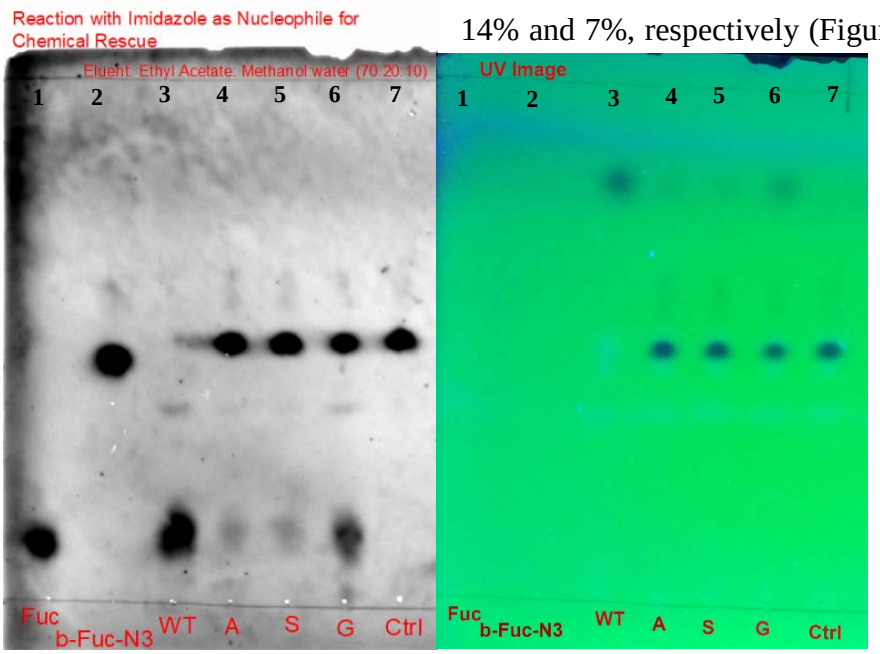


Figure 4-5: Chemical rescue using imidazole as external nucleophile after 18.5 hours, lane 1: fucose marker, lane 2: β-fuc-N_3, lane 3, 4, 5, 6: samples with TmαFuc_WT, D224A, D224S, D224G, lane 7: control, left figure: TLC plate, right figure: Image under UV light to visualize the pNP groups. Lane 6 clearly shows hydrolysis of pNP-fucose as we see spots in the same vertical height as the marker 1 indicating presence of fucose and release of pNP as seen on the UV image of the plate. This suggests that imidazole acts as a good external nucleophile.

nucleophile. In our study (refer Figure 4-5), the D224G mutant was able to hydrolyze pNP-fucose also in the presence of imidazole as indicated by the presence of a clear spot confirming fucose released as observed in lane 6 of the TLC image. This enhanced action of imidazole was likely resulting due to its small size and structural compatibility that allows it to readily access the enzyme active site. Also, imidazole has a pKa value of 6.95 and hence this molecule is expected to be mostly deprotonated at pHs higher than neutral pH to display increasing basicity. Note that imidazole is also chemically similar to Histidine and might be able to form stacking interactions with other aromatic amino acids within the active site to facilitate chemical rescue unlike other external nucleophiles.

Further experimentation and modeling will be needed to reveal detailed structural features between the enzyme and imidazole that are important for chemical rescue. It is possible that trace amounts of imidazole could be present along with the IMAC purified protein if the desalting/buffer exchange step is not performed properly (refer section 3.1). This potential pseudo-nucleophile could result in higher % hydrolysis for the mutant proteins. In order to avoid this issue, all protein purification experiments were carefully carried out to make sure the desalting columns and all FPLC plumbing lines were properly flushed with the elution buffer to avoid any imidazole contamination during protein elution.

After running the reaction for 18.5 hours, both a NaOH assay (to monitor pNP release) and a DNS assay (to monitor fucose release) were performed on identical reaction samples to close the mass balance on pNP-fucose hydrolyzed using two independent assays. For all samples, the concentration of pNP (in molar equivalents) obtained from the NaOH assay was greater than the concentration of fucose obtained from DNS assay. This is likely to have happened because of under prediction of the low concentrations of fucose using DNS assay. The assay was performed using fucose standards in the range of 0.25 to 5 g/L, i.e., 1.5 to 30.5 mM (Figure 4-4). However, considering most of the fucose released would be in the range of 0 to 4 mM, the DNS assay method is less sensitive and likely results in a lower prediction of fucose than present. Since 1 mole of pNP-fucose upon hydrolysis is expected to yield equimoles of pNP and fucose, the differences in concentration from both assay methods was assumed to be largely due to the lower sensitivity of DNS assay in the performed experimental range. Only cyanate anions showed an anomalous interference during DNS assay, as the blank controls without fucose also resulted in a color change of the DNS reagent. However, it is also possible that the reducing sugar reacts with the external nucleophile in some cases to give a lower response as well. To further confirm hydrolysis of pNP-fucose during chemical rescue experiments, we ran TLC analysis as well.

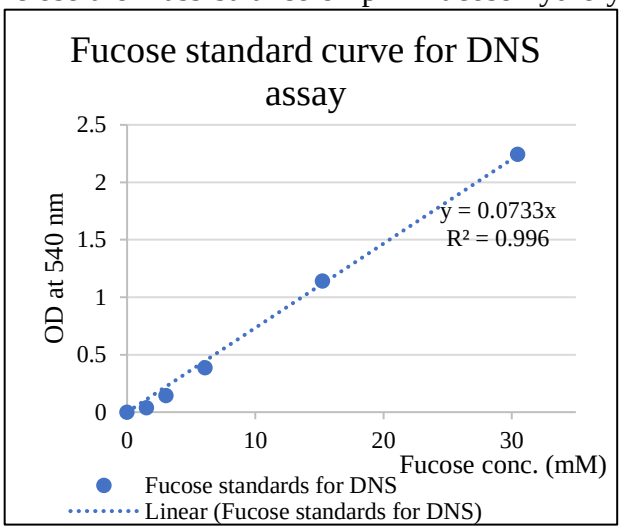


Figure 4-4: Fucose standard curve for DNS assay

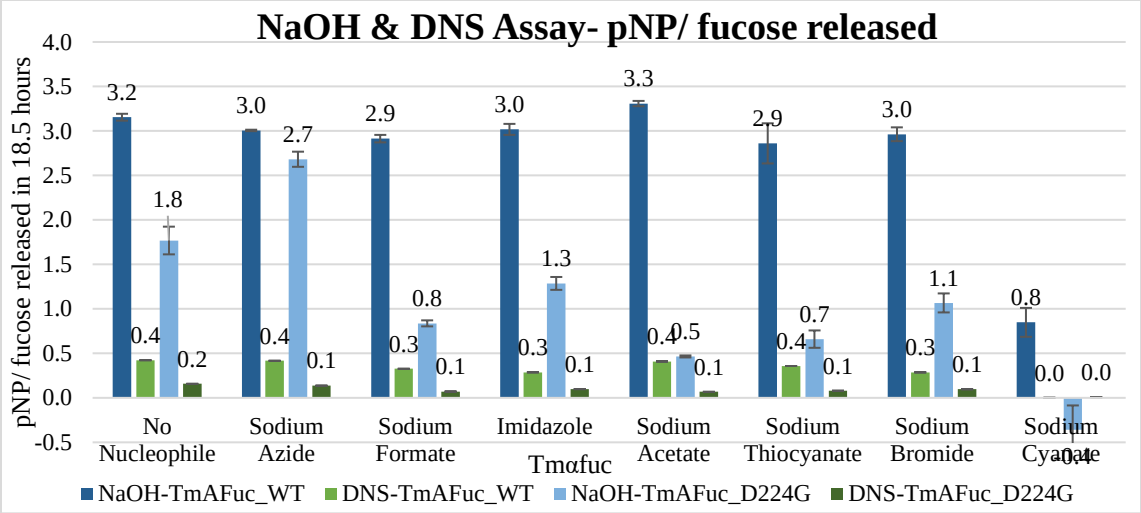


Figure 4-5: Comparison of amount of pNP released to fucose released with NaOH and DNS assay, respectively for chemical rescue experiments with TmαFuc_WT and D224G mutant in the presence of 4 mM pNP-fucose. The experiment was performed using duplicates and the error bars depict for a single standard deviation from the mean value reported for biological replicates.

When using sodium azide as an external nucleophile with the mutant enzyme, it was observed that 2.7 mmoles of pNP and 0.1 mmoles of fucose were released. From this, it could be inferred that this difference in pNP-fucose hydrolysis prediction could arise partly due to the insensitivity of the DNS assay but also due to secondary reactions of the released fucose with added external nucleophiles in some cases. The released fucose was assumed to react with the azide ions

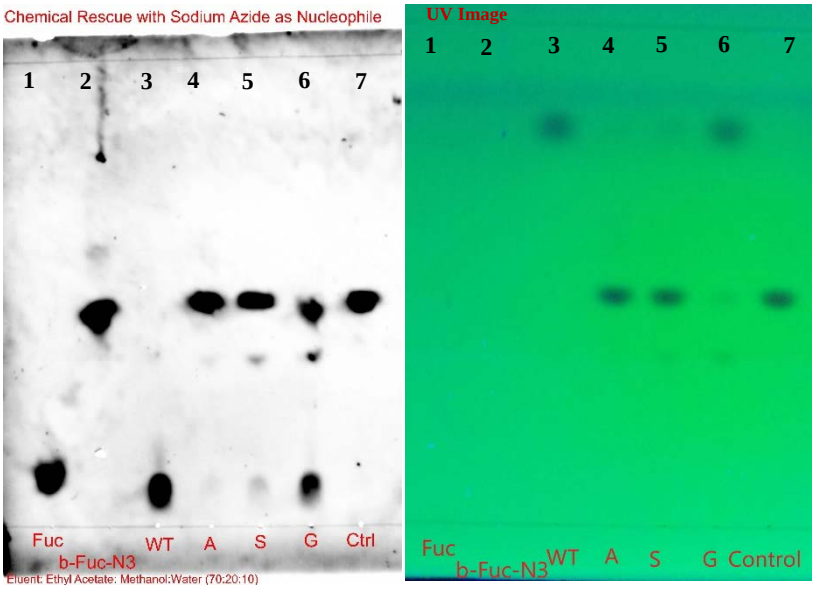


Figure 4-6: Chemical rescue using azide as external nucleophile after 18.5 hours, lane 1: fucose marker, lane 2: β-fucosyl azide, lane 3, 4, 5, 6: samples with TmαFuc_WT, D224A, D224S, D224G, lane 7: control (without any enzyme). Lane 6 clearly shows a spot in the same vertical height as the lane 2 marker indicating the formation of β-fucosyl azide during the chemical rescue experiment.

present in the solution resulting in the formation of β-fucosyl azide as confirmed using TLC analysis (Figure 4-6). These results also suggest that β-fucosyl azide could be a potential donor sugar for mutant fucosynthases to synthesize FL.

4.2.3 Glycosynthase reactions of catalytic nucleophilic mutants of TmαFuc

It was inferred from the chemical rescue experiments (section 4.2.1 and 4.2.2) that the D224G enzyme could regain hydrolytic activity in the presence of sodium azide. Thus, β-fucosyl azide was used as a donor for the glycosynthase reaction. Lactose and pNP-lactose were first tested as acceptors with the D224G mutant protein to synthesize fucosyl lactose. This glycosynthase reaction was expected to produce 2'-FL since TmαFuc protein is an α-1,2-fucosidase. Albeit, even after 72-hours of incubation time the TLC plate analysis did not suggest formation of any products of TG or GS reaction (Figure 4-7).

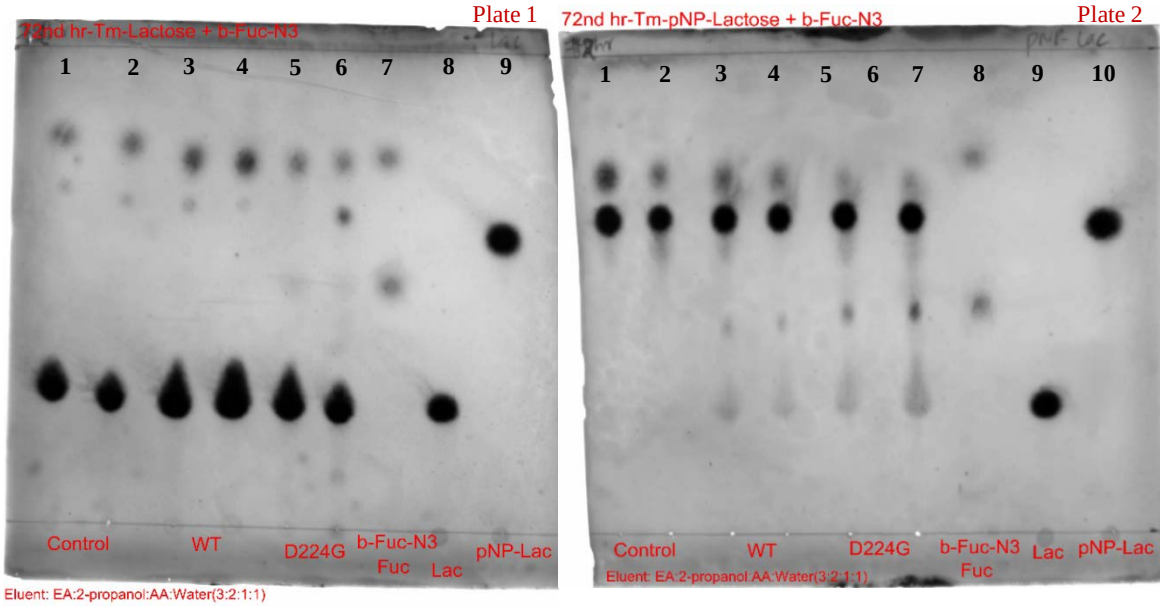


Figure 4-7: TLC plate image of glycosynthase reaction using β-fucosyl azide as donor and lactose (plate 1-see left)/ pNP-lactose (plate 2- see right) as acceptor with WT and D224G mutant. The markers fucose (lane 7), β-fucosyl azide (lane 7), lactose (lane 8) and pNP-lactose (lane 9) were also spotted. The experiment was conducted in duplicates and lanes 1, 2 are control without any enzyme, lanes 3, 4 are samples with both donor and acceptor with TmαFuc_WT enzyme and lanes 5, 6 are samples with both donor and acceptor with TmαFuc_D224G enzyme, respectively for each plate. Both plates show no GS products, the expected products are fucosyllactose and fucosyllactose-pNP, respectively for each plate.

In case of reaction with β-fucosyl azide and lactose in the presence of D224G enzyme, the expected GS product was 2'-FL which runs a little lower than lactose on a TLC plate. For the enzymatic reaction between β-fucosyl azide and pNP-lactose, the expected product was 2-fucosyl-4-nitrophenyllactose. The expected GS products were not obtained. Although, this was planned to be a GS reaction for the D224G mutant, TmαFuc_WT was used here as control to check if the wild-type enzyme showed any TG products during the time course. However, no TG products were detected for the wild-type enzyme as well. It was hypothesized that additional mutations were necessary to facilitate better docking of the desired acceptor sugars. The next sets of studies, using bioinformatics and *in silico* ligand docking modeling, were devoted to determining additional mutations for the D224G enzyme that would facilitate GS/TG reactions.

Chapter 5 : *In-silico* modeling study for identification of additional mutations

5.1 Bioinformatics

To investigate the evolutionary structural relationship of TmαFuc and BiAfc among other enzymes of the GH29 family, a high speed multiple sequence alignment (MSA) program meant for large DNA alignments named Multiple Alignment using Fast Fourier Transform (MAFFT) was used. To perform MSA, first all available GH29 family bacterial proteins (2835) were extracted from the CAZy database as a fasta file and the redundant sequences were detected and eliminated using 'Find Duplicates' tool in the Geneious software leaving a total of 2238 protein sequences. These sequences were aligned using the 'MAFFT' alignment type under 'Multiple Align' using Geneious version 11 software (Figure 5-1).

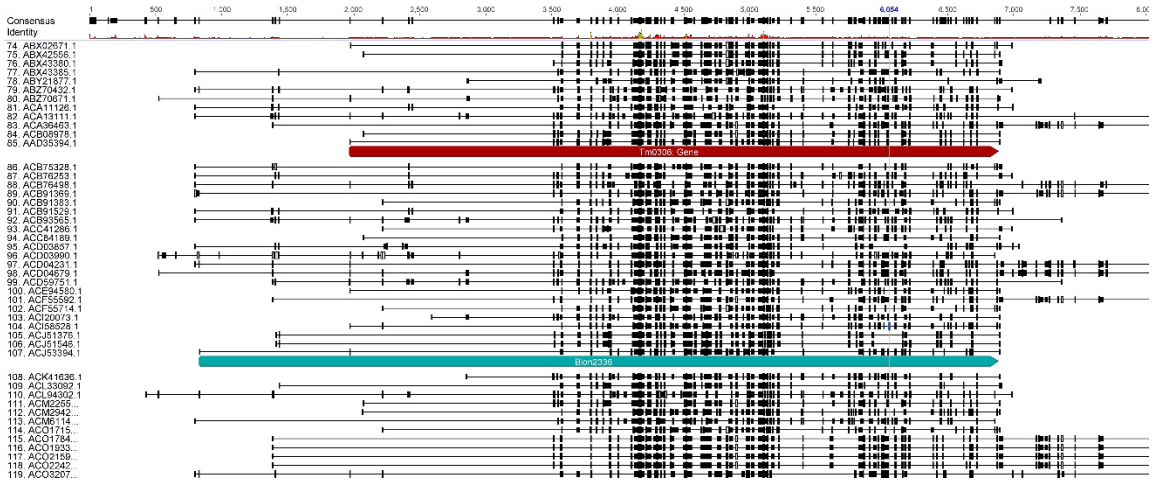


Figure 5-1: MAFFT sequence alignment of 2238 protein sequences of the GH29 family using Geneious v.11.

In an MSA, the targeted protein sequences (Tm0306 & Blon2336) are compared to a group of other sequences from evolutionarily related organisms that belong to the same family i.e. GH29 members which perform similar functions in different organisms [43]. Proteins from the same family share the main characteristic sequence features and the tertiary

structure. A specific set of amino acid residues within the protein family are likely conserved to conserve protein function in a set of corresponding species/target clades. Alignment techniques for MSA have been developed to be able to evaluate the degree of similarity of these GH29 sequences [44]. Further, these MSA are used to build a phylogenetic tree (Figure 5-2) that gives us a more detailed overview of the taxonomic diversity to help understand the conserved sequence characteristics among sub-clades. The range of characteristics represented by a subset of species provides a more detailed understanding of a species (or corresponding target proteins) ability to adapt to changing conditions and their divergent evolution. This type of analysis can be helpful to identify conserved structural/functional residues in GH29 family that could aid in picking residues for mutagenesis to make a more efficient GS/TG. Here, a phylogenetic tree was generated by means of the Neighbor-Joining (NJ) method using the aligned sequences. This method produces a unique final tree under the principle of minimum evolution based on distance of the specific sequence's tip to the root of the generated tree [45] using an algorithm that uses lesser time to analyze a large number of sequences compared to other commonly used methods like Maximum Likelihood and Maximum Parsimony. A study conducted in 1991 by Vane-Wright *et al.* [46] suggests that this level of diversity can be designated by the cladistic (phylogenetic) relationships amongst proteins/species as observed for Tm0306 (highlighted in red) and Blon2336 (highlighted in green) shown in Figure (5-2). Here, a clade is a group that includes a common ancestry and the branches are segment of the tree, lying between two nodes, where the nodes can be terminal nodes or internal branching points [47]. In the developed phylogenetic tree, both the proteins are part of two different clades (Figure 5-2).

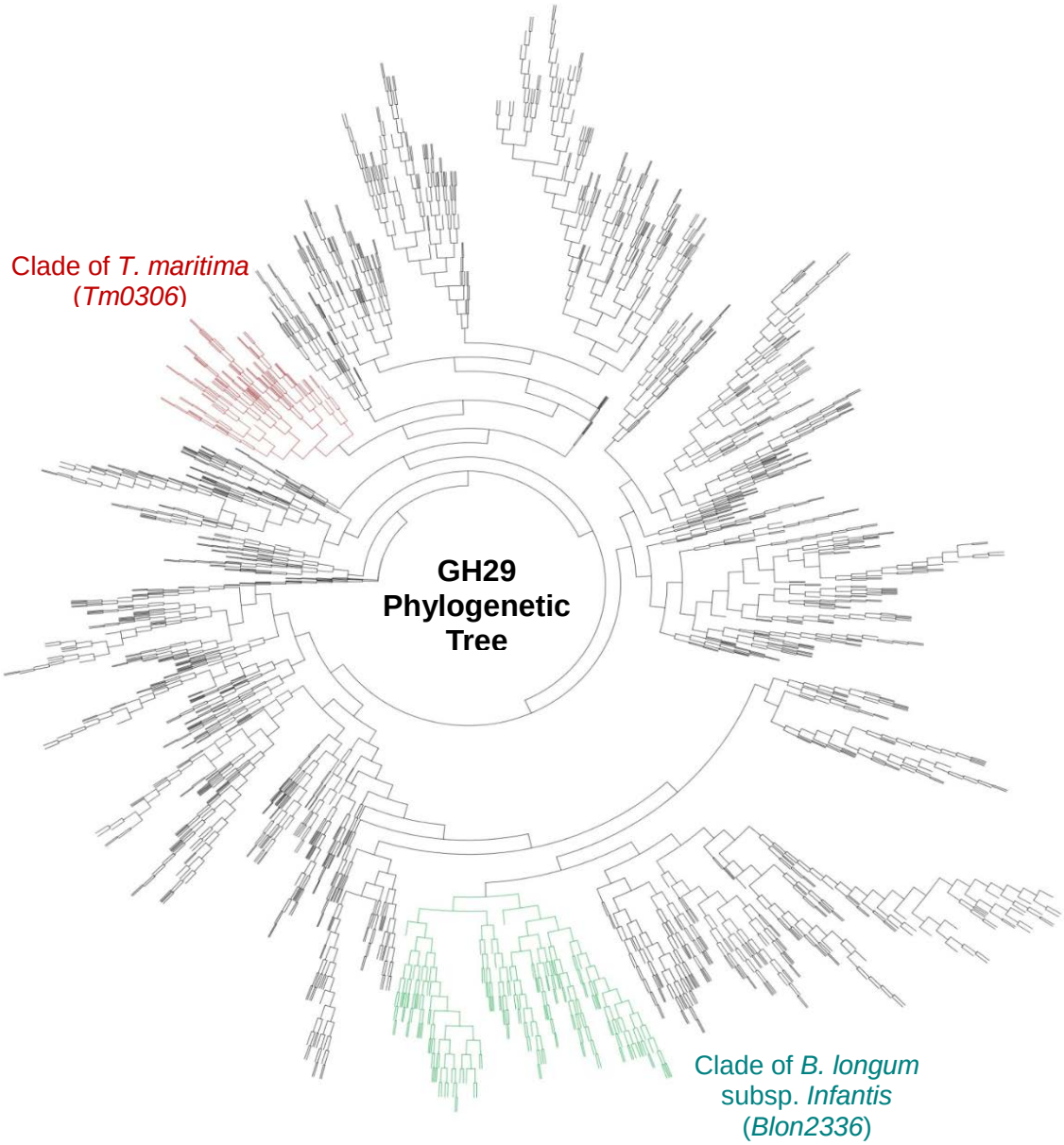


Figure 5-2: The GH29 phylogenetic tree constructed using NJ method with the multiple sequence alignment (using MAFFT). The green highlighted clade belongs to Blon2336 and the red one belongs to Tm0306 indicating the difference in their taxonomic difference and evolution.

One possible reason could be that these proteins belong to two different subfamilies of fucosidase enzymes [48]. Tm0306 is associated to the subfamily A which acts on α-1,2-fucose bonds whereas Blon2336 is linked to subfamily B that acts on α-1,3/4-fucose bonds. All sequences belonging to each clade were next extracted and re-aligned by MSA this

time using the CLUSTAL W algorithm using Geneious software and results shown in Figure 5-3/5-4. The CLUSTAL W algorithm [49] aligns the most similar sequences first and develops a consensus sequence for the sequence pair. The consensus sequences include only the conserved amino acids present in all sequences or partially represented in majority of the sequences (75%). The algorithm then gradually aligns other similar sequences to the consensus sequence. A sequence logo was then developed using these conserved residues and is depicted by amino acids (single alphabet notations) below. Here, the height of the logo at each site is equal to the total information at that site and the height of each symbol in the logo is proportional to its contribution to the information content. The highly conserved residues were then identified and extracted from the consensus sequence after completing the CLUSTAL W alignment for each clade of GH 29 proteins (see Table 7 for Tm0306; data for Blon2336 is presented in appendix).

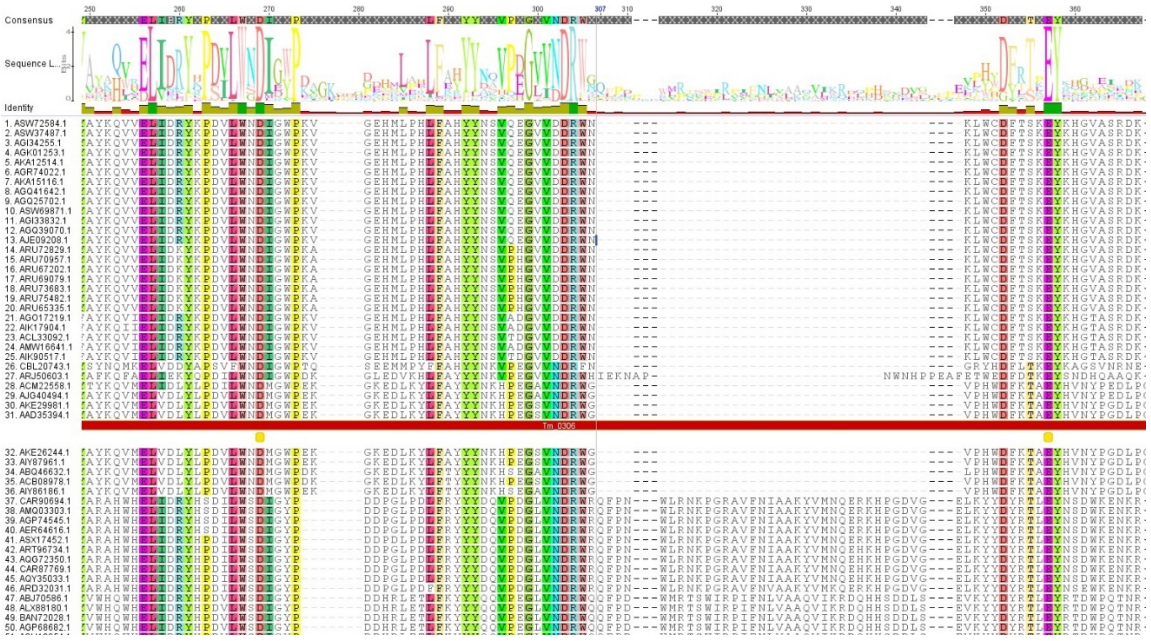


Figure 5-3: CLUSTAL W alignment of Tm0306 clade of sequences developed using Geneious software. The sequences are colored to illustrate significant conserved residues which have also been show been shown as a sequence logo. The red annotated horizontal line represents the Tm0306 sequence. The yellow annotated residues are: left, catalytic nucleophile (shown as a red "D" alphabet) and right, acid/base residue of Tm0306 (shown as a magenta "E" alphabet).

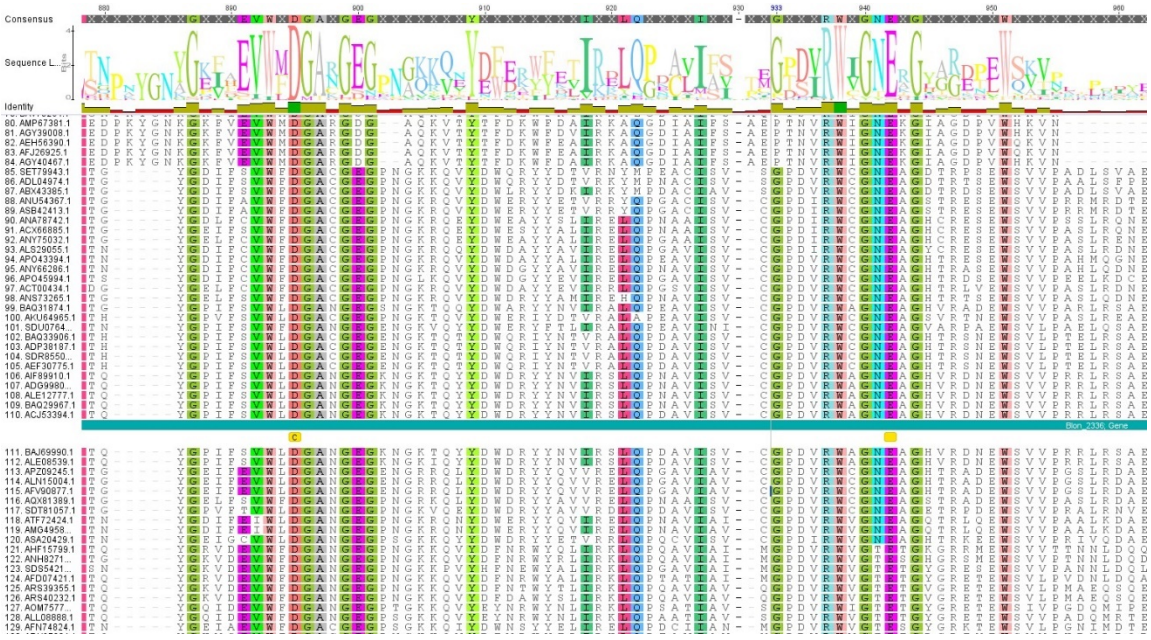


Figure 5-4: CLUSTAL W alignment of Blon2336 clade of sequences developed using Geneious software. The sequences are colored to illustrate significant conserved residues which have also been show been shown as a sequence logo. The green annotated horizontal line represents the Blon2336 sequence. The yellow annotated residues are: left, catalytic nucleophile (shown as a red "D" alphabet) and right, acid/base residue of Blon2336 (shown as a magenta "E" alphabet).

5.2 Results and Discussion

Tm0306		
Residue Type	Consensus Sequence	AAD35394.1 Sequence
W	48	23
G	55	30
H	59	34
Y	99	64
E	101	66
N	105	70
Y	128	93
Y	159	121
G	169	131
G	209	169
Y	211	171
W	218	178
L	257	212
W	267	222
D	269	224
R	304	254
E	357	266
Y	358	267
R	375	284
G	376	285
S	380	289
F	381	290
N	384	293
G	410	319
L	413	322
G	417	326

Table 7: Conserved residues of Tm0306 obtained after MSA of Tm0306 clade from the GH29 phylogenetic tree. The yellow highlighted rows are the conserved catalytic nucleophile and acid/base residue. The green highlighted Y64 residue is discussed later in detail under section 5.3.2.

The conserved residues that showed a 100% identical match were listed and analyzed. For Tm0306, 26 residues were highly conserved (Table 7) and the catalytic nucleophile and acid/base residue were amongst these. This further indicated that these residues are highly conserved to facilitate enzyme function. However, considering several residues were identified, it was difficult to choose only a single residue to mutate other than their catalytic

nucleophile to make them function as more efficient GS. Additional, 3D structural modeling and ligand docking analysis was performed next to identify possible target residues for mutation.

In the case of Blon2336, as the protein sequence was longer (478 amino acid residues), and the conserved residues were also proportionally a lot more and were therefore not listed here (shown in Appendix K). Due to these results and other reasons stated in section 3.5, Blon2336 was not studied further as a protein to engineer GS/TGs. Interestingly, the MSA of Blon2336 clade showed its catalytic nucleophile D172 as a conserved residue but the identified acid/base residue E217 showed up as a residue that is conserved only 30 to 99% of the times. Among the 191 protein sequences present in the Blon2336 clade, 10 sequences showed a Serine residue instead of Glutamic acid, this might be because of the selection of the clade. A smaller clade excluding a few branches would have yielded in a 100% match between the consensus and Blon2336 sequence. It could also be an indicative of the possible mutation that could make the enzyme a better transfucosidase instead of the native hydrolase enzyme.

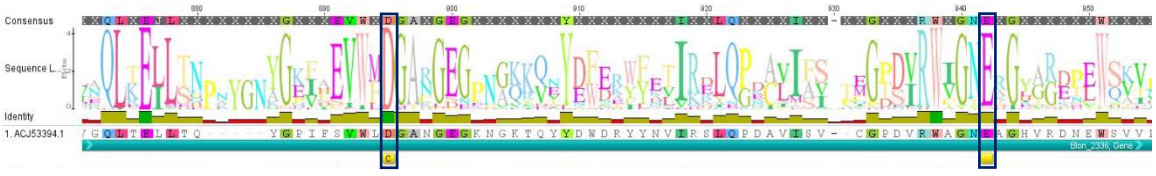


Figure 5-5: MSA of the Blon2336 enlarged to show the D172 conserved residue but E217 not conserved indicating evolutionary changes inside the same clade.

It also implies that either E217 residue got mutated during evolution into a Serine residue and it lost its importance as an acid/base residue when compared to the other members of

the same clade or it could be the other way around and the need for an acid/base residue to breakdown fucose moieties would have led to formation of E217.

5.3 Molecular docking simulations

Next, protein and ligand docking simulations were performed to predict the position and orientation of a targeted ligand (e.g., donor or acceptor sugars) when it is bound within the enzyme active site. The docked position of the ligand will provide further understanding and information about the exact amino acid residues interacting with the ligand. Selected interacting residues were next mutated *in silico* and docking of relevant ligands was further analyzed to confirm if the mutations resulted in 'better docking' based on a suitable docking parameter (e.g., docking free energy).

AutoDock software was used to perform all molecular modeling-based ligand docking simulations. A measure of 'better docking' other than the orientation of the ligand was the binding free energy (in kcal/mol), which is a scoring function used to rank and select the best docked conformations for further analysis. The ideal conformation should have the same or lower binding energy as the natural conformation of the complex. AutoDock calculates the binding energy by finding the minimum global energy of protein, which is measured by the bonds formed between the ligand/s and protein complex through intermolecular interactions drive by ionic bonds, hydrogen bonds and van der Waals type forces. AutoDock combines the empirical free energy force field with a Lamarckian Genetic Algorithm (GA), which was chosen since it is an efficient and reliable stochastic algorithm providing fast prediction of bound conformations with predicted free energies of

binding [50]. Since such dockings are computational, user validated precise protein and ligand structures are typically required as input and multiple interactions can be analyzed simultaneously. An experimental validation is also essential as these docking simulations cannot mimic a physical system completely and so thus cannot provide reliable binding results without user inputs. There are some experimental methods available to obtain more accurate results of the structure of protein-ligand complex like X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and electron microscopy. The most frequently used experimental technique to study protein-ligand interactions is X-ray crystallography. Available X-ray crystal structures for GH29 (with or without bound ligands) were used to validate the AutoDock simulation conformations with the lowest free energy of binding, when feasible.

5.3.1 Material and methods

The relevant GH29 protein crystal structure saved as 1HL8 and 2ZWY in pdbqt format were downloaded from the RCSB protein databank (PDB) and cleaned up to be devoid of any extra unit of protein, ligands, or water using PyMol software. Similarly, the ligand was either preferably extracted from an already docked .pdb file or prepared manually by converting a 3D structure available on PubChem (https://pubchem.ncbi.nlm.nih.gov/) as .sdf file into a .pdbqt file which contains rotatable bonds supported by AutoDock. The prepared protein macromolecule was initially fixed in a grid box of dimension (36, 40, 40) in x, y and z direction centering the catalytic nucleophile (-18.527, 24.798, 53.11) such that the catalytic pocket of the enzyme is enclosed within the grid box with a grid spacing of 0.375 Å. Next, the ligand could be placed at a given grid point with a random initial

position. A receptor-ligand interaction energy was next calculated and stored using the formula given below:

Estimated Free Energy of Binding = (1) + (2) + (3) - (4)

Where;

(1) = Final Intermolecular Energy = ΔG_{vdw} + $\Delta G_{Hbond} + \Delta G_{desol} + \Delta G_{elec}$ Energy

(2) = Final total Internal Energy

(3) = Torsional Energy = ΔG_{tor}

(4) = Unbound System's Energy [= (2)]

Here, ΔG_{vdw} stands for the energy for van der Waals forces, ΔG_{Hbond} represents hydrogen bond contribution, ΔG_{desol} models the solvent entropy changes relevant contribution, ΔG_{elec} is for electrostatic contributions and ΔG_{tor} models the internal and external rotation restriction relevant contributions to free energy of the system. The next step involves using AutoGrid to pre-calculate the energy of a specific atom at regular points over a 3D space around the receptor. These energies are saved as 'grid maps'. The initial position of ligand was set to be random. General arrangement (GA) parameters were set to search number of GA runs to 50, population size to 300 and so on, with all other default parameters. After this, .dpf files were saved containing the docking parameters and instructions for Lamarckian GA docking [50]. With all parameters set and saved, the .dpf files was set as input to run AutoDock. Finally, the data obtained was analyzed using ADT (AutoDock Tools), PyMol (for 3D structure viewing), and LigPlot (to check interacting residues in 2D view).

5.3.2 Results and discussion

Docking simulations were performed first with the 1HL8 PDB structure of targeted Thermotoga maritima fucosidase. 1HL8 structure was chosen over 1ODU as the latter has a few amino acid sequences missing compared to the former. Initial dockings were performed using substrates (as ligands) for TmαFuc such pNP-fucose and 2'-FL that are known to show some activity with the wild-type enzyme. The wild type protein was then mutated to create its catalytic nucleophile D224G using the 'Mutagenesis' option under Wizard in PyMol. The structure was stored as .pdbqt file and was used as detailed in section 5.3.1 to dock other ligands. The ligand or activated donor sugar (β-fucosyl azide) was then docked and hereafter, this docked .pdbqt conformation with the protein was used to dock other acceptor sugar ligands like lactose and pNP-lactose. Based on the docking results and protein-ligand interactions, we identified sites for possible secondary mutations to the D224G enzyme that would facilitate engineering of a more efficient GS/TG. Therefore the. pdbqt D224G mutant protein was next mutated at F59 position to Y and Y64 position to F/W using PyMol software and the double ligand docking simulations with β-fucosyl azide and lactose/pNP-lactose were conducted again. The results obtained from each of these docking simulations are described below:

1) **Docking pNP-fucose into 1HL8_WT**

   Most of the conformations obtained upon simulations were oriented the same way within the catalytic pocket of 1HL8. The minimal energy conformation (-7.64 kcal/mol) obtained after docking of pNP-fucose into 1HL8 was chosen only after validating the orientation of the docked fucose moiety with its conformation reported in the 1ODU structure. This manner of validation of the autodock conformation was

considered ideal since the 1ODU structure was obtained from x-ray crystallography [31].

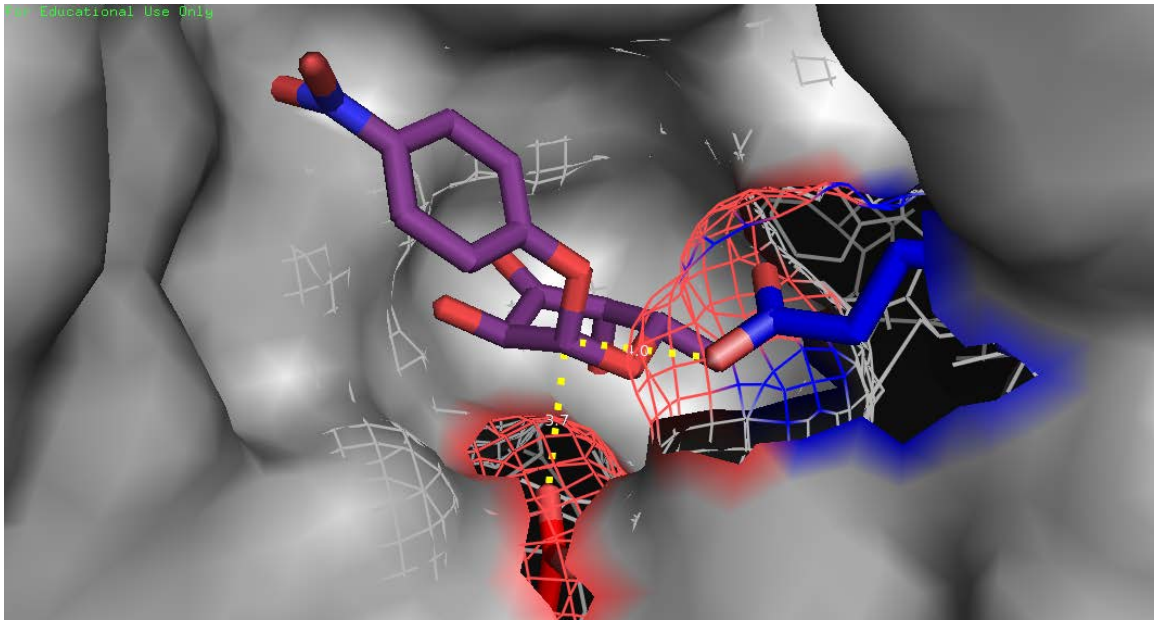


Figure 5-6: Docking pNP-fucose into 1HL8 using AutoDock (viewed using PyMol). The purple ligand is the pNP-fucose and the grey colored background region is the catalytic pocket of the protein shown as surface/mesh. The catalytic nucleophile (D224) is shown in red and acid/base residue (E266) is shown in blue with a mesh indicating the surface. The distance of the C1 anomeric carbon of the pNP-fucose molecule is located 3.7 Å from D224 and 4.0 Å from E266 residue, as shown by the dotted yellow lines.

The distance of the C1 anomeric carbon for the sugar moiety in pNP-fucose is located 3.7 Å from D224 and 4.0 Å from E266 residues. This conformation is therefore well within the bond formation/modification interaction range of transition state relevant ligand-enzyme complexes expected during cleavage of the glycosidic bond. Due to the presence of the pNP group, the fucose of the 1ODU structure did not fully align with the fucose moiety in pNP-fucose. This could be due to the stacking effect of the pNP group over Y64 residue (also confirmed in the Cobucci-Ponzano *et al.* study [40]) and as seen in the LigPlot developed figures showing 2D interactions.

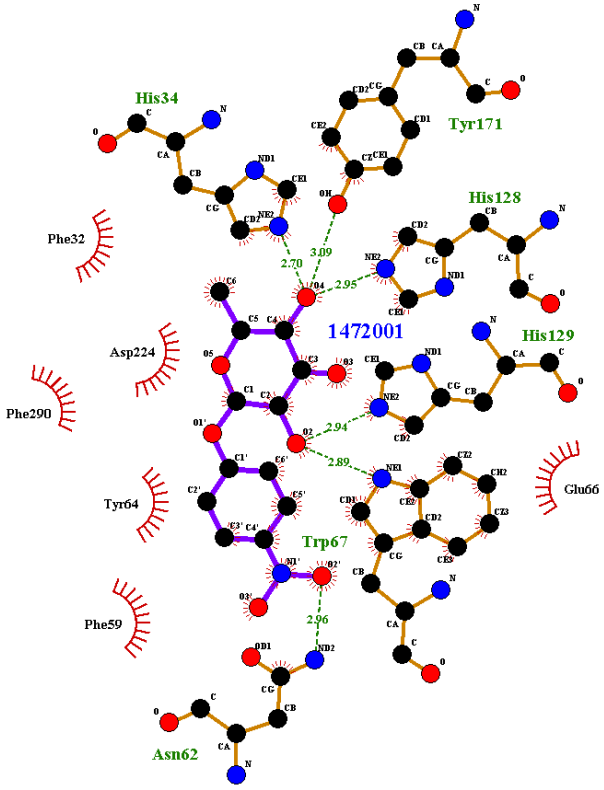


Figure 5-7: 2D interaction between pNP-fucose (compound shown using purple bonds) and nearby residues in the 1HL8 catalytic pocket (drawn using LigPlot); D224 clearly interacts with the C1 of fucose moiety to drive cleavage of that bond and pNP group is seen stacking over the Y64 residue. The H-bonds and their corresponding lengths are denoted using green dashed lines.

2) Docking 2'-FL into 1HL8_WT

The docking of 2'-FL into the catalytic pocket of 1HL8 resulted in various possible conformations. Interestingly, the minimum binding energy conformation (-9.09 kcal/mol) was the most ideal docking conformation based on the orientation of 2'-FL. This docking was first validated with the fucose conformation as seen also in the 1ODU structure. This

conformation of 2'-FL was next used to identify the likely orientations of lactose/pNP-lactose docked into this wild-type enzyme as well.

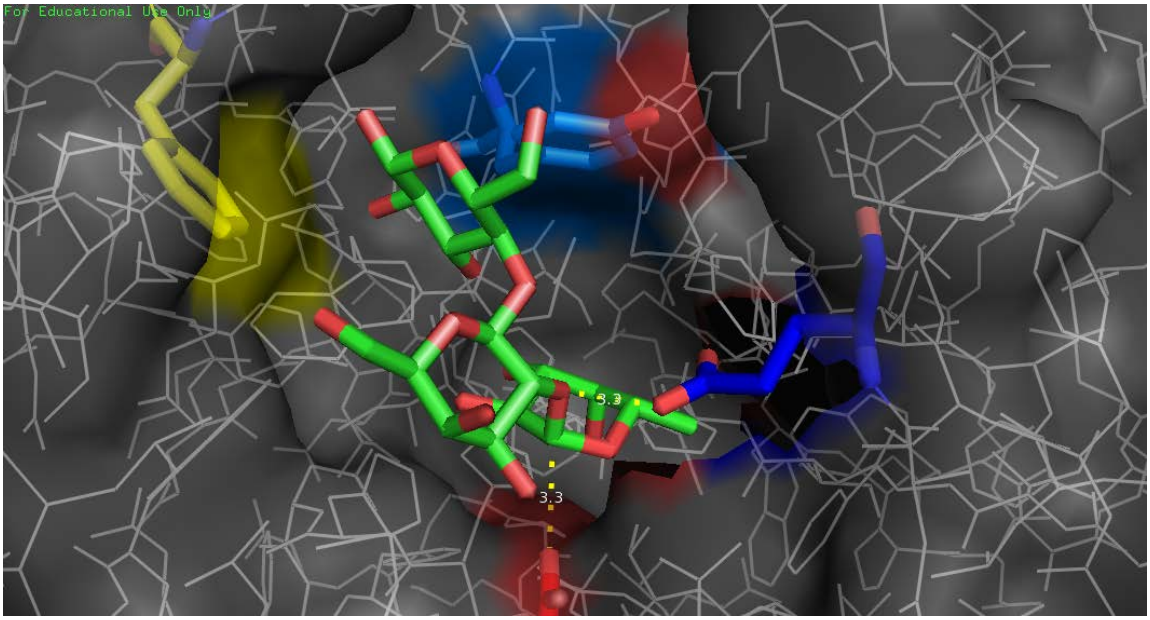


Figure 5-8: Docking 2'-FL into 1HL8 using AutoDock (viewed using PyMol). The green ligand is 2'FL and the grey colored region is the catalytic pocket of the protein shown as surface with transparency setting of 0.5. The catalytic nucleophile (D224) is shown in red and acid/base residue (E266) is shown in blue. The fucose moiety orients perfectly well with the crystal structure 1ODU also showing docked fucose. The distance of the C1 anomeric carbon of the 2'FL is located 3.3 Å from D224 and 3.3 Å from E266 residue. The glucose moiety of 2'-FL interacts with Y64 (indicated in blue) and F59 (indicated in yellow) residues.

This docked structure for 2'-FL (Figure 5-8) highlights possible future mutations suggested for improving the GS activity using β-fucosyl azide and lactose that will likely need to be conducted, as the lactose moiety of 2'-FL is constricted due to the small catalytic pocket. These are confirmed by the hydrophobic interactions between the ligand and protein residues displayed in the 2D LigPlot (Figure 5-9). Further mutations that remove the occupying residues to smaller sized amino acid residues are hypothesized to increase accessibility to the lactose moiety. This could also open possibilities to facilitate docking of larger ligands into this catalytic pocket for functionality.

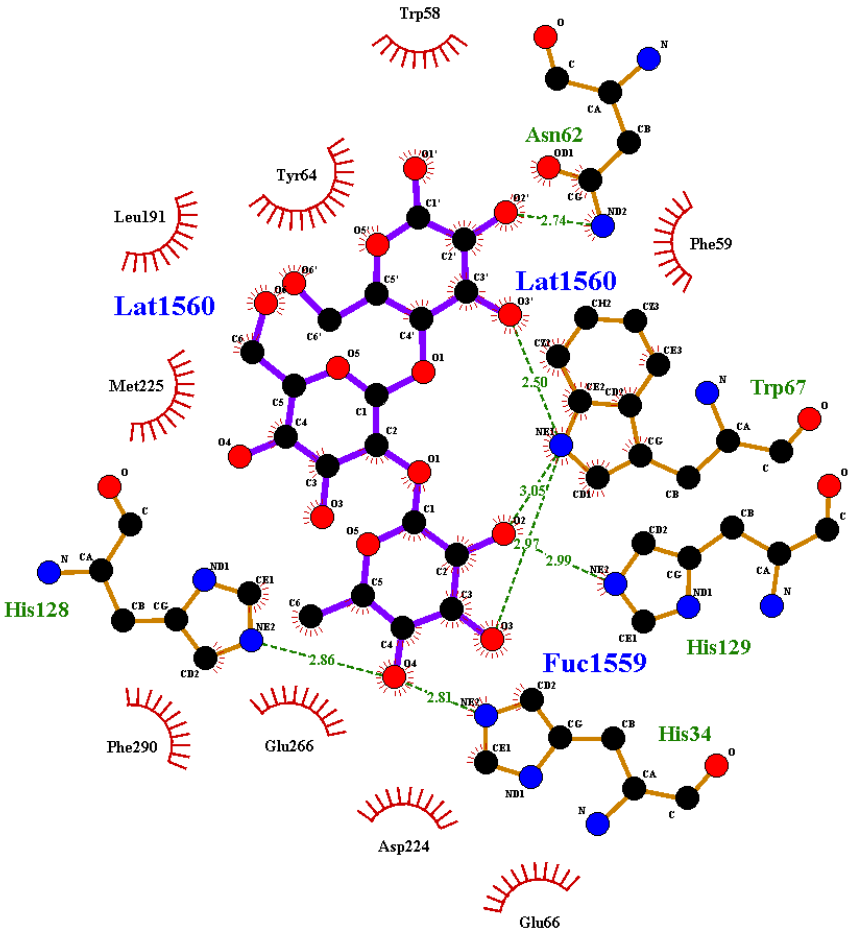


Figure 5-9: 2D interaction between 2'-FL and nearby residues in the catalytic pocket (drawn using LigPlot); The H-bonds and their corresponding lengths are denoted using green dashed lines. The glucose moiety has some hydrophobic interactions with the residue F59 and Y64.

3) Docking β-fucosyl azide into 1HL8_D224G mutant

To create an *in silico* 1HL8_D224G mutant, the mutagenesis option on PyMol was used and the D224 residue was mutated to a Glycine residue. The optimum minimal energy conformation that matched the requirements of mimicking the enzyme-substrate complex was obtained with the 9th energy conformation with a binding energy of -6.36 kcal/mol. The Cobucci-Ponzano *et al.* study [40] have also used the same mutant with β-fucosyl azide to perform GS reactions experimentally but using pNP-xylose as the acceptor sugar. In order to validate the simulations using their experimental results, we intend to dock β-

fucosyl azide along with the pNP-xylose acceptor sugar within the active site of the mutant. However, ligand-docking simulations using two different ligands at the same time is not yet feasible in AutoDock. Thus, we docked the two ligands one at a time into 1HL8_D224G mutant enzyme structure. The results for fucosyl azide docked structure are shown in Figure 5-10 (validated using experimental crystal structure 2WSP to fucose orientation alone), while the next docking of pNP-xylose into this structure is discussed in the next section

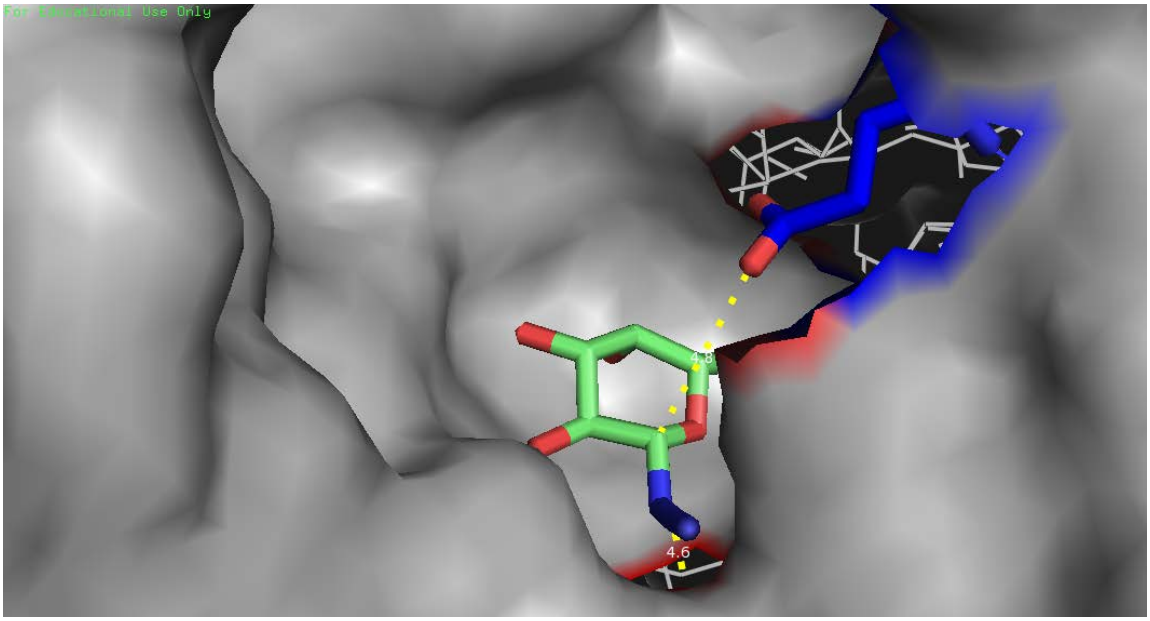


Figure 5-10: Docking β-fucosyl azide into 1HL8_D224G using AutoDock (viewed using PyMol). The green ligand is β-fucosyl azide and the grey color background is the catalytic pocket region of the protein shown as a surface/mesh. The catalytic nucleophile (D224) is shown in red and acid/base residue is shown in blue. The fucose moiety for the donor azide sugar orients according to the published experimental crystal structure of 2WSP docked with α-L-Fuc-(1-2)-β-L-Fuc-N3. The distance of the C1 anomeric carbon of fucose is 4.6 Å from D224G and 4.8 Å from E266 residue, respectively. The C1 carbon of the fucose moiety is closer to the D224G site and the azide arm mimics the enzyme-substrate complex within the empty pocket that would have formed a bond between C1 carbon of fucose and D224 residue in the wild-type enzyme

4) Docking pNP-xylose into 1HL8_D224G mutant docked with β-fucosyl azide

The second acceptor ligand was docked into the previous docking structure, which had optimum conformation of β-fucosyl azide in 1HL8_D224G. The top 15 conformations were carefully evaluated based on their binding affinity values for each enzyme and sugar orientations. The research conducted in the Cobucci-Ponzano et al. paper showed that they obtained equimolar product yields of α-(1-4) and α-(1-3) linkage formation between β-fucosyl azide and pNP-xylose. Based on this work, the O- atom of the E266 residue should be in close proximity of the hydrogen attached to -OH group located on 3rd or 4th carbon of pNP-xylose. This also indicates that the O- present at the 3rd or 4th carbon on pNP-xylose should be within bond formation range to attack the C1 anomeric carbon of β-fucosyl azide. A minimum distance between these groups was found for conformation 12 that had a binding free energy of -4.94 kcal/mol. Even though this binding free energy was closer to zero, this conformation was chosen based on the minimal distance between the bond forming atoms as the other conformations in this set of simulation were in the range of -5.11 kcal/mol to -3.29 kcal/mol binding free energies. The difference in the minimal energy conformation and the chosen conformation was within ~0.2 kcal/mol.

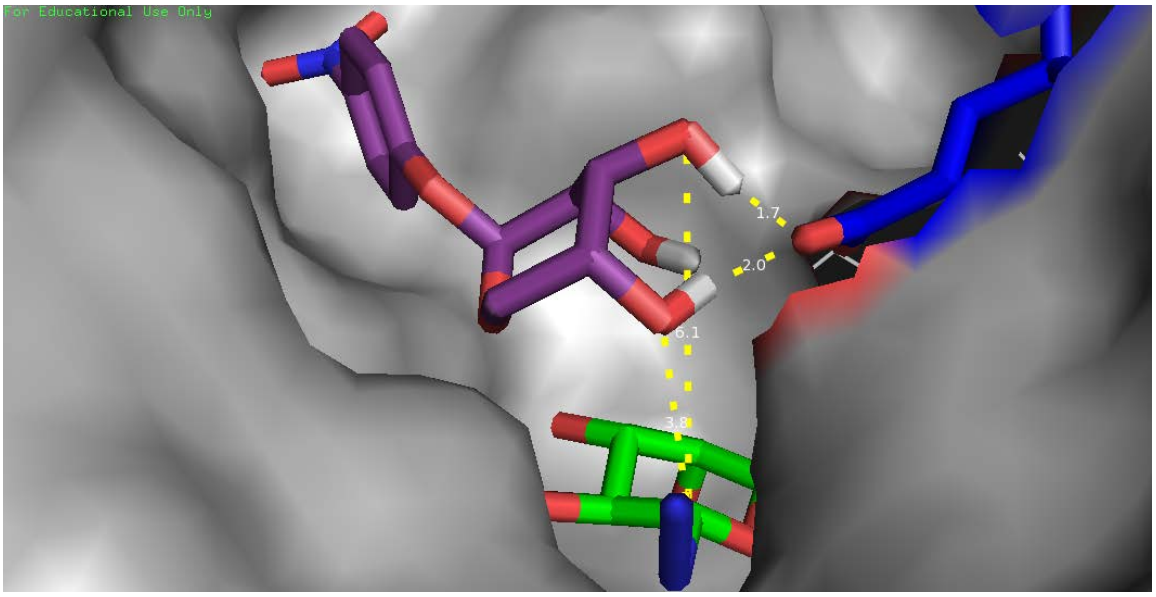


Figure 5-11: Docking pNP-xylose into 1HL8_D224G with previously docked β-fucosyl azide using AutoDock (viewed using PyMol). The green ligand is β-fucosyl azide and the grey colored is the catalytic pocket of the protein shown as a surface. The acid/base is shown in blue. The xylose moiety is oriented so that it can form α-(1-4) and α-(1-3) bonds with β-fucosyl azide. The distance of the O- atom of E266 was 1.7 Å from 3rd and 2.0 Å from 4th hydrogen of pNP-xylose. The distance of the C1 anomeric carbon of fucose is 3.8 Å from the oxygen group on the 4th carbon and 6.1 Å from the 3rd carbon oxygen group of pNP-xylose.

This specific docking also provides detailed knowledge about the acceptor-binding site for this mutant enzyme. The location at which pNP-xylose was docked would be the most probable co-ordinates in which lactose/pNP-lactose moieties might dock if these sugars were to participate in a similar GS reaction. The interactions made by pNP-xylose at the catalytic site in the double ligand docked enzyme complex were analyzed using the 2D interaction plot. The 2D interaction plot also reaffirms our analysis based on the 3D structure viewed in PyMol. Distinct interactions between the O- atom of E266 residue are observed with the oxygen of pNP-xylose, 3rd and 4th carbon. The pNP group of the acceptor seems to have hydrophobic interactions and stacking with respect to the Y64, F59 and W67 aromatic residues.

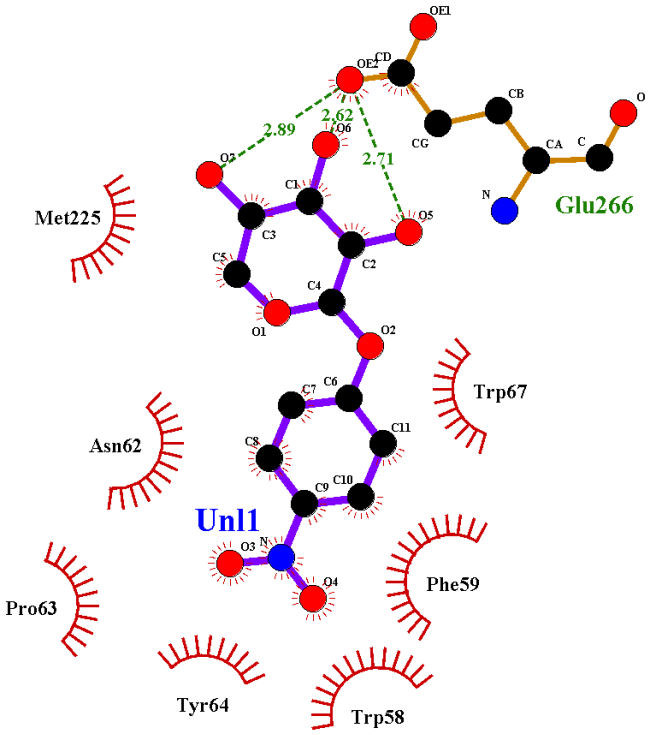


Figure 5-12: 2D interaction plot between pNP-xylose (compounds shown using purple bonds) and nearby residues in the catalytic pocket already docked with β-fucosyl azide in the D224G mutant (drawn using LigPlot). The H-bonds and their corresponding lengths are denoted using green dashed lines between E266 and xylose moiety of pNP-xylose. The pNP group displays stacking interactions between Y64, F59 and W67 residues.

5) Docking lactose into 1HL8_D224G mutant docked with β-fucosyl azide

After identifying the likely acceptor-binding region in the 1HL8_D224G mutant with β-fucosyl azide, secondary ligand lactose was docked instead of pNP-xylose. Next, pNP-lactose was also tested as an acceptor sugar, but the results have not been discussed here since the resulting binding energies were poorer. The GS assays conducted to create 2'-FL using β-fucosyl azide and lactose/pNP-lactose had not yielded any products (section 4.2.3). One reason for this is the poor affinity of these acceptors to be docked within the acceptor site. To verify this, in silico simulations were performed and analysis of the minimal energy dockings were found to show oriented conformations that would likely not favor the

formation of 2'-FL. Also, the obtained binding energies were poor, with the galactose moiety docked farther away than glucose.

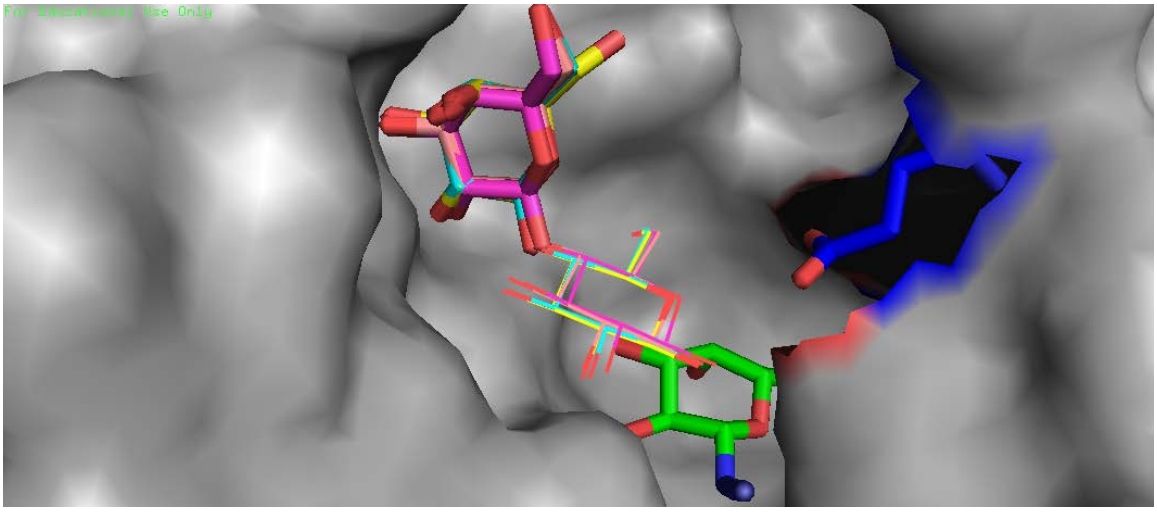


Figure 5-13: Docking lactose into 1HL8_D224G with previously docked β-fucosyl azide using AutoDock (viewed using PyMol). The green ligand is β-fucosyl azide and the grey colored region is the catalytic pocket of the protein shown as a surface. The acid/base is shown in blue. The lactose moiety orients such that the glucose moiety (shown as lines) docked closer to the β-fucosyl azide ligand while the galactose moiety (shown as sticks) docked farther away.

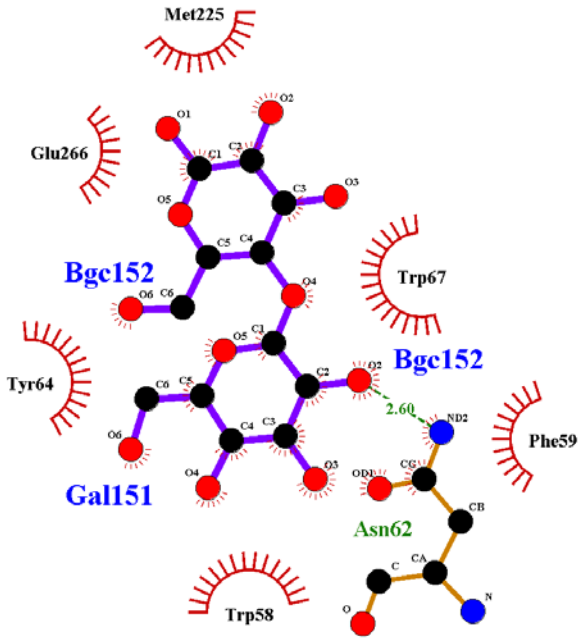


Figure 5-14: 2D interaction plot between lactose (compound shown using purple bonds) and nearby residues in the catalytic pocket already docked with β-fucosyl azide in the D224G mutant (drawn using LigPlot); Lactose shows hydrophobic interactions between Y64, F59 and W67 residues.

An interaction plot showed the residues that formed hydrophobic interactions with the galactose moiety to stabilize it at the incorrectly docked location. Residues F59, Y64 and W67 seem to be playing a key role in this docking and were also discussed earlier. The galactose and glucose in the docked lactose needed to be flipped 180° so that they orient correctly in place to likely form bonds with the fucose moiety and E266 residue. Thus, secondary mutations were made *in silico* such that the

hydrophobicity of the Y64 residue would increase slightly upon mutation of this position. Tyrosine residue has an -OH group that stabilized galactose, thus it was mutated to Phenylalanine that has a hydrophobic aromatic end group and Tryptophan that also has a non-polar indole side chain. Another secondary mutation was created by altering the F59 residue to tyrosine. This was done to increase the possibility of formation of H-bonds by adding a -OH group to the Phenylalanine residue position to allow better docking of glucose near it. After re-docking β-fucosyl azide into these three separate double mutants, it was found that Y64F showed better docking results and hence, only those results have been discussed further. The other mutants displayed higher binding energies and/or incorrect orientations of the docked sugars.

6) Docking lactose into 1HL8_D224G_Y64F double mutant docked with β-fucosyl azide

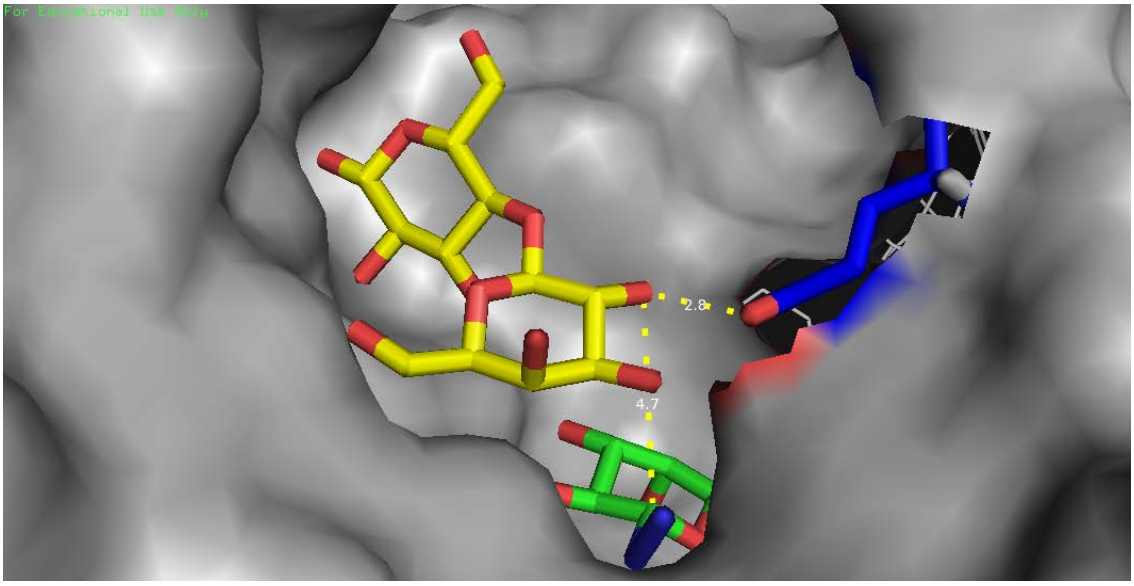


Figure 5-15: Docking lactose into 1HL8_D224G_Y64F double mutant with previously docked β-fucosyl azide using AutoDock (viewed using PyMol). The green ligand is β-fucosyl azide and the grey colored is the catalytic pocket of the protein shown as a surface. The acid/base is shown in blue. The lactose moiety (shown as yellow sticks) is assumed to orient correctly with the glucose moiety docking more towards the top left and the galactose moiety docking closer to the E266 residue. The distance of the O^- atom of E266 was 2.8 Å away from hydrogen present on the 2nd carbon of galactose moiety. The distance of the C1 anomeric carbon of fucose is 4.7Å from the oxygen group on the 2^{nd} carbon.

The ideal minimum energy conformation was found orienting in accordance with the lactose seen in 2'-FL docking results and it had a binding energy of -6.17 kcal/mol (Figure 5-15). Based on the 2D interaction plot (Figure 5-16) we hypothesized that this was the most appropriate conformation for lactose docking that will likely lead to the formation of 2'-FL in a GS reaction between lactose and fucosyl azide.

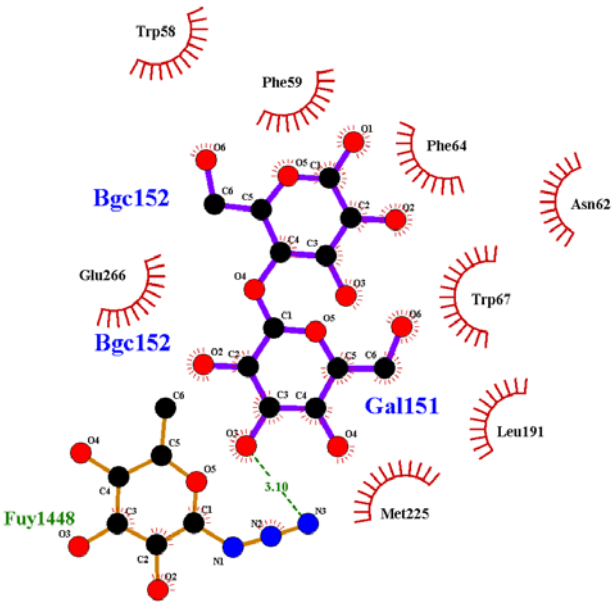


Figure 5-16: 2D interaction plot between lactose (compound shown using purple bonds) and nearby residues in the catalytic pocket already docked with β-fucosyl azide (compound shown using light brown bonds) in the D224G_Y64F double mutant (drawn using LigPlot); Lactose was hypothesized to show correct orientation.

This mutation is likely ideal for the 1HL8 based protein structure but on closer verification of the protein sequence of TmAFuc_WT available in the lab, it was found that 1HL8 crystallographic structure is missing a few important amino acid residues. The missing loops are amino acid backbone chains from 47 to 55 and 268 to 273. Upon additional research of the PDB database, the 2ZWY pdb structure was identified from the X-ray crystallographic study performed by Wu *et al.* [51]. This structure was more complete and did not have missing loops near the active site of this enzyme.

These missing loops (Figure 5-17) are likely to impact the binding energies and the final orientation of the docked ligands. Thus, mutations to the TmαFuc_D224G enzyme were not attempted as additional simulations based on using 2ZWY structure are currently on going.

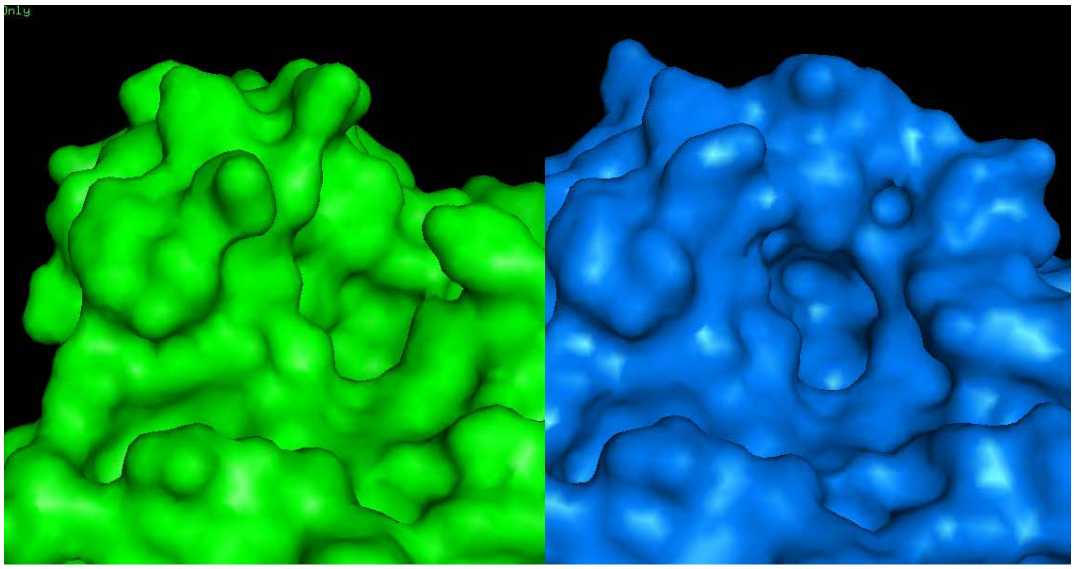


Figure 5-17: PyMol images of protein 1HL8 and 2ZWY showing the varying size of the catalytic pocket due to absence of certain amino acid loops (in 1HL8)

Chapter 6 : Conclusions and future work

This research was aimed at; (1) Design enzymes by reverse engineering the active site residues of native GH to make them function as effective GS that can synthesize designer oligosaccharides, and (2) Study how specific designer oligosaccharides can modulate the growth of selectively engineered bacteria that utilize a specific oligosaccharide as a carbon source.

Growth studies: During the course of this thesis work (for aim 2), it was concluded that selective growth was indeed observed with the engineered BL21_pET26b+_Blon2336_V374A_WT strain line in the presence of appropriate substrate, 3'-FL that could be hydrolyzed by the expressed protein, BiAfc_WT, and the released lactose could be hence metabolized by the cells. However, a similar trend was not observed in the case of the engineered BL21_pEC_Tm0306_WT cell line. It was hypothesized that the expressed enzymes could not hydrolyze 2'-FL due to inappropriate growth temperature of 37 °C, which is low for a thermostable enzyme like TmαFuc. This was also proven by conducting enzyme assays at lower temperatures. For further studies, to aid growth of these engineered cells expressing thermostable enzymes on 2'-FL either higher concentrations of the substrate should be used and/or improved protein expression might help drive increased hydrolysis rate on the substrate.

Designing effective GS enzymes for oligosaccharides synthesis:

As a general conclusion to this section of the thesis (aim 1), a workflow was developed for chemo-enzymatic synthesis of designer oligosaccharides using a combination of enzymatic

assays, bioinformatics, and molecular modeling approaches. Two genes coding for fucosidase enzymes that were intended to be used to synthesize fucosylated oligosaccharide, 2'-FL and 3'FL, were analyzed as part of this workflow as summarized below.

***Bifidobacterium longum* subsp. *infantis*:** The genus Bifidobacterium is associated with infant colon gut microbiome, where these organisms benefit their host by supplementing fed nutrients by actively participating in the digestion of HMOs, that are otherwise not available to the host, to produce beneficial by-products such as short chain fatty acids. Here, one such Bifidobacterium gene, *Blon_2336* which codes for an α-1,3/4-fucosidase (BiAfc) was studied. The hydrolysis assays conducted using BiAfc_WT confirmed that this enzyme was active on the substrates 2Cl-4NP-fucose and 3'-FL but inactive on pNP-fucose and 2'-FL, as shown also previously [52]. The catalytic nucleophile mutants D172A/S/G were then created but weren't pursued further as it was challenging to select a readily available activated donor (pNP-fucose) that could be used as substrate for chemical rescue reactions. The bioinformatics study conducted by aligning multiple sequences of the same clade of the phylogenetic tree as BiAfc or ACJ53394.1 resulted in identification of numerous conserved residues. However, since it was arduous to decide on additional mutations to create a GS from BiAfc aimed at synthesis of 3'-FL, we focused on the TmαFuc enzyme instead.

***Thermotoga maritima*:** The α-1,2-fucosidase (TmαFuc) produced by the *Tm_0306* gene is the closest bacterial relative that shares only 38% identity with the mammalian α-L-

fucosidases [37]. The bottleneck to large-scale production of this protein was its expression in the form of inclusion bodies. Attempts were made to improve expression using various plasmids or strains of *E. coli* that would help solubilize this protein to obtain more soluble protein yield. After many failed attempts, the protein was expressed using the cells BL21_pEC_Tm0306_WT in a larger culture volume instead to isolate sufficient active protein for *in vitro* assays. During hydrolysis assays, TmαFuc_WT showed hydrolytic activity on pNP-fucose, 2Cl-4NP-fucose and 2'-FL but was also found to be inactive on 3'-FL due to the specificity (α-1,2-fucosidase) of its enzyme towards 2'-FL. The hydrolysis assays also confirmed that pNP-Fucose could be used as a substrate for chemical rescue experiments.

Here, we further studied in detail the effect of various externally added nucleophiles and varying concentrations to identify the best external nucleophile. We found 1 M sodium azide behaved as an ideal nucleophile that helped the TmαFuc_D224G mutant recover its hydrolysis activity on pNP-fucose. For GS reactions, the azide ion plays the role of a good leaving group when present as β-fucosyl azide [40]. Another interesting result was obtained while analyzing the sodium azide chemical rescue reactions using TLC analysis. The reaction of TmαFuc_D224G with sodium azide in the presence of pNP-fucose resulted in the formation of the activated donor β-fucosyl azide. Further analysis of this product needs to be conducted using mass spectroscopy techniques and if the compound structure is established then the purchase of this expensive donor could be avoided, and the GS experimental protocol could be modified accordingly. Further studies need to be conducted to optimize the conditions for donor sugar preparation and the reaction henceforth. Studies

to understand if the released sodium and pNP groups in the former reaction could also affect the GS reaction need to also be evaluated.

Among all other external nucleophiles tested, surprisingly imidazole also displayed the ability to recover the hydrolytic activity of this mutant. Imidazole has been used as an external nucleophile with many other enzyme families [53-56] but has never been used for glycosyl hydrolases. Further chemical rescue studies should be based on researching varying concentrations of imidazole for varying pH conditions to optimize chemical rescue. Imidazole could be also used in conjugation with fucose as another possible leaving group for GS reactions.

Aiming to synthesize 2'-FL, GS reaction was carried out with β-fucosyl azide as the activated donor and lactose or pNP-lactose as acceptor sugars. The results were not fruitful, and it was hypothesized that this might be owing to the large size of acceptor sugars compared to the limited space available in the catalytic pocket. Further mutations to create space within the catalytic pocket where the acceptors could dock more efficiently were sought after. A bioinformatics and molecular ligand docking modeling approach was used to find additional mutants in *in silico* experiments.

A phylogenetic tree generated using multiple sequence alignment of all available sequences of GH29 family of enzymes. The respective clade of Tm0306 protein was identified and another multiple sequence alignment of the proteins within this clade alone resulted in identification of few critical residues that were highly conserved within clade. Conserved

residues were identified, and a few residues were chosen to be mutated which were in proximity of the catalytic pocket. Docking simulations were next carried out using available GH29 pdb structures (e.g., 1HL8) to initially dock the active substrates pNP-fucose and 2'-FL. This simulation was first validated for the TmαFuc_D224G mutant by verifying the possible orientations of the docked ligands that would facilitate the reaction between the activated donor β-fucosyl azide and pNP-xylose. The results helped us identify the acceptor site of this enzyme that could potentially dock lactose or pNP-lactose. The binding free energy and orientations of the docked ligands were the key factors that were then closely monitored when swapping the acceptor sugars. The simulations demonstrated the need for a double mutant that would further reduce the ligand binding free energy and allow better orientation of the acceptor sugar (lactose) with the donor sugar to facilitate the reaction. However, upon additional investigation, it was identified that the 2ZWY pdb structure is a better crystallographic structure to setup the docking simulations than 1HL8. The protocol developed for docking these ligands is currently being applied to 2ZWY using AutoDock prior to making additional GH29 mutants.

Abbreviations

HMO: Human Milk Oligosaccharides

GH: Glycosyl hydrolase

GT: Glycosyl transferase

GS: Transglycosidase

2'-FL: 2'- Fucosyllactose

3'-FL: 3'- Fucosyllactose (Fucose(α1-3)- Galactose(β1-4)- Glucose)

3-FL: 3- Fucosyllactose (Galactose β1-4 (Fucose(α1-3))- Glucose)

3'-SL: 3'-Sialyllactose

FUT2: α-1,2-Fucosyltransferase

pNP: Para Nitrophenyl group

2Cl-4NP: 2-Chloro-4-Nitrophenyl group

LDFT: Lacto-difucotetraose

FOS: Fructo- Oligosaccharide

LB: Luria-Bertani

OD_{600}: Optical Density at 600 nm

WT: Wild-Type

MM: MasterMix

PIPE: Polymerase Incomplete Primer Extension (PIPE) Cloning Method

SLIC: Sequence- and Ligation-Independent Cloning

PCR: Polymerase Chain Reaction

SDM: Site directed mutagenesis

PDB: Protein Data Bank

SDS: Sodium Dodecyl Sulfate

IB: Inclusion Bodies

B-PER II: Bacterial-Protein Extraction Reagent (2X)

PIC: Protein Inhibitor Cocktail

IPTG: Isopropyl β-D-1-thiogalactopyranoside

FPLC: Fast Protein Liquid Chromatography

IMAC: Immobilized Metal Affinity Chromatography

MES: 2-(N-Morpholino)ethane Sulfonic acid

TLC: Thin Layer Chromatography

DNS: 3,5-Dinitrosalicylic acid

T_m: Melting Temperature (°C)

QM-MM: Quantum Mechanics/ Molecular Mechanics

MSA: Multiple Sequence Alignment

NJ: Neighbor-Joining

MAFFT: Multiple Alignment using Fast Fourier Transform

NMR: Nuclear Magnetic Resonance Spectroscopy

GA: General Arrangement

REFERENCES

1. Yu, Z.-T., et al., *The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota.* Glycobiology, 2012. **23**(2): p. 169-177.
2. Wu, S., et al., *Annotation and structural analysis of sialylated human milk oligosaccharides.* Journal of proteome research, 2011. **10**(2): p. 856-868.
3. Wu, S., et al., *Development of an annotated library of neutral human milk oligosaccharides.* Journal of proteome research, 2010. **9**(8): p. 4138-4151.
4. Kobata, A., *Structures and application of oligosaccharides in human milk.* Proceedings of the Japan Academy, Series B, 2010. **86**(7): p. 731-747.
5. Mariño, K., et al., *Method for milk oligosaccharide profiling by 2-aminobenzamide labeling and hydrophilic interaction chromatography*. Glycobiology, 2011. **21**(10): p. 1317-1330.
6. Engfer, M.B., et al., *Human milk oligosaccharides are resistant to enzymatic hydrolysis in the upper gastrointestinal tract–*. The American journal of clinical nutrition, 2000. **71**(6): p. 1589-1596.
7. Bode, L., *Human milk oligosaccharides: every baby needs a sugar mama.* Glycobiology, 2012. **22**(9): p. 1147-1162.
8. Ruhaak, L.R. and C.B. Lebrilla, *Advances in Analysis of Human Milk Oligosaccharides–*. Advances in Nutrition, 2012. **3**(3): p. 406S-414S.
9. Newburg, D.S., et al., *Innate protection conferred by fucosylated oligosaccharides of human milk against diarrhea in breastfed infants.* Glycobiology, 2003. **14**(3): p. 253-263.
10. Castanys-Muñoz, E., M.J. Martin, and P.A. Prieto, *2'-Fucosyllactose: an abundant, genetically determined soluble glycan present in human milk.* Nutrition reviews, 2013. **71**(12): p. 773-789.
11. Thurl, S., et al., *Detection of four human milk groups with respect to Lewis blood group dependent oligosaccharides.* Glycoconjugate journal, 1997. **14**(7): p. 795-799.
12. Zivkovic, A.M. and D. Barile, *Bovine milk as a source of functional oligosaccharides for improving human health.* Advances in Nutrition, 2011. **2**(3): p. 284-289.
13. Hindsgaul, O., et al., *Synthesis of type 2 human blood-group antigenic determinants. The H, X, and Y haptens and variations of the H type 2 determinant as probes for the combining site of the lectin I of Ulex europaeus.* Carbohydrate research, 1982. **109**: p. 109-142.
14. Chin, Y.W., et al., *Metabolic engineering of Escherichia coli to produce 2'-fucosyllactose via salvage pathway of guanosine 5'-diphosphate (GDP)-l-fucose.* Biotechnology and bioengineering, 2016. **113**(11): p. 2443-2452.
15. Chin, Y.-W., et al., *Improved production of 2'-fucosyllactose in engineered Escherichia coli by expressing putative α-1, 2-fucosyltransferase, WcfB from Bacteroides fragilis.* Journal of biotechnology, 2017. **257**: p. 192-198.
16. CHIN, Y.-W., et al., *P0013 Enhanced Production of 2'-Fucosyllactose in Engineered Escherichia coli BL21 Star (DE3) by Modulation of Lactose Metabolism and Fucosyltransferase.*

17. Lee, W.-H., et al., *Whole cell biosynthesis of a functional oligosaccharide, 2'-fucosyllactose, using engineered Escherichia coli.* Microbial cell factories, 2012. **11**(1): p. 48.
18. Taylor, D.E., et al., *Lack of correlation between Lewis antigen expression by Helicobacter pylori and gastric epithelial cells in infected patients.* Gastroenterology, 1998. **115**(5): p. 1113-1122.
19. Breton, C., et al., *Structures and mechanisms of glycosyltransferases.* Glycobiology, 2006. **16**(2): p. 29R-37R.
20. Sela, D.A., et al., *Bifidobacterium longum subsp. infantis ATCC 15697 α-fucosidases are active on fucosylated human milk oligosaccharides.* Applied and environmental microbiology, 2012. **78**(3): p. 795-803.
21. Lodish, H., et al., *Growth of Microorganisms in Culture*. 2000.
22. Whitehead, T.A., et al., *Negatively supercharging cellulases render them lignin-resistant.* ACS Sustainable Chemistry & Engineering, 2017. **5**(7): p. 6247-6252.
23. Deuschle, U., et al., *Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures.* The EMBO journal, 1986. **5**(11): p. 2987-2994.
24. Blommel, P.G., et al., *Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition.* Biotechnology progress, 2007. **23**(3): p. 585-598.
25. Murphy, V., *Expression, purification and characterization of carbohydrate-binding proteins*. 2014.
26. Gräslund, S., et al., *Protein production and purification.* Nature methods, 2008. **5**(2): p. 135.
27. Wang, Y. and Y.P. Zhang, *Overexpression and simple purification of the Thermotoga maritima 6-phosphogluconate dehydrogenase in Escherichia coli and its application for NADPH regeneration.* Microbial cell factories, 2009. **8**(1): p. 30.
28. Yadava, R.S., R. Kumar, and P.K. Yadava, *Expression of lexA targeted ribozyme in Escherichia coli BL-21 (DE3) cells.* Molecular and cellular biochemistry, 2005. **271**(1-2): p. 197-203.
29. Bornhorst, J.A. and J.J. Falke, *[16] Purification of proteins using polyhistidine affinity tags*, in *Methods in enzymology*. 2000, Elsevier. p. 245-254.
30. Hochuli, E., et al., *Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent.* Nature biotechnology, 1988. **6**(11): p. 1321.
31. Lim, S., S.P. Chundawat, and B.G. Fox, *Expression, purification and characterization of a functional carbohydrate-binding module from Streptomyces sp. SirexAA-E.* Protein expression and purification, 2014. **98**: p. 1-9.
32. Zhang, H., H. Wu, and H. Zhang, *A novel high-copy plasmid, pEC, compatible with commonly used Escherichia coli cloning and expression vectors.* Biotechnology letters, 2007. **29**(3): p. 431-437.
33. Li, M.Z. and S.J. Elledge, *SLIC: a method for sequence-and ligation-independent cloning*, in *Gene synthesis*. 2012, Springer. p. 51-59.
34. Jeong, J.-Y., et al., *One-step sequence-and ligation-independent cloning (SLIC): rapid and versatile cloning method for functional genomics studies.* Applied and environmental microbiology, 2012: p. AEM. 00844-12.

35. Froger, A. and J.E. Hall, *Transformation of plasmid DNA into E. coli using the heat shock method.* Journal of visualized experiments: JoVE, 2007(6).
36. Rinas, U., et al., *Bacterial inclusion bodies: discovering their better half.* Trends in biochemical sciences, 2017. **42**(9): p. 726-737.
37. Sulzenbacher, G., et al., *Crystal Structure of Thermotoga maritima α-L-Fucosidase insights into the catalytic mechanism and the molecular basis for fucosidosis.* Journal of Biological Chemistry, 2004. **279**(13): p. 13119-13128.
38. Yen, T.Y., et al., *Characterization of cysteine residues and disulfide bonds in proteins by liquid chromatography/electrospray ionization tandem mass spectrometry.* Journal of mass spectrometry, 2000. **35**(8): p. 990-1002.
39. Kong, B. and G.L. Guo, *Soluble expression of disulfide bond containing proteins FGF15 and FGF19 in the cytoplasm of Escherichia coli.* PLoS One, 2014. **9**(1): p. e85890.
40. Cobucci-Ponzano, B., et al., *β-Glycosyl azides as substrates for α-glycosynthases: preparation of efficient α-l-fucosynthases.* Chemistry & biology, 2009. **16**(10): p. 1097-1108.
41. Ingle, R.A., *Histidine biosynthesis.* The Arabidopsis Book/American Society of Plant Biologists, 2011. **9**.
42. Huang, S. and S.-C. Tu, *Identification and characterization of a catalytic base in bacterial luciferase by chemical rescue of a dark mutant.* Biochemistry, 1997. **36**(48): p. 14609-14615.
43. Katoh, K., et al., *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.* Nucleic acids research, 2002. **30**(14): p. 3059-3066.
44. Shaikh, F.A., et al., *Identifying the catalytic acid/base in GH29 α-l-fucosidase subfamilies.* Biochemistry, 2013. **52**(34): p. 5857-5864.
45. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Molecular biology and evolution, 1987. **4**(4): p. 406-425.
46. Vane-Wright, R.I., C.J. Humphries, and P.H. Williams, *What to protect?—Systematics and the agony of choice.* Biological conservation, 1991. **55**(3): p. 235-254.
47. Faith, D.P., *Conservation evaluation and phylogenetic diversity.* Biological conservation, 1992. **61**(1): p. 1-10.
48. Saumonneau, A., et al., *Design of an α-L-transfucosidase for the synthesis of fucosylated HMOs.* Glycobiology, 2015. **26**(3): p. 261-269.
49. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic acids research, 1994. **22**(22): p. 4673-4680.
50. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of computational chemistry, 1998. **19**(14): p. 1639-1662.
51. Wu, H.J., et al., *Structural Basis of α-Fucosidase Inhibition by Iminocyclitols with Ki Values in the Micro-to Picomolar Range.* Angewandte Chemie, 2010. **122**(2): p. 347-350.

52. Ashida, H., et al., *Two distinct α-L-fucosidases from Bifidobacterium bifidum are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates.* Glycobiology, 2009. **19**(9): p. 1010-1017.
53. Chen, W.-J., et al., *Probing the catalytic mechanism of bovine pancreatic deoxyribonuclease I by chemical rescue.* Biochemical and biophysical research communications, 2007. **352**(3): p. 689-696.
54. Hung, J.E., et al., *Chemical rescue and inhibition studies to determine the role of Arg301 in phosphite dehydrogenase.* PloS one, 2014. **9**(1): p. e87134.
55. Lehoux, I.E. and B. Mitra, *(S)-Mandelate dehydrogenase from Pseudomonas putida: mutations of the catalytic base histidine-274 and chemical rescue of activity.* Biochemistry, 1999. **38**(31): p. 9948-9955.
56. Williams, D.M., D. Wang, and P.A. Cole, *Chemical rescue of a mutant protein-tyrosine kinase.* Journal of Biological Chemistry, 2000. **275**(49): p. 38127-38130.

APPENDIX

A. Cell growth studies conducted on Fucose, Lactose, and Fucose + Lactose as carbon sources using either BL21 control and BL21 with *lacI* containing plasmid (BL21_GFP)

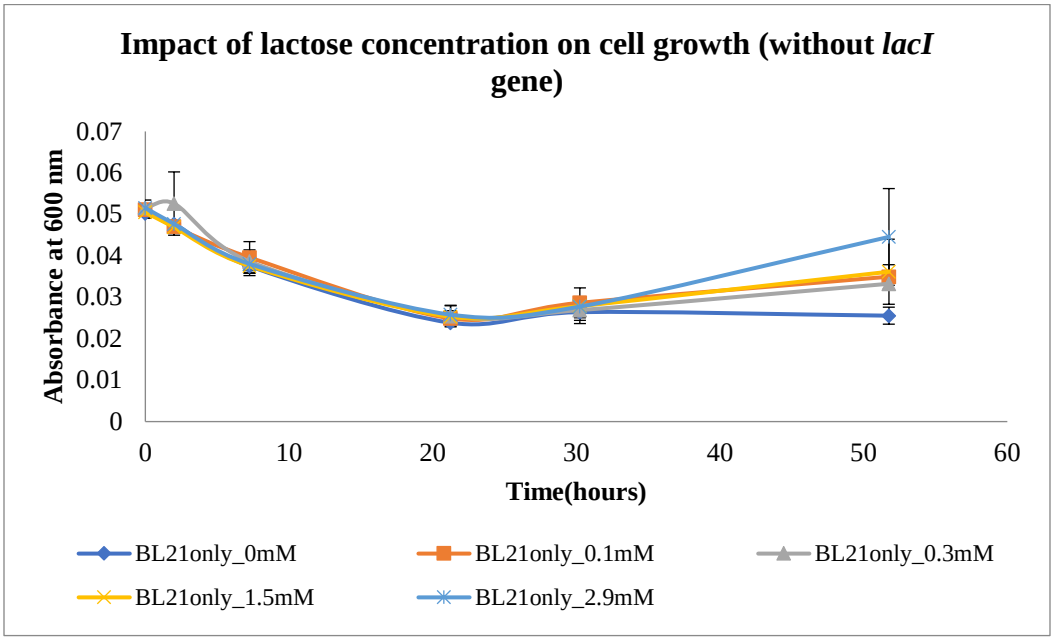


Figure A- 1: Growth curve of BL21 control cells in the presence of varying concentrations of lactose is shown. Here, the cells show no significant growth due to their inability to efficiently use lactose as a carbon food source in absence of lacI.

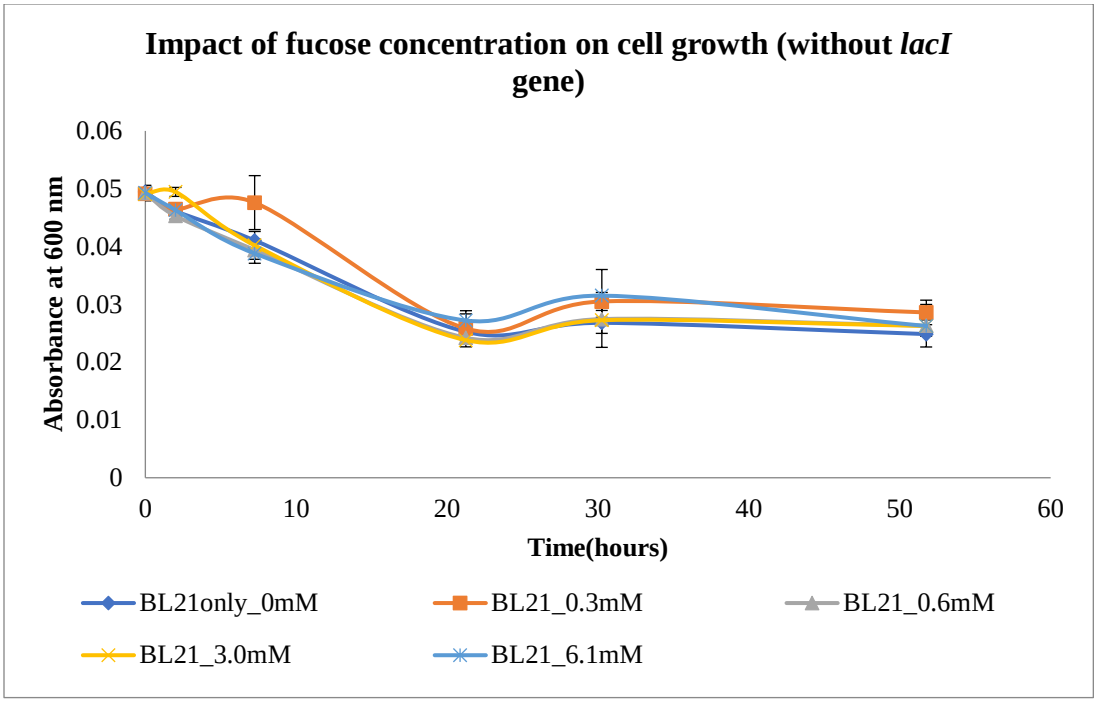


Figure A- 2: Growth curve of BL21 control cells in the presence of varying concentrations of fucose is shown. Here, the cells show no significant growth due to their inability to use fucose as a carbon food source.

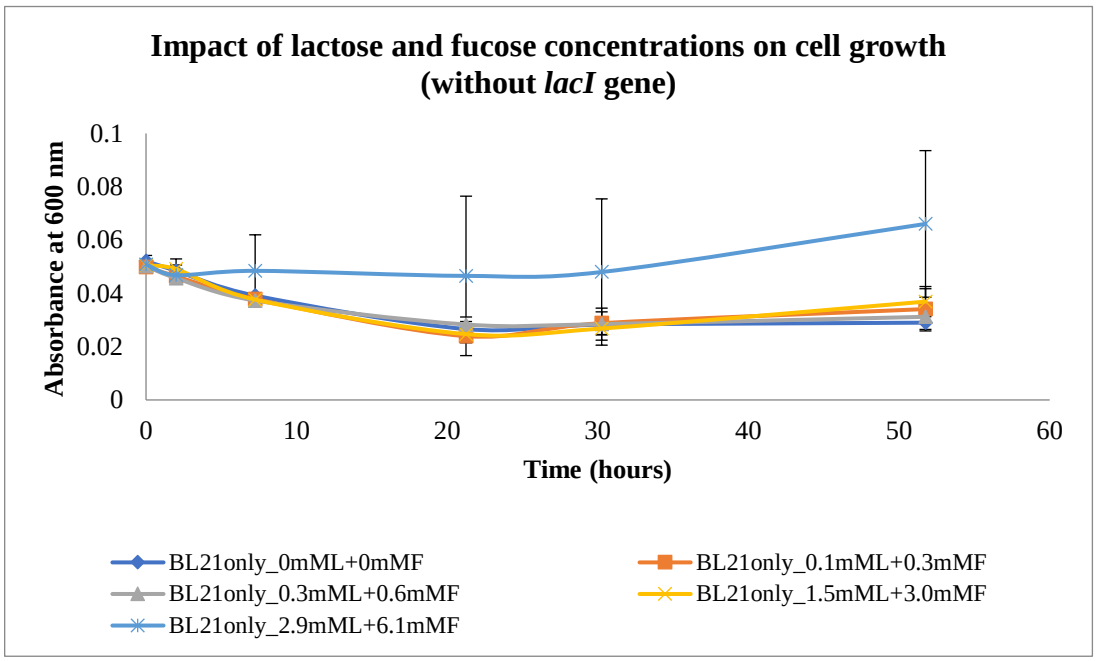


Figure A- 3: Growth curve of BL21 control cells in the presence of varying concentrations of both lactose or fucose is shown. Here, the cells show no significant growth due to their inability to efficiently use lactose + fucose as a carbon food source in the absence of lacI.

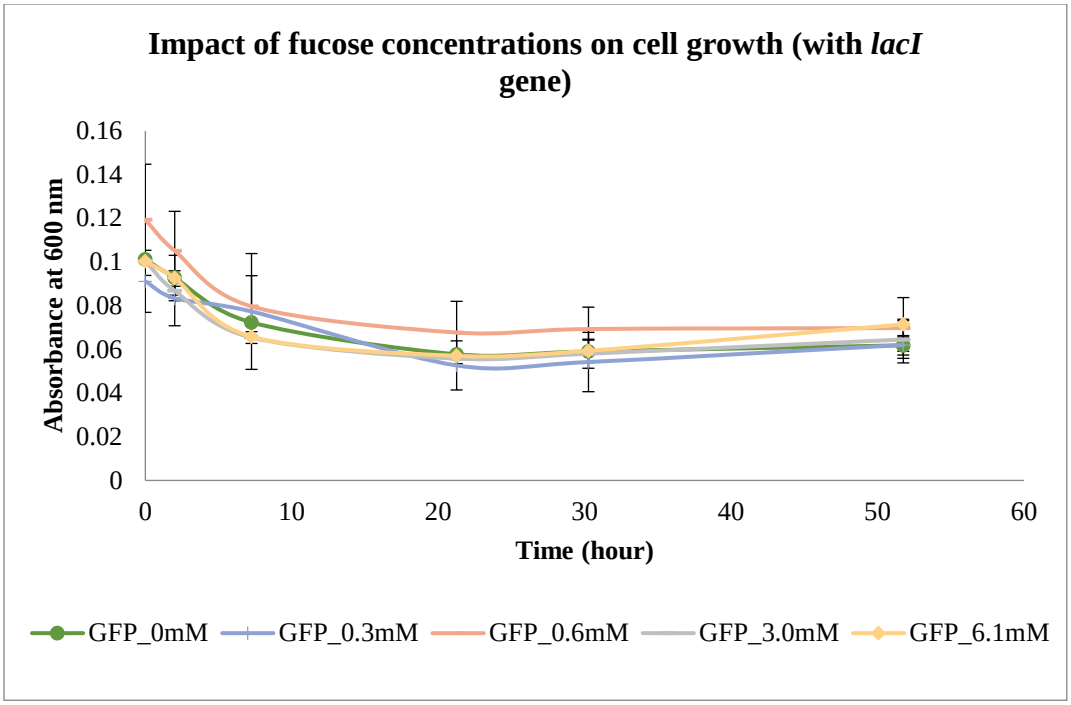


Figure A- 4: Growth curve of BL21_GFP (with plasmid) cells in the presence of varying concentrations of fucose is shown. Here, the cells show no significant growth due to their inability to use fucose as a carbon food source in the presence of lacI.

B. Cell growth on 2'-FL as carbon source alone is shown below for three different cell lines at 0^{th} hour and after ~5 days of growth.

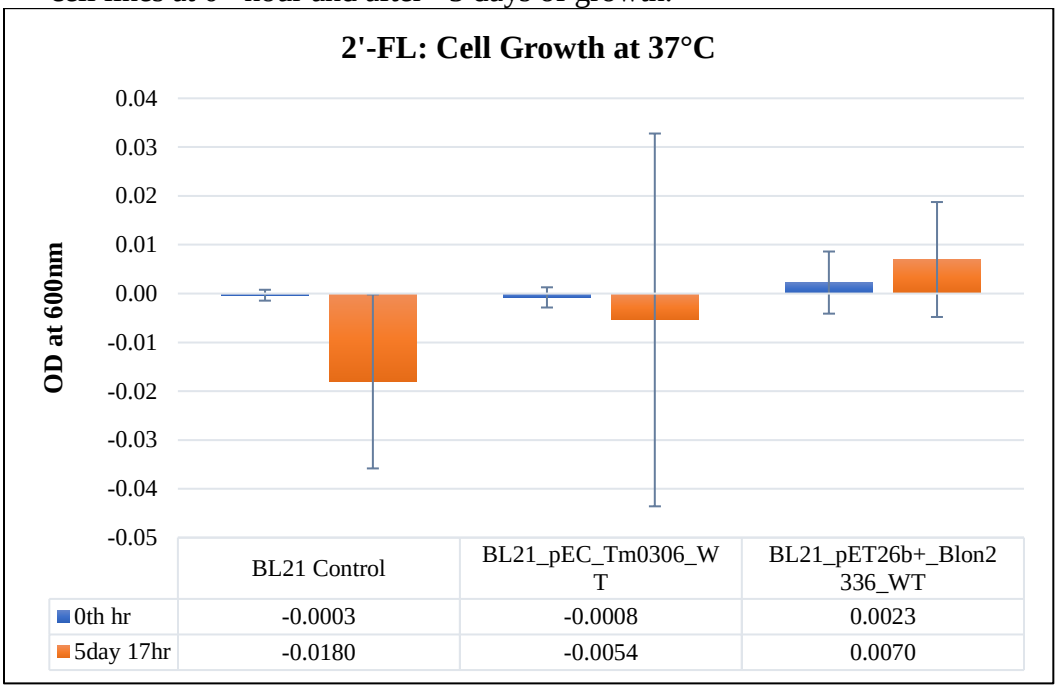


Figure B- 1: BL21 control, BL21_pEC_Tm0306_WT and BL21_pET26b+_Blon2336_WT cell growth in minimal media with 2mM 2'-FL. Note that all three cultures could not metabolize 2'-FL and hence displayed cell death.

C. Cell growth curve on 3'-FL as carbon source alone is shown below for three different cell lines from 0th hour to ~5 days of growth duration.

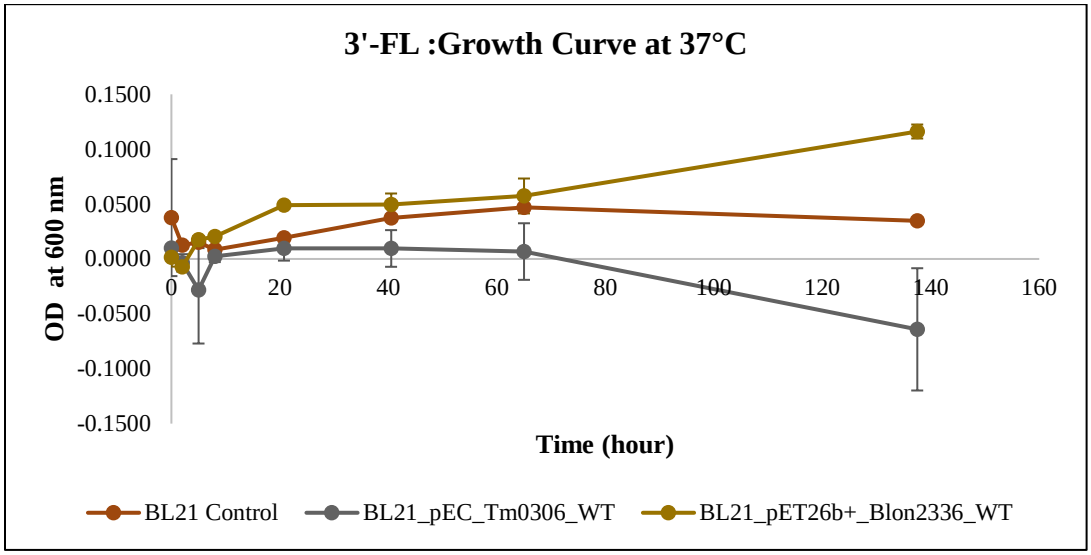


Figure C- 1: BL21 control, BL21_pEC_Tm0306_WT and BL21_pET26b+_Blon2336_WT cell growth in minimal media with 2mM 3'-FL. Note that BL21 control cultures showed no further cell growth whereas BL21_pEC_Tm0306_WT could not metabolize 3'-FL and hence displayed cell death. Only BL21_pET26b+_Blon2336_WT displayed significant growth.

D. TLC of all standards sugars used

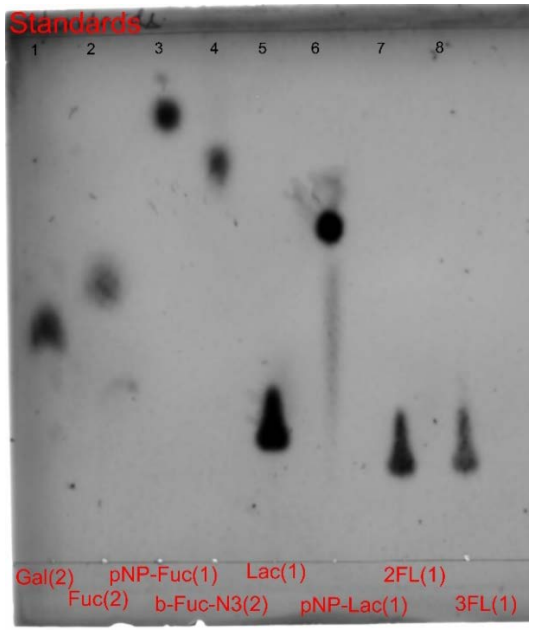


Figure D- 1: TLC image showing spotting of all sugars used in the current thesis. Lane 1: galactose, lane 2: fucose, lane 3: pNP-fucose, lane 4: β-fucosyl azide, lane 5: lactose, lane 6: pNP-lactose, lane 7: 2'-FL and lane 8: 3'-FL. The numbers shown within parentheses for each of these standards suggests the volumes in µL of solution spotted on the TLC plate.

E. PIPE/ SLIC primers to create pET26b+_Blon2336_V374A_WT

PIPE/SLIC Primers	
Insert Forward	5’-TAATACGACTCACTATAGG-3’
Insert Reverse	5’-GCTAGTTATTGCTCAGCGG-3’
Vector Forward	5’-TGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGC-3’
Vector Reverse	5’-TATCCGCTCACAATTCCCCTATAGTGAGTCGTATTA-3’

Table 8: PIPE/SLIC primers used to create pET26b+_Blon2336_V374A_WT using pET26b+ (vector) & Blon2336_V374A_WT (insert)

F. DNA Sequences of wild-type and each nucleophile mutant

Thermotoga maritima or Tm_0306

WT:

MGHHHHHHHHASENLYFQAIAMISMKPRYKPDWESLREHTVPKWFDKAKFGIFI
HWGIYSVPGWATPTGELGKVPMDAWFFQNPYAEWYENSLRIKESPTWEYHVKT
YGENFEYEKFADLFTAEKWDPQEWADLFKKAGAKYVIPTTKHHDGFCLWGTKY
TDFNSVKRGPKRDLVGDLAKAVREAGLRFGVYYSGGLDWRFTTEPIRYPEDLSY
IRPNTYEYADYAYKQVMELVDLYLPDVLWNDMGWPEKGKEDLKYLFAYYYNK
HPEGSVNDRWGVPHWDFKTAEYHVNYPGDLPGYKWEFTRGIGLSFGYNRNEGP
EHMLSVEQLVYTLVDVVSKGGNLLLNVGPKGDGTIPDLQKERLLGLGEWLRKY
GDAIYGTSVWERCCAKTEDGTEIRFTRKCNRIFVIFLGIPTGEKIVIEDLNLSAGTV
RHFLTGERLSFKNVGKNLEITVPKKLLETDSITLVLEAVEE*

D224A:

MGHHHHHHHHASENLYFQAIAMISMKPRYKPDWESLREHTVPKWFDKAKFGIFI
HWGIYSVPGWATPTGELGKVPMDAWFFQNPYAEWYENSLRIKESPTWEYHVKT
YGENFEYEKFADLFTAEKWDPQEWADLFKKAGAKYVIPTTKHHDGFCLWGTKY

TDFNSVKRGPKRDLVGDLAKAVREAGLRFGVYYSGGLDWRFTTEPIRYPEDLSY
IRPNTYEYADYAYKQVMELVDLYLPDVLWNAMGWPEKGKEDLKYLFAYYYNK
HPEGSVNDRWGVPHWDFKTAEYHVNYPGDLPGYKWEFTRGIGLSFGYNRNEGP
EHMLSVEQLVYTLVDVVSKGGNLLLNVGPKGDGTIPDLQKERLLGLGEWLRKY
GDAIYGTSVWERCCAKTEDGTEIRFTRKCNRIFVIFLGIPTGEKIVIEDLNLSAGTV
RHFLTGERLSFKNVGKNLEITVPKKLLETDSITLVLEAVEE*

D224S:

MGHHHHHHHHASENLYFQAIAMISMKPRYKPDWESLREHTVPKWFDKAKFGIFI
HWGIYSVPGWATPTGELGKVPMDAWFFQNPYAEWYENSLRIKESPTWEYHVKT
YGENFEYEKFADLFTAEKWDPQEWADLFKKAGAKYVIPTTKHHDGFCLWGTKY
TDFNSVKRGPKRDLVGDLAKAVREAGLRFGVYYSGGLDWRFTTEPIRYPEDLSY
IRPNTYEYADYAYKQVMELVDLYLPDVLWNSMGWPEKGKEDLKYLFAYYYNK
HPEGSVNDRWGVPHWDFKTAEYHVNYPGDLPGYKWEFTRGIGLSFGYNRNEGP
EHMLSVEQLVYTLVDVVSKGGNLLLNVGPKGDGTIPDLQKERLLGLGEWLRKY
GDAIYGTSVWERCCAKTEDGTEIRFTRKCNRIFVIFLGIPTGEKIVIEDLNLSAGTV
RHFLTGERLSFKNVGKNLEITVPKKLLETDSITLVLEAVEE*

D224G:

MGHHHHHHHHASENLYFQAIAMISMKPRYKPDWESLREHTVPKWFDKAKFGIFI
HWGIYSVPGWATPTGELGKVPMDAWFFQNPYAEWYENSLRIKESPTWEYHVKT
YGENFEYEKFADLFTAEKWDPQEWADLFKKAGAKYVIPTTKHHDGFCLWGTKY
TDFNSVKRGPKRDLVGDLAKAVREAGLRFGVYYSGGLDWRFTTEPIRYPEDLSY
IRPNTYEYADYAYKQVMELVDLYLPDVLWNGMGWPEKGKEDLKYLFAYYYNK
HPEGSVNDRWGVPHWDFKTAEYHVNYPGDLPGYKWEFTRGIGLSFGYNRNEGP

EHMLSVEQLVYTLVDVVSKGGNLLLNVGPKGDGTIPDLQKERLLGLGEWLRKY
GDAIYGTSVWERCCAKTEDGTEIRFTRKCNRIFVIFLGIPTGEKIVIEDLNLSAGTV
RHFLTGERLSFKNVGKNLEITVPKKLLETDSITLVLEAVEE*

Bifidobacterium longum* subspecies *infantis* or *Blon_2336

WT:

MGSSHHHHHHSSGRENLYFQGHMNNPADAGINLNYLANVRPSSRQLAWQRME
MYAFLHFGMNTMTDREWGLGHEDPALFNPRNVDVDQWMDALVAGGMAGVIL
TCKHHDGFCLWPSRLTRHTVASSPWREGKGDLVREVSESARRHGLKFGVYLSP
WDRTEESYGKGKAYDDFYVGQLTELLTQYGPIFSVWLDGANGEGKNGKTQYY
DWDRYYNVIRSLQPDAVISVCGPDVRWAGNEAGHVRDNEWSVVPRRLRSAELT
MEKSQQEDDASFATTVSSQDDDLGSREAVAGYGDNVCWYPAEVDTSIRPGWFY
HQSEDDKVMSADQLFDLWLSAVGGNSSLLLNIPPSPEGLLAEPDVQSLKGLGRR
VSEFREALASVRCEARTSSASAAAAHLADGNRDTFWRPDADDAAPAITLTLPQP
TTINAIVIEEAIEHGQRIEHLRVTGALPDGTERVLGQAGTVGYRRILRFDDVEVSS
VTLHVDGSRLAPMISRAAAVRI*

D172A:

MGSSHHHHHHSSGRENLYFQGHMNNPADAGINLNYLANVRPSSRQLAWQRME
MYAFLHFGMNTMTDREWGLGHEDPALFNPRNVDVDQWMDALVAGGMAGVIL
TCKHHDGFCLWPSRLTRHTVASSPWREGKGDLVREVSESARRHGLKFGVYLSP
WDRTEESYGKGKAYDDFYVGQLTELLTQYGPIFSVWLAGANGEGKNGKTQYY
DWDRYYNVIRSLQPDAVISVCGPDVRWAGNEAGHVRDNEWSVVPRRLRSAELT
MEKSQQEDDASFATTVSSQDDDLGSREAVAGYGDNVCWYPAEVDTSIRPGWFY
HQSEDDKVMSADQLFDLWLSAVGGNSSLLLNIPPSPEGLLAEPDVQSLKGLGRR

VSEFREALASVRCEARTSSASAAAAHLADGNRDTFWRPDADDAAPAITLTLPQP
TTINAIVIEEAIEHGQRIEHLRVTGALPDGTERVLGQAGTVGYRRILRFDDVEVSS
VTLHVDGSRLAPMISRAAAVRI*

D172S:

MGSSHHHHHHSSGRENLYFQGHMNNPADAGINLNYLANVRPSSRQLAWQRME
MYAFLHFGMNTMTDREWGLGHEDPALFNPRNVDVDQWMDALVAGGMAGVIL
TCKHHDGFCLWPSRLTRHTVASSPWREGKGDLVREVSESARRHGLKFGVYLSP
WDRTEESYGKGKAYDDFYVGQLTELLTQYGPIFSVWLSGANGEGKNGKTQYYD
WDRYYNVIRSLQPDAVISVCGPDVRWAGNEAGHVRDNEWSVVPRRLRSAELTM
EKSQQEDDASFATTVSSQDDDLGSREAVAGYGDNVCWYPAEVDTSIRPGWFYH
QSEDDKVMSADQLFDLWLSAVGGNSSLLLNIPPSPEGLLAEPDVQSLKGLGRRV
SEFREALASVRCEARTSSASAAAAHLADGNRDTFWRPDADDAAPAITLTLPQPTT
INAIVIEEAIEHGQRIEHLRVTGALPDGTERVLGQAGTVGYRRILRFDDVEVSSVT
LHVDGSRLAPMISRAAAVRI*

D172G:

MGSSHHHHHHSSGRENLYFQGHMNNPADAGINLNYLANVRPSSRQLAWQRME
MYAFLHFGMNTMTDREWGLGHEDPALFNPRNVDVDQWMDALVAGGMAGVIL
TCKHHDGFCLWPSRLTRHTVASSPWREGKGDLVREVSESARRHGLKFGVYLSP
WDRTEESYGKGKAYDDFYVGQLTELLTQYGPIFSVWLGGANGEGKNGKTQYY
DWDRYYNVIRSLQPDAVISVCGPDVRWAGNEAGHVRDNEWSVVPRRLRSAELT
MEKSQQEDDASFATTVSSQDDDLGSREAVAGYGDNVCWYPAEVDTSIRPGWFY
HQSEDDKVMSADQLFDLWLSAVGGNSSLLLNIPPSPEGLLAEPDVQSLKGLGRR
VSEFREALASVRCEARTSSASAAAAHLADGNRDTFWRPDADDAAPAITLTLPQP

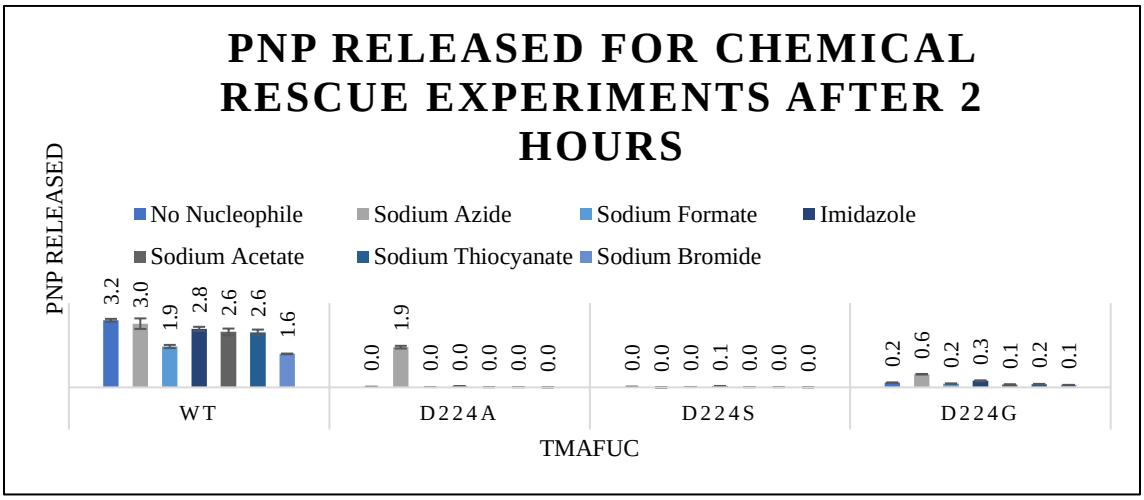
TTINAIVIEEAIEHGQRIEHLRVTGALPDGTERVLGQAGTVGYRRILRFDDVEVSS

VTLHVDGSRLAPMISRAAAVRI*

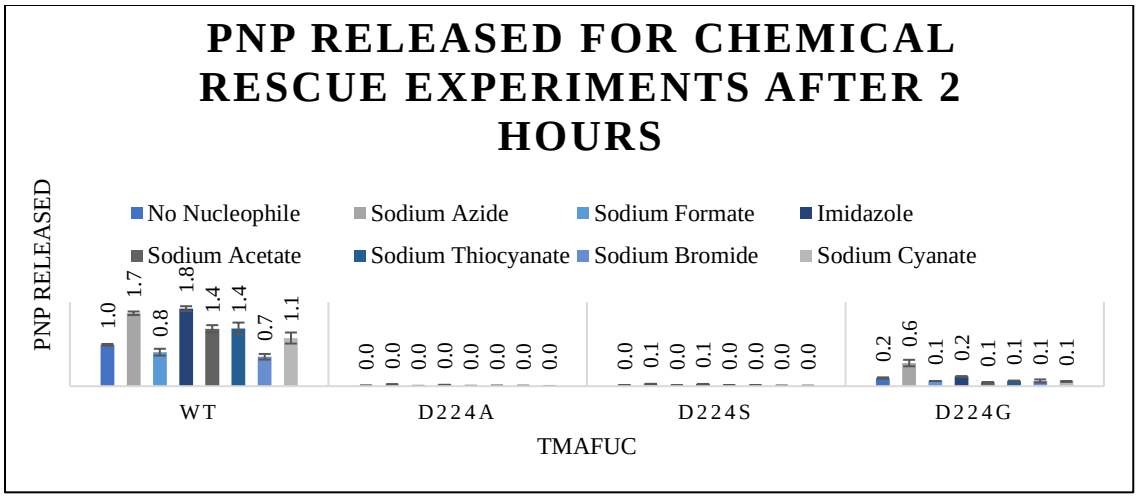
G. Chemical rescue reactions with varying external nucleophile

We performed this experiment thrice, the results shown in the section 4.2.1 of the thesis are the results obtained from the last set of sampled time points from a replicate assay. We performed both TLC & DNS assays for the samples isolated after 18.5 hours for only the last time point. For the previous two earlier sampled time points we only analyzed the samples by taking an OD reading to estimate the released pNP concentration at 2 hours and then terminated the experiment. The amounts of fucose released are provided in the next section of the appendix:

Trial 1: Not all nucleophiles were tested and D224A showed an anomalous result for azide unlike the D224G mutant. Hence, the assays were repeated.



Trial 2: There was some cross contamination while removing the top cover off the plate. WT was showing lower values without nucleophiles than D224A.



H. Fucose concentration measurements using DNS assay for third trail of varying nucleophile chemical rescue (Raw OD data) are provided below. The fucose concentration calculated using fucose standard curve.

Fucose released after reaction	No Nucleophile	Sodium Azide	Sodium Formate	Imidazole	Sodium Acetate	Sodium Thiocyanate	Sodium Bromide	Sodium Cyanate
WT	0.50	0.50	0.40	0.36	0.50	0.43	0.36	0.00
D224A	0.08	0.12	0.07	0.09	0.09	0.08	0.08	0.00
D224S	0.13	0.12	0.07	0.08	0.09	0.08	0.09	0.00
D224G	0.23	0.22	0.15	0.17	0.16	0.15	0.17	0.01
Control	0.07	0.08	0.08	0.07	0.09	0.07	0.07	0.00
Standard Deviation (OD)	**No Nucleophile**	**Sodium Azide**	**Sodium Formate**	**Imidazole**	**Sodium Acetate**	**Sodium Thiocyanate**	**Sodium Bromide**	**Sodium Cyanate**
WT	0.02	0.00	0.05	0.02	0.01	0.03	0.04	0.00
D224A	0.00	0.02	0.01	0.00	0.02	0.00	0.01	0.00
D224S	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D224G	0.03	0.02	0.01	0.00	0.00	0.04	0.00	0.00
Control	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00

I. Enzymatic hydrolysis of 2'-FL using TmαFuc_WT at 37°C

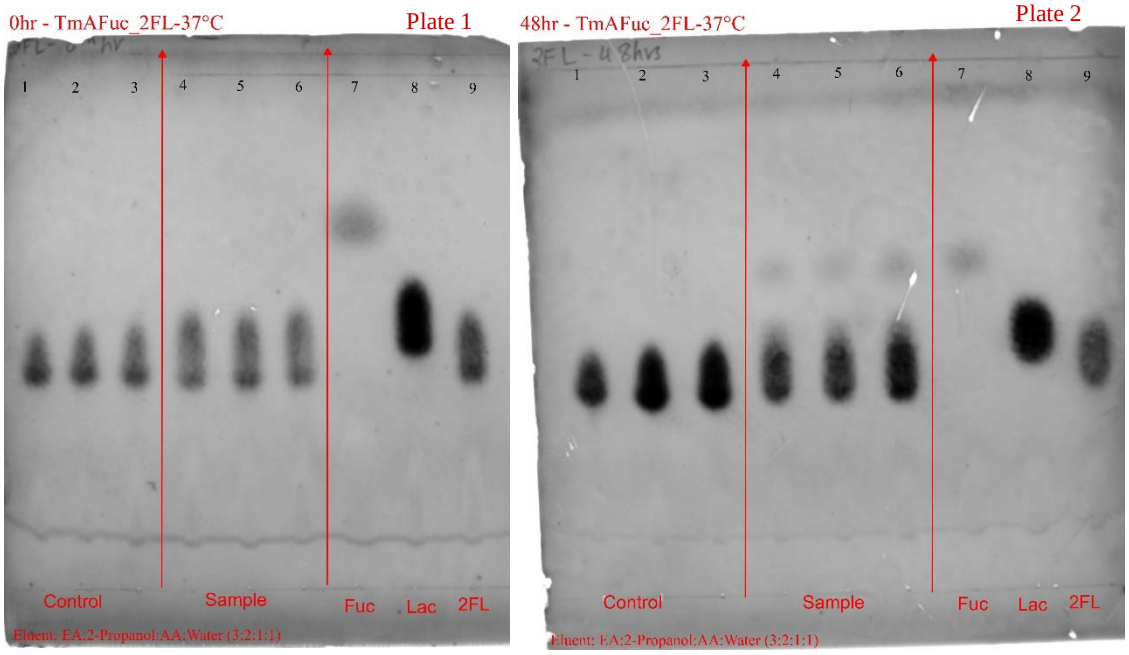


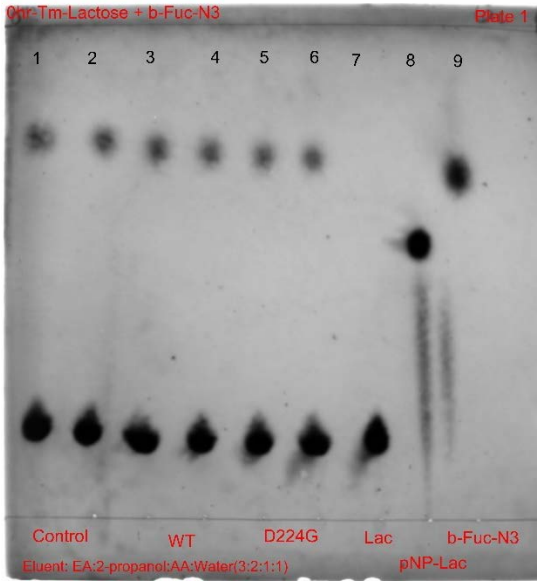
Figure I- TLC plate images showing the slow hydrolysis of 2'-FL in the presence of TmαFuc_WT at 37°C and hence, validating no significant growth of BL21_pEC_Tm0306_WT cell lines for growth studies discussed under section 2.2. 2 µL of each sample was spotted on the TLC plate. Plate 1 (left) shows the zeroth hour reading showing presence of 2'-FL. Plate 2 (right) shows very low hydrolysis of 2'-FL even after a duration of 48 hours. Lanes 1, 2, 3 are the control with no enzyme, lanes 4, 5, 6 are the reaction mixture at respective timepoint of the reaction, lanes 7, 8 and 9 are the fucose, lactose and 2'-FL markers.

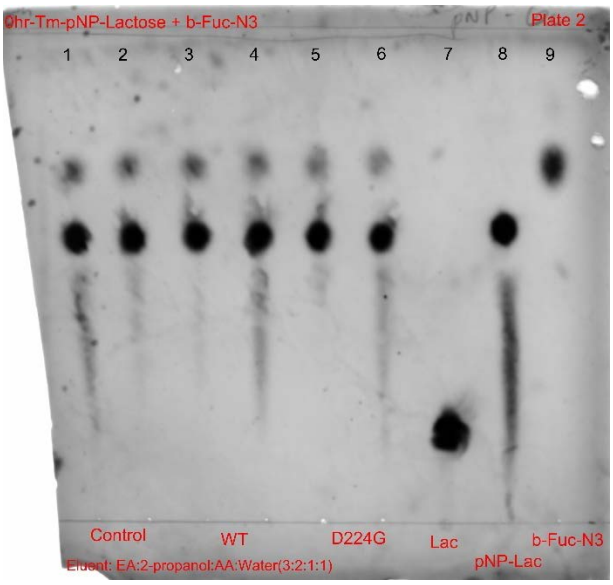
J. GS reaction using β-fucosyl azide and lactose/pNP-lactose

Timepoint readings were taken at 0^{th} hour, 24 hours, 48 hours and 72 hours (shown under section 4.2.3) of the GS reaction. TLC plate images of the GS reaction using β-fucosyl azide as donor and lactose (plate 1-see left)/ pNP-lactose (plate 2- see right) as acceptor with WT and D224G mutant are shown here. The markers lactose (lane 7) and pNP-lactose (lane 8) and β-fucosyl azide (lane 9) were also spotted. The experiment was conducted in duplicates and lanes 1, 2 are control without any enzyme, lanes 3, 4 are samples with both donor and acceptor with TmαFuc_WT enzyme and lanes 5, 6 are samples with both donor and acceptor with TmαFuc_D224G enzyme, respectively for each plate. Both plates show

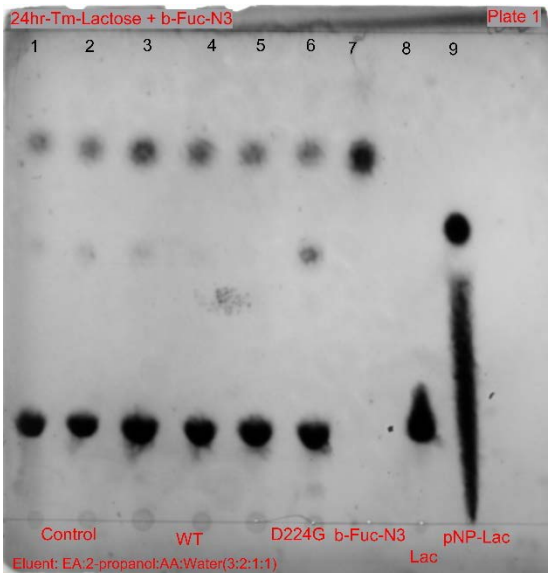
no GS products, the expected products are fucosyllactose and fucosyllactose-pNP, respectively for each plate.

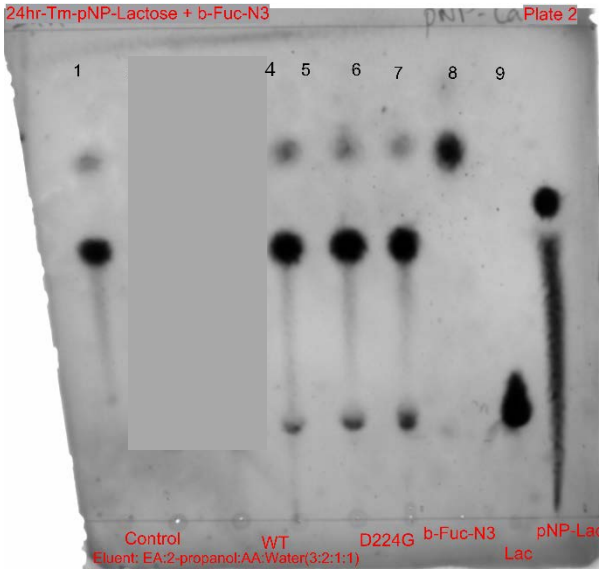
0th hour reading:





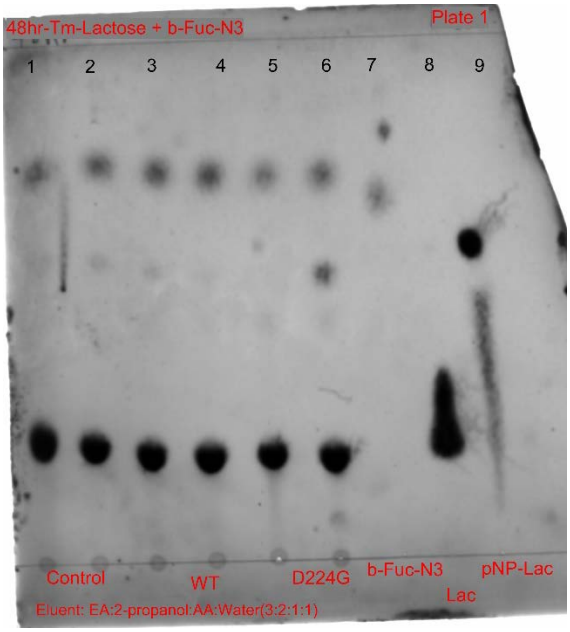
24-hour reading:

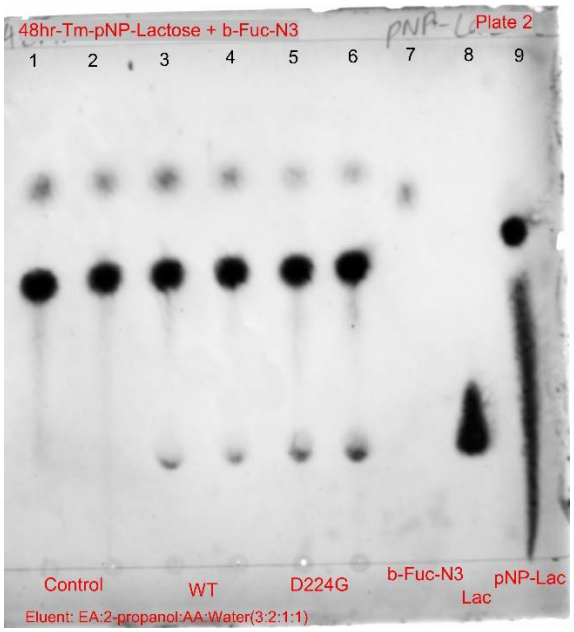




Lane 2 and 3 spotted wrongly on plate 2.

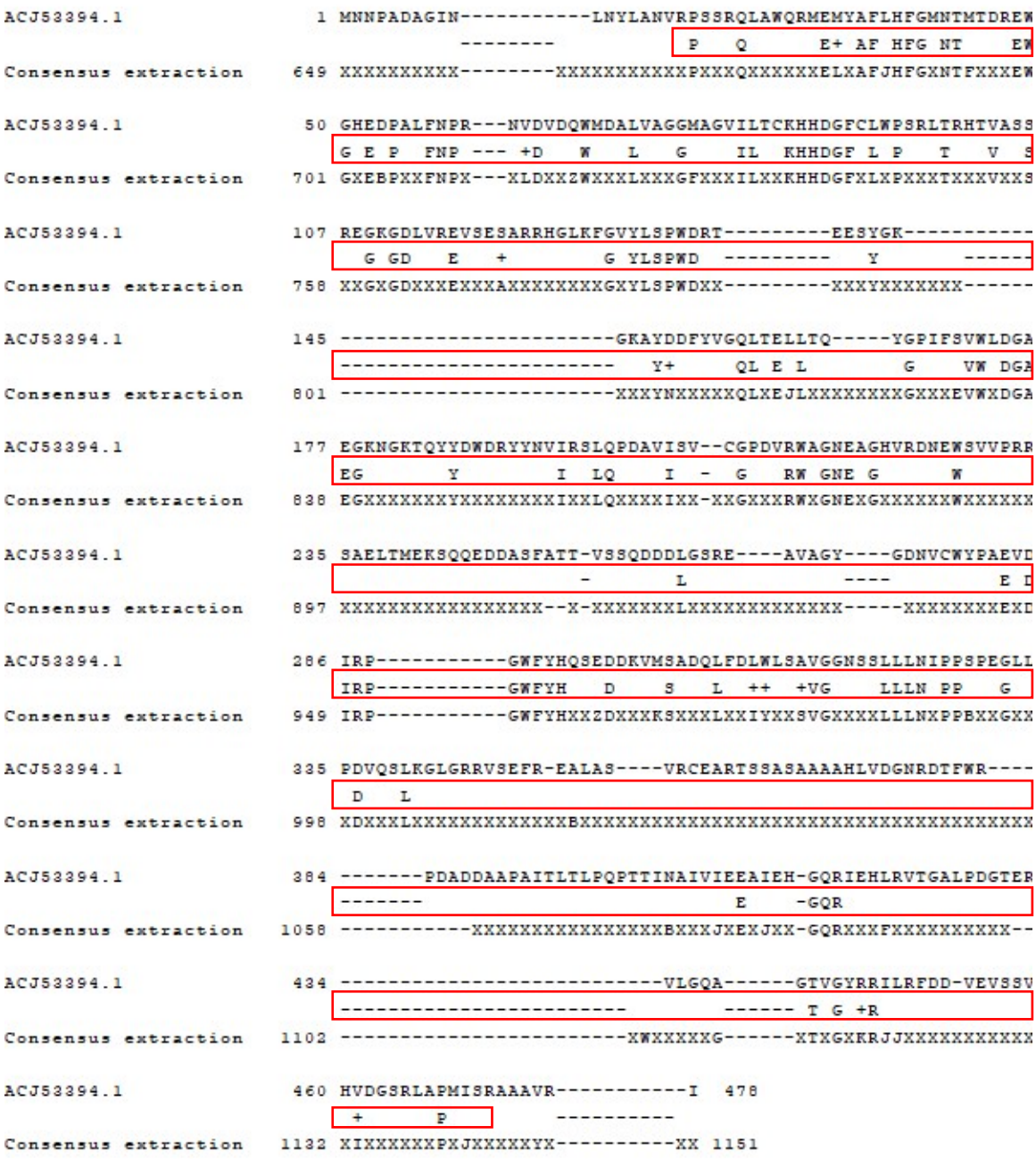
48-hour reading:





K. Conserved residues of *Blon2336* clade found using MSA (Clustal W)

Here ACJ53394.1 represents the protein produced by the *Blon2336* gene and the consensus sequence is developed based on the most common amino acid sequence in the clade of *Blon2336* extracted from the phylogenetic tree and later aligned using Clustal W alignment



technique. The conserved residues are shown in the between the two sequences shown using a red box.

L. AutoDock detailed procedure:

Make sure your working directory path does not have any kind of "run" in the name.

Running the simulations using Autodock:

1. Generate both protein and ligand.pdb files. Both the pdb files should be separate.
2. Open Autodock Tool 1.5.6
3. First lets us change the default directory of the autodock to our directory.
    a. File --> preferences --> set
    b. Under the "startup directory" option, put your working directory file path and click set.
    c. Dismiss

Protein setup:

1. Now, we need to load the protein.pdb file into the software
    a. File --> Read Molecule
    b. Load the protein.pdb file
2. We need to add hydrogen to the protein.
    a. Edit --> Hydrogens --> Add
    b. In the dialogue box which appears, choose "Polar only" and click ok
3. We need to add charges to protein
    a. Edit --> Charges --> Compute Gasteiger --> Ok

4. We can save the charged protein in pdbqt format. (Autodock always save them as pdbqt format because it has been tuned to use that format)

    a. Grid --> Macromolecule --> Choose

    b. In the dialogue box which appears, choose the protein and click ok

    c. A new dialogue box appears to save the file. Save as protein.pdbqt

Grid Box setup:

1. Now, we need to define the grid box. This is the arena of action where the ligand will be docked and different conformations will be tested.

    a. Grid --> Grid Box

    b. In the dialog box, input the "Center grid box" coordinates. This is chosen based on the catalytic nucleophile or catalytic acid/base atom residue coordinates from pdb file. View the pdb file in word and get the coordinates and input them here.

    c. Next, change the grid box dimensions to fit the active site completely.

    d. File --> Close saving changes. (This is very important. If you close without doing this, the changes will not be saved)

2. Save the grid you have chosen.

    a. Grid --> Output --> Save GPF

    b. Save it as grid.gpf

Ligand setup:

1. Click on Ligand --> Input --> Open
2. In the dialog box, change file type as "all" and choose the ligand.pdb
3. Add hydrogens and charges same as protein
4. Click on Ligand --> torsion tree --> Detect root

(In case we need to change the torsions, we can use the additional options in Torsion tree option)

5. Now, we need to save the ligand
    a. Ligand --> Ouput --> Save as pdbqt

Running the Grid:

1. Run --> AutoGrid
2. In the dialog box,
    a. Choose the program pathname and click on Autogrid4 executable file
    b. Choose the parameter filename for grid.gpf file. (Once you browse the filepath, the "Log Filename" entry will be automatically be filled.
    c. Click on Launch
    d. This step should take some time to run (around 2-3 minutes). If the simulation ends immediately, there is some problem which can be view in the python shell window (black window which looks like cmd window).

To know if the grid ran properly, look in the working directory, a number of "*.map" files would have been generated.

Running the Autodock:

1. Before running the autodock, we need to perform the following steps.
    a. Docking --> Macromolecule --> Set Rigid Filename --> Choose the protein.pdbqt file
    b. Docking --> Ligand --> Choose --> Choose the ligand name in the dialogue box which appears.
    c. Docking --> Search parameter --> Genetic Algorithm.

(This is the place where we can choose number of configurations etc. Other parameters can be played around for changing the simulation algorithm)

d. Docking --> Output --> Lamarckian GA --> Save the file as dock.dpf

2. Run --> Autodock

a. Similar to Autogrid, choose the appropriate files and click launch.

b. This again should take some time

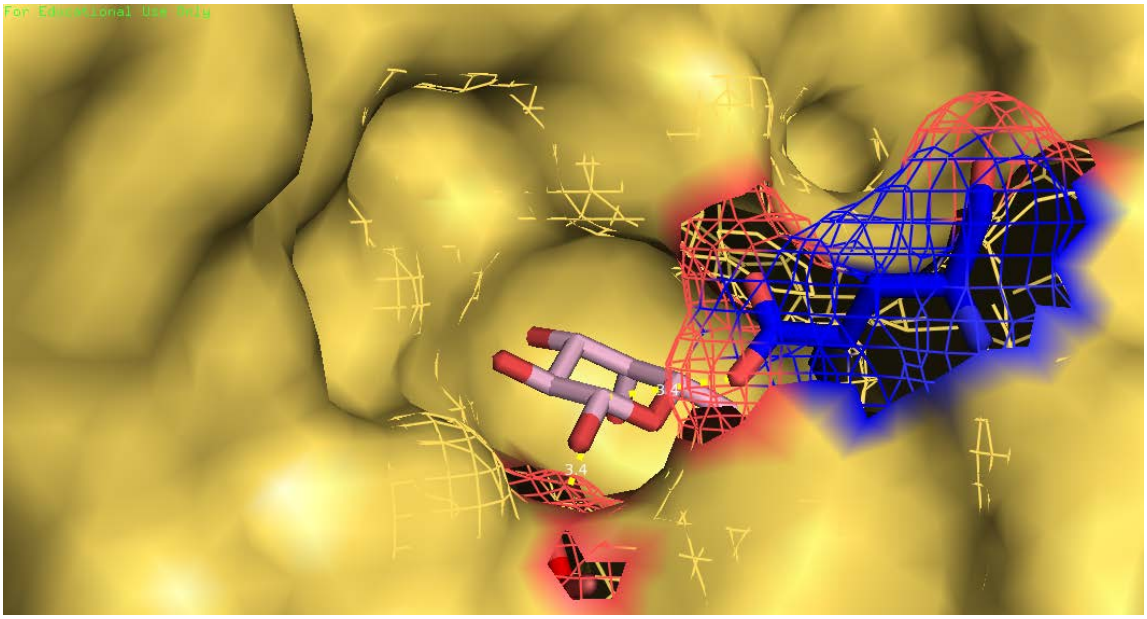
<u>Visualizing the results</u>:

1. Click on Analyze --> Docking --> Open --> Choose dock.dlg

2. Analyze --> Macromolecule --> Choose --> Choose protein

3. Analyze --> Conformations --> Play, ranked by energy

M. Validation structure for Fucose:



Fucose in 1ODU-crystal structure