工 學 博 士 學 位 論 文

Strategy for protein engineering of glycosyltransferase using hybrid approach and biosynthesis of sialyl and fucosyl oligosaccharides

시알릴/푸코 올리고당 생산을 위한 당전이 효소의 하이브리드 단백질 공학적 변이 전략 및 생합성 개발에 관한 연구

2016 년 2 월

서울大學校 大學院

化學生物工學部

崔 允 姬

Strategy for protein engineering of glycosyltransferase using hybrid approach and biosynthesis of sialyl and fucosyl oligosaccharides

A Thesis
Submitted to the Faculty of Seoul National University
by

Yun Hee Choi

In Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy

Advisor: Professor Byung-Gee Kim, Ph. D.

February, 2016

Interdisciplinary Program for Bioengineering
Seoul National University

ABSTRACT

Strategy for protein engineering of glycosyltransferases using hybrid approach and biosynthesis of sialyl and fucosyl oligosaccharides

Yun Hee Choi

Interdisciplinary Program for Bioengineering
The Graduate School
Seoul National University

For the efficient production of sialyl- and fucosyllactose from human milk oligosaccharides, protein engineering of glycosyltransferases using hybrid approach was developed and their *in vivo* /*in vitro* biosynthesis were studied.

Firstly, protein engineering of sialyltransferases (STs) was performed to increase the yield and productivity for the *in vitro* enzymatic synthesis of α2,3-sialyllactose (Neu5Ac(α2,3)Galβ1,4Glc (3′-SL)) and α2,6-sialyllactose (Neu5Ac(α2,6)Galβ1,4Glc (6′-SL)). Each SL could be produced by using a combination of the enzymatic synthesis of cytidine monophosphate *N*-acetylneuraminic acid (CMP-Neu5Ac) and ST reaction. In the large-quantity production of 3′-SL and 6′-SL using STs, there are major hurdles to overcome for further improvement in yield and productivity of the enzyme reactions. Specifically, *Pasteurella multocida* α2,3-sialyltransferase (α2,3PST) forms a by-product to a certain extent, owing to its multifunctional activity at pH below 7.0 for optimum pH of CMP-Neu5Ac synthesis, and *Photobacterium damselae* α2,6-sialyltransferase (α2,6PdST) shows relatively low ST activity. In this study, α2,3PST and α2,6PdST were successfully engineered using a hybrid approach that

combines rational design with site-saturation mutagenesis. Narrowly focused on the substrate-binding pocket of the STs, putative functional residues were selected by multiple sequence alignment and alanine scanning, and subsequently subjected to site-saturation mutagenesis. In the case of α2,3PST, R313N single mutation improved its activity slightly (by a factor of 1.5), and further improvement was obtained by generating the double mutants (R313N/T265S and R313H/T265S) resulting in an overall 2-fold improvement in its specific α2,3 ST activity, which is mainly caused by the increase in k_{cat}. It was revealed that the R313 mutations to N, D, Y, H or T greatly reduced the α2,6 ST side-reaction activity of α2,3PST at below pH 7.0. In the case of α2,6PdST, single-mutation L433S/T and double-mutation I411T/L433T exhibited 3- and 5-fold enhancement of the α2,6 ST specific activity compared with the wild-type, respectively. These mutants of α2,3PST and α2,6PdST were used efficiently to improve the reaction process through the construction of one-pot batch reaction and enhancement of the yield.

Secondly, α1,2- and α1,3-fucosyltranferases (α1,2-FucT and α1,3-FucT) were cloned from *Helicobacter pylori* and their regio-selectivities were investigated for the production of α1,2-fucosyllactose (Fuc(α1,2)Galβ1,4Glc, 2′-FL) and α1,3-fucosyllactose (Galβ1,4Fuc(α1,3)Glc, 3-FL), respectively. For economic production of FLs, one-pot reaction was constructed by combining with FucT and GDP-fucose synthesis using an enzyme of salvage pathway originated from *Bacteroids fragilis*. Engineering of FucTs was conducted to improve the activity and soluble expression level in *E. coli* since the FucT reactions were rate-limiting steps in the one-pot reaction. In the case of α1,2-FucT, the yield of 2′-FL was increased up to 90% based on the 5 mM of GDP-fucose and productivity was also improved by a factor of 14 through codon optimization and fusion protein expression. For the α1,3-FucT, codon optimization and systematic truncation of the protein at the C-terminus were conducted to yield 150-200 mg/L of soluble protein

of α1,3-FucT and resulting in more than an 18-fold increase in the 3-FL yield. To improve the low level of enzyme catalytic activity for lactose, focused directed evolution was attempted using a semi-rational approach that combines structure-guided computational analysis and subsequent iterative saturation mutagenesis (ISM). In order to select the functional residues in active site/substrate binding site, docking simulation was used together with HotSpot Wizard. The selected residues from each α-helix were clustered, and ISM was performed for each cluster in parallel. As a result, a mutant with quadruple mutations was generated, which showed the synergistic effects, i.e. 15.2-fold improvement in specific activity relative to the truncated wild-type. The mutation increased its binding affinity for lactose and k_{cat} values for lactose and GDP-fucose. The quadruple mutant was successfully applied in the *in vitro* synthesis of 3-FL with an improved yield and productivity (>96% yield based on 5 mM of GDP-Fuc within 1 h).

Finally, metabolic engineering was carried out for the *in vivo* production of 3-FL in engineered *E. coli* with the α1,3-FucT mutant. The genes responsible for the production of GDP-fucose were overexpressed, while genes related with degradation of substrates were knocked-out. In result, 2 g/L of 3-FL could be produced in *E. coli* after 76 h by fed-batch culture from the glycerol as a carbon source when lactose and L-fucose were used for substrates.

The results demonstrate that the protein engineering strategies for hybrid approach can be utilized for the development of mutants of various glycosyltransferases for engineered substrate specificities and improved catalytic activity. In addition, the generated mutants and economical biosynthesis of sialyl- and fucosyloligosaccharides could be used efficiently for the mass production of various functional oligosaccharides with enhanced productivity and yield.

Keywords: Human milk oligosaccharides, Sialyllactose, Fucosyllactose,

Sialyltransferase, Fucosyltransferase, Protein engineering, Hybrid approach

Student number: 2010-30265

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1.

Introduction

1.1 Overview of human milk oligosaccharides (HMOs)

1.1.1 Compositions and structure of HMOs

Human milk has been reported to have a lot of biological functions, which are closely related to its composition (Figure 1.1). In particular, human milk oligosaccharides (HMOs) are the third most abundant component in human milk by accounting for 5-15 g/L after lactose and lipid (Bode 2012; Zivkovic et al. 2011). It has been known that HMOs have important healthy effects to breast fed infants.

HMOs are composed of the five monosaccharides glucose (Glc), galactose (Gal), N-acetylglucosamine (GlcNAc), fucose (Fuc), and N-acetylneuraminic acid (Neu5Ac). The glycan units allows structural diversity of oligosaccharide, resulting over 200 different HMOs structures that differ in size, charge, and sequence (Bode 2012; Kunz et al. 2000; Wu et al. 2010b). The diversity and composition of HMO are dependent on expression of certain glycosyltransferase (GTs) by blood group characteristics for individuals.

HMOs contained a lactose (Gal β1,4Glc) reducing end elongated with β1-3- or β1-6-linked lacto-*N*-biose (Galβ1-3GlcNAc-, type 1 chain) or *N*-acetyllactosamine (Galβ1-4GlcNAc-, type 2 chain). Lactose or the elongated oligosaccharides can be fucosylated in α1-2, α1-3, or α1,4 linkage and/or sialylated in α2-3 or α2-6 linkage (Bode and Jantscher-Krenn 2012; Ninonuevo et al. 2006) (Table 1.1). As shown in Table 1.2, Sialyloligosaccharides (SOs) account for 10 to 20% of total HMO. The major SOs in HMO include 3′-sialyllactose (3′-SL), 6′-sialylactose (6′-SL), sialyl lacto-*N*-tetraose a (LST a), sialyl lacto-*N*-tetraose b (LST b), sialyl lacto-*N*-tetraose c

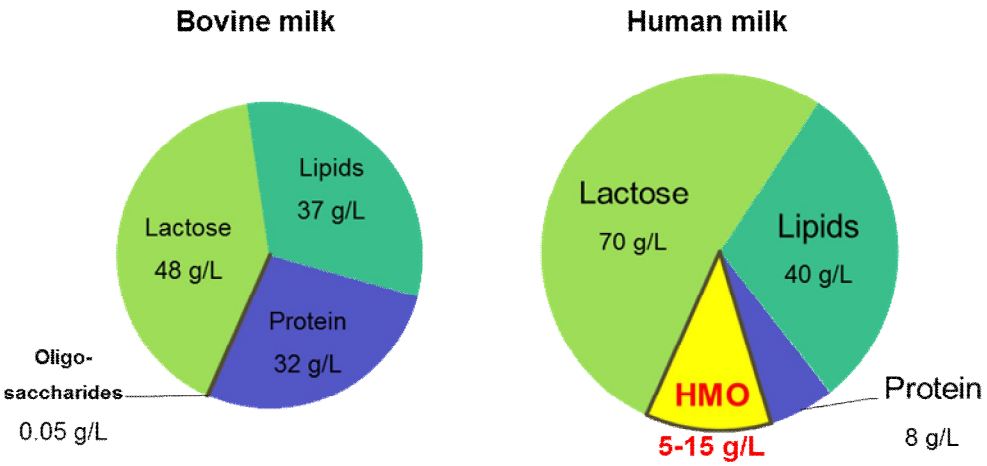


Figure 1.1 Composition of Human milk and bovine milk

In human milk, oligosaccharides account for 5-15 g/L indicating 100-300 times higher concentration than those of bovine milk.

1 **Table 1.1 Structure and their common abbreviations**

HMO	Abbreviation	Structure
2′-fucosyllactose	2′-FL	Fuc-α-(1→2)-Gal-β-(1→4)-Glc
3-fucosyllactose	3-FL	Gal-β-(1→4)-[Fuc-α-(1→3)]-Glc
lacto-*N*-tetraose	LNT	Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
lacto-*N*-neotetraose	LNnT	Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
lacto-*N*-fucopentaose I	LNFP I	Fuc-α-(1→2)-Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
lacto-*N*-fucopentaose II	LNFP II	Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
lacto-*N*-fucopentaose III	LNFP III	Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
lacto-*N*-difucohexaose I	LNDFH I	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
lacto-*N*-difucohexaose II	LNDFH II	Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-[Fuc-α-(1→3)]-Glc
3′-sialyllactose	3′-SL	Neu5Ac-α-(2→3)-Gal-β-(1→4)-Glc
6′-sialyllactose	6′-SL	Neu5Ac-α-(2→6)-Gal-β-(1→4)-Glc
sialyl lacto-*N*-tetraose a	LST a	Neu5Ac-α-(2→3)-Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc

sialyl lacto-*N*-tetraose b	LST b	Gal-β-(1→3)-[Neu5Ac-α-(2→6)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
sialyl lacto-*N*-tetraose c	LST c	Neu5Ac-α-(2→6)-Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
disialyl lacto-*N*-tetraose	DS-LNT	Neu5Ac-α-(2→3)-Gal-β-(1→3)-[Neu5Ac-α-(2→6)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc
	Type I chain	Gal-β-(1→3)-GlcNAc
	Type II chain	Gal-β-(1→4)-GlcNAc

Table 1.2 Major HMOs and their compositions

Components	Amount (g/liter)	
	Human milk	**Cow's**
Lactose	55-70	40-50
Oligosaccharides	**5-15**	**0.05**
Low M.W. oligosaccharides	5.0-8.0	minute
Lacto-N-tetraose (LNT, type1)	0.5-1.5	minute
% Fucosylated OS/ total OS	**50-80%**	**~1%**
2′-Fucosyllactose (2′-FL)	2.2-2.6	-
3-Fucosyllactose (3-FL)	0.8-1.0	-
Lacto-N-fucopentaose I (LNFP I)	1.2-1.7	-
Lacto-N-difucohexaose I (LNDFH I)	0.1-0.2	-
Lacto-difucotetraose (LDFT)	0.4	-
% Sialylated OS/ total OS	**10-20%**	**~70%**
6′-Sialyllactose (6′-SL)	0.3-0.5	0.03-0.06
3′-Sialyllactose (3′-SL)	0.1-0.3	
LS-Tetrasaccharide a (LST a)	0.03-0.2	minute
LS-Tetrasaccharide c (LST c)	0.1-0.6	minute
Disialyllacto-N-tetraose (DS-LNT)	0.2-0.6	minute

(LST c), and disialyl lacto-*N*-tetraose (DS-LNT). Among the SOs, 3′-SL and 6′-SL have been reported to show various biological function, and they are contained 0.1 to 0.5 g/L in HMO, showing 3-16 times higher amount relative to those of cow milk (Kobata 2010; Kunz et al. 2000).

Fucosyloligosaccharides (FOs) are major components of HMO and account for 50 to 80% of the total 5-15 g/L of HMO (Bode 2012; Zivkovic et al. 2011). The major FOs in HMO contain 2′-fucosyllactose (2′-FL), 3-fucosyllactose (3-FL), lacto-*N*-fucopentose (LNFP I, II, III, V), lacto-*N*-difucohexaose (LNDFH I, II), and lacto-difucotetraose (LDFT) (Chaturvedi et al. 2001; Newburg and Grave 2014). FLs, fucose linked lactose, recently have been studied for their production and a diversity of biological functions.

1.1.2 Biological functions of sialyllactose (SL) and fucosyllactose (FL)

As illustrated in Figure 1.2, sialyl- and fucosyloligosaccharides are known to have various important functions in human health and disease. Particularly, sialyllactose (SL) moieties are found in gangliosides in neuron and brain tissues, which are a component of cell plasma membrane that modulates cell signal transduction events such as brain development and memory formation (Antoine et al. 2003; Wang 2009). Sialyl Lewis X/A molecules containing sialic acid moiety are well-known carbohydrate ligand molecules recognized by selectins expressed on endothelial cells, and their bindings result in inflammation to recruit leucocytes by helping the adhesion of the leukocytes to the inner wall of a blood vessel (Lowe 2003).

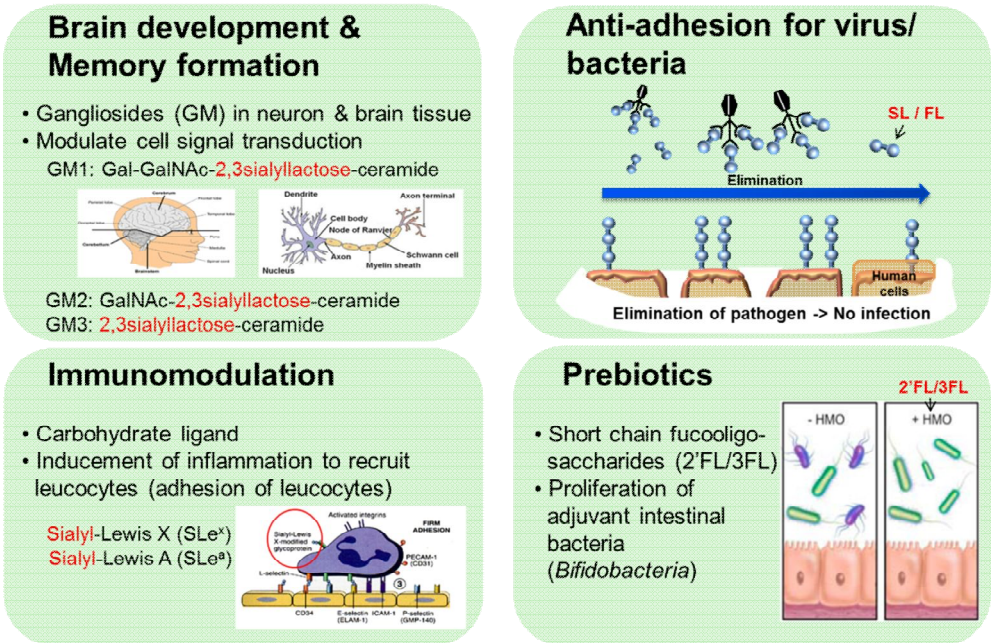


Figure 1.2 Biological functions of SLs and FLs of sialyl and fucosyloligosaccharies

SLs and FLs have various biological functions and human healthy effects.

Sialyllactose is known as one of the major components of human milk oligosaccharides (HMO), and can also function as a neutralizing reagent of toxins produced by enteric bacteria (Idota et al. 1995; Sinclair et al. 2008). They are also known as potent inhibitors of bacterial or viral adhesion to the epithelial surface in the initial stages of the infection process (Kunz et al. 2000; Lehmann et al. 2006). 3′-SL and 6′-SL containing Neu5Ac(α2,3)Gal and Neu5Ac(α2,6)Gal sequences are differentially recognized by avian and human influenza viruses, respectively (Hideshima et al. 2013; Lehmann et al. 2006; McCullough et al. 2012), through interaction with viral hemagglutinins. Synthetic SLs can be used for various applications, such as functional food ingredients, nutraceutical medicine, and virus adsorbent materials, so that their large-scale enzymatic production is in great demand (Drouillard et al. ; Endo et al. 2000; Priem et al. 2002).

Fucosyloligosaccharides (FOs) have been reported to have several important functions in human health and disease, including anti-adhesive effects against pathogens and prebiotic functionality. Among the FOs, fucosyllactoses (FLs) are known to be potent inhibitors of bacterial or viral adhesion to human epithelial cells by acting as decoys that lure pathogens and viruses. 2′-FL, the most abundant α1,2-fucosyloligosaccharide in HMO, is reported to exhibit anti-adhesion activity against for *Campylobacter jejuni* of diarrhea pathogen (Morrow et al. 2004; Ruiz-Palacios et al. 2003), *Salmonella* enteric serotype *Typhimurium* (Chessa et al. 2009), Enterotoxigenic *Escherichia coli* (Newburg et al. 1990), *Helicobacter pylori* (Magalhaes and Reis 2010) and noroviruses (Newburg et al. 2005). The anti-adhesive activities of 3-FL that is α1,3-fucosyloligosaccharide in HMO, have been also described for *Pseudomonas aeruginosa* (Weichert et al. 2013), enterotoxigenic

Escherichia coli, *Salmonella fyris* diarrheal pathogen (Coppa et al. 2006), and the respiratory viruses (H1N1 influenza virus, RSV, HPIV3) (Buck et al. 2014).

In addition, FLs play a role as prebiotics to construct a healthy gut ecosystem by selectively stimulating the growth of beneficial microorganisms such as *Bifidobacterium longum*, *B. bifidum* secreting α1,2-fucosidases (Katayama et al. 2004) or α1,3-fucosidases (Asakuma et al. 2011; Ashida et al. 2009) while also suppressing the growth of harmful microorganisms such as *Clostridium perfringens*, *Enterobacter cloacae*, and *E. coli* (Yu et al. 2013a; Yu et al. 2013b). The prebiotic effect of FOs is due to the HMO-related gene cluster for only *Bificobacteria*, in particular for *B. longum* and *B. bifidum*. The gene cluster includes glycosidase (sialidase, fucosidase, galactosidase, and hexosamidase) and carbohydrate transporters for importing and metabolizing HMOs. Once bifidobacteria deconstruct HMOs such as FOs via glycosidase, metabolized monosaccharides are used for central metabolic pathway (Zivkovic et al. 2011).

In addition, Lewis X (Le^x), which is lewis motif corresponding to homologous structure of LNFP III (Galβ1,4[Fuc(α1,3)]GlcNAc(β1,3)Gal(β1,4)Glc) of HMO, play an important role as carbohydrate epitope involved in inflammation as well as adhesion and metastasis of cancer cells. Particularly, sialyl Le^x is the minimum glycan epitope required to facilitate selectin-ligand interaction for leukocyte extravasation. The sialyl Le^x could be an inhibitor of selectin-mediated cell adhesion of abnormal immune system that triggers an inappropriate inflammatory signal and causes damage to its own tissues (Bode and Jantscher-Krenn 2012; Soriano del Amo et al. 2010).

Synthetic FLs can be used for various applications, such as functional food ingredients, nutraceutical medicine, and virus adsorbent materials, so that their large-scale production is in great demand.

1.2 Glycosyltransferase (GT) technology

1.2.1 Structural folds in GTs

Glycosyltransferases (GTs; EC 2.4.x.y) catalyze the regio- and stereo-specific transfer of sugar from an activated sugar donor to a variety of acceptors including glycans, proteins, lipid, and small molecules (Chang et al. 2011). According to the structures, reaction mechanisms, and sequences, GTs are classified from folds into clans, families, and subfamilies, as illustrated in Figure 1.3 (Coutinho et al. 2003). Subfamily level could provide maximal functional prediction but have hundreds of different categories. At the other end of the spectrum, fold level comprising GT-A and GT-B enables the detection of distant evolutionary relationship, but provide little functional predictive power (Coutinho et al. 2003).

To date, X-ray crystal structures are available for about 140 GTs. According to the structural information, GTs have been classified into two general fold, called GT-A and GT-B. GT-A fold is consisted of two closely abutting β/α/β Rossmann-like domains. The β/α/β domains are associated tightly, which resulted in formation of flanked β-sheet in active site (Breton et al. 2006; Lairson et al. 2008). All GT-A fold enzymes have DxD motif (Asp-X-Asp) required for a divalent cation for activity (Breton and Imberty 1999; Breton et al. 2006). There are distinct domains of donor nucleotide and acceptor binding sites for GT-A although two domains are

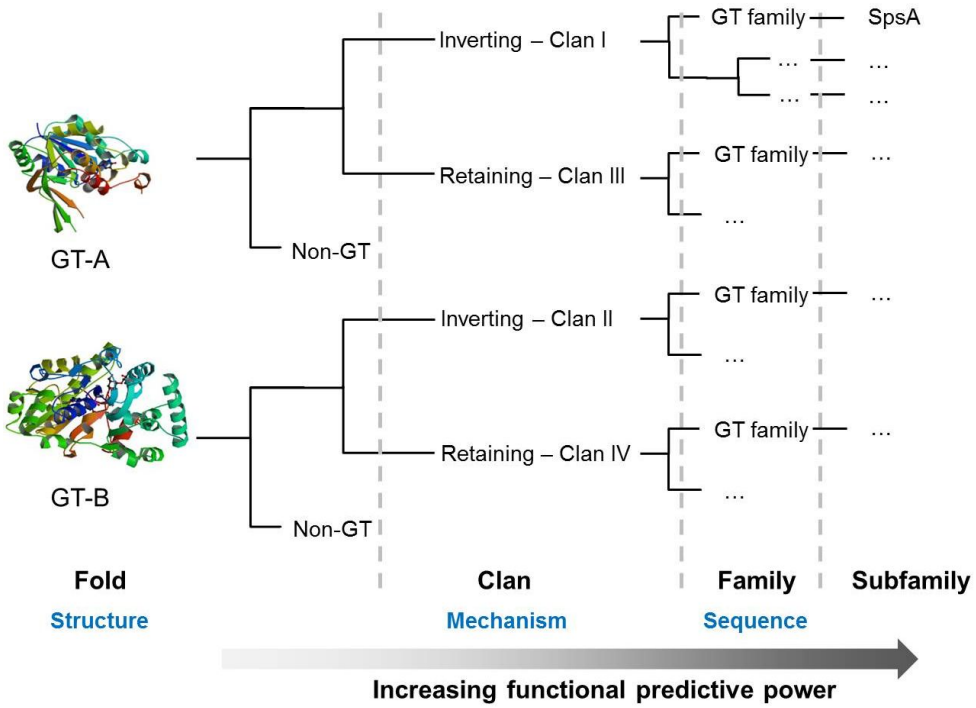


Figure 1.3 Classification of GTs

GT-A and GT-B fold are classified into clan according to the reaction mechanism (i.e. inverting or retaining). 97 GT families belonging to each clan are divided into subfamily.

associated tightly (Unligil and Rini 2000).

On the other hand, GT-B fold consists of two separate Rossmann domains with a connecting linker region and a catalytic site located between the domains. Like the GT-A fold, the GT-B fold enzymes consist of two β/α/β Rossmann-like domains, however, there is a distinct difference with GT-A fold. In the case of GT-B fold, the two domains are less tightly associated and facing each other with the active site in the center (Lairson et al. 2008). Particularly, the two domains divided distinctly into N-terminal and C-terminal domain are responsible for binding of acceptor and donor substrates, respectively.

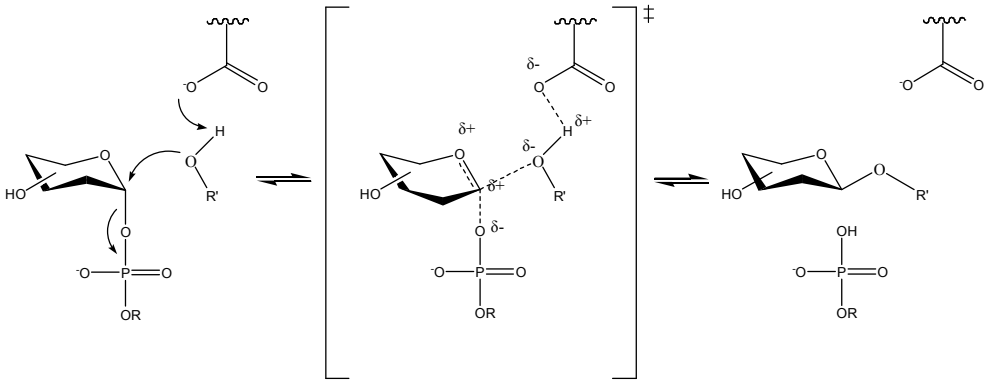
In addition, novel folds of GTs have been identified for several GT families, which comprise a large hydrophobic transmembrane helices and utilize a lipid phohphate-activated donor substrates (Breton et al. 2012).

1.2.2 Reaction mechanisms and stereospecificity of GTs

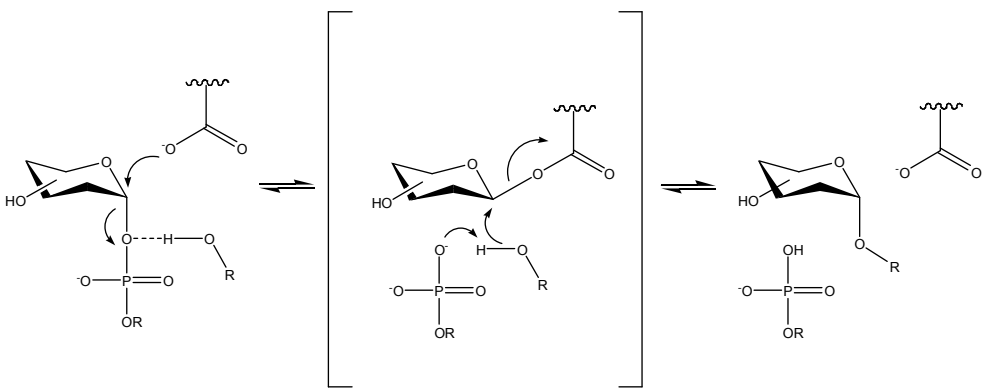
GTs can be classified as either retaining or inverting enzymes according to the stereochemistry of the substrates and reaction products. The anomeric configuration of the product can either be retained or inverted relative to that of donor substrate. The transfer of sugars by GTs is regiospecific and stereospecific by the stereochemical mechanisms as mentioned above (Breton et al. 2012).

As illustrated in Figure 1.4 (A), inverting GT reactions are suggested to occur in a single displacement S_N2 mechanism via an oxocarbenium-ion like transition state. The side chain of acidic catalytic key residue such as aspartate/glutamate (Ni et al. 2007; Ramakrishnan and Qasba 2010; Sun et al. 2007) or histidine (Lazarus et al. 2011) serves as a base catalyst that deprotonates the incoming nucleophile of

A



B



C

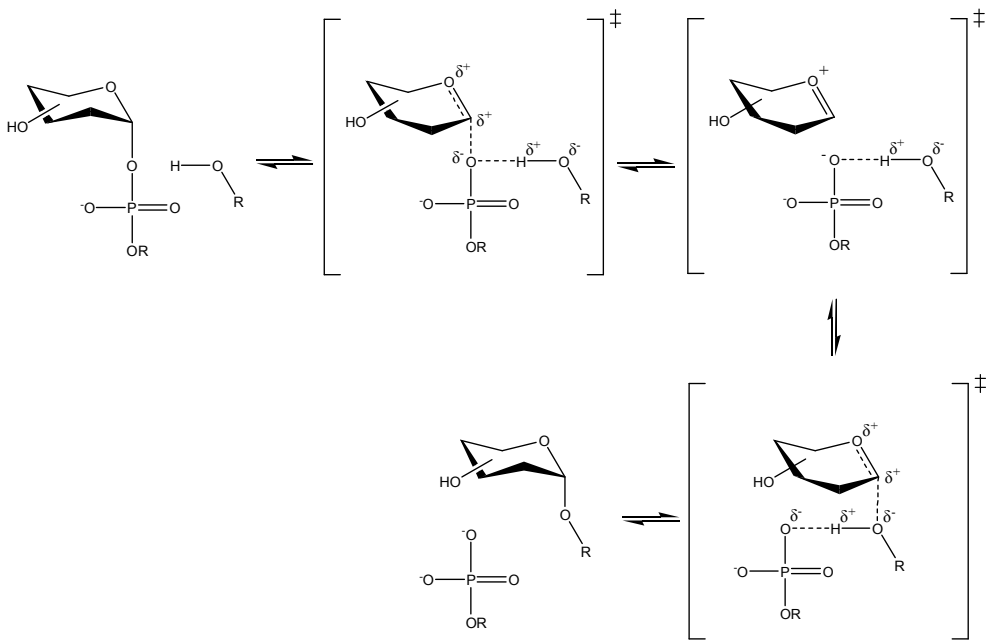


Figure 1.4 Reaction mechanism proposed for inverting and retaining GTs

(A) Inverting GT reactions are proposed to occur in single displacement with formation of an oxocarbenium-ion like transition state. (B) Double displacement mechanism for retaining GTs involves formation of a covalently bound glycosyl-enzyme intermediate. (C) S_Ni-like mechanism for retaining GT involves formation of discrete ion pair intermediates.

acceptor, facilitating direct displacement of the activated phosphate leaving group (Chang et al. 2011; Lairson et al. 2008). For GT-B fold enzyme, conserved catalytic key residue is located in the cleft between the two Rossmann domains.

The mechanism for retaining GTs is controversial and is proposed to occur via double-displacement mechanism or a S_Ni-like mechanism. The double-displacement mechanism involves the formation of a covalently bound glycosyl-enzyme intermediate as shown in Figure 1.4 (B) (Breton et al. 2012; Lairson et al. 2008). When the active site of GT-A or GT-B lacks nucleophile suitably positioned for double displacement, S_Ni-like mechanism for retaining GT occurs by formation of discrete ion pair intermediates (Figure 1.4 (C)). For the S_Ni-like mechanism (internal return), decomposition of leaving group causes production of a nucleophile to allow attack on the same face as the leaving group (Errey et al. 2010).

1.2.3 Sequence and structure similarities between GTs

GTs are classified into families based on amino acid similarities. Currently, there are almost 204000 GTs in Carbohydrate-Active Enzymes database (CAZy, http:// ww.cazy.org) comprising 97 families (Lombard et al. 2014). Classification according to sequence similarities allows logical grouping of enzymes of different EC numbers into specific families and offers insights into the divergent evolution of enzyme families (Campbell et al. 1997).

In general, sequence similarities between families are extremely low (Breton et al. 2012; Campbell et al. 1997). On several occasions, GTs acting on similar substrates with the same mechanism are classified into different families. For instance, α1,2-

fucosyltransferase (α1,2-FucT) and α1,3-fucosyltransferase (α1,3-FucT) from *H. pylori* belong to the different family, GT 11 and GT10, respectively. Sequence similarities of them are extremely low although these enzymes transfer fucose from guanosine 5′-diphospho-β-L-fucose (GDP-Fuc) to an acceptor using the same inverting mechanism. Multiple sequence alignment with the sequences in the same family or between families allows for evolutionary analysis.

3D-structure comparison is a powerful method to detect evolutionary relationship in company with the sequence similarities. In the GT-A fold, the central β-sheet core of the C-terminus is topologically conserved, but significant variation in C-terminal β-strand (α/β), α-helices and/or loops contributes to acceptor binding and acceptor specificity (Chang et al. 2011). On the other hand, there is a noticeable structural conservation between protein members of the GT-B fold family. In particular, C-terminal domain corresponding to the nucleotide binding is quite conserved, but variations are observed in the loops and helices of N-terminal domain which is responsible for acceptor binding (Albesa-Jove et al. 2014; Breton et al. 2006). The variation of N-terminus is due to the evolution to accommodate very different acceptors. In macrolide GT OleD of GT-B fold, loops of N-terminus undergo a disordered to ordered transition in the presence of ligand, and neighboring α-helices form an internal hydrophobic cavity (Bolam et al. 2007). In addition, GTs undergo significant conformational changes upon substrate binding. Structural studies reveal that one or two flexible loop at the substrate binding site of the GTs undergo a significant conformational change from an open to a closed form. The loop as a lid covering the bound donor substrate creates an acceptor-binding site through the conformational changes by the binding of donor

substrate (Qasba et al. 2005). In the case of α1,4-galactosyltransferase LgtC from *Neisseria meningitidis* (GT-A fold, retaining), the two loops from opposite sides fold over the bound nucleotide donor substrate (Persson et al. 2001).

Similarly, the binding of donor to the GT-B fold causes a large movement of two Rossmann domains to make a closed conformation (Albesa-Jove et al. 2014; Ni et al. 2006), which lead to subsequent formation of the binding pocket for an acceptor. In the example of α2,3-sialyltransferase (α2,3-ST) from *Pasteurella multocida* for GT-B fold enzyme, nucleotide donor CMP only interact with the C-terminus, and the binding of CMP causes a large closure movement of the N-terminal domain toward the C-terminal domain. Upon binding of CMP, in addition, a short α-helix near the active site becomes disordered and trigger the closed conformation to form the binding pocket for the acceptor substrate (Ni et al. 2006).

1.3 Synthesis of functional oligosaccharides by GTs

1.3.1 Sialyltransferase (ST) and Fucosyltransferase (FucT)

α2,3-/ α2,6-Sialyllactose (3′-SL/6′-SL) can be synthesized by STs (EC 2.4.99) which transfer *N*-acetylneuraminic acid (Neu5Ac) from cytidine monophosphate *N*-acetylneuraminic acid (CMP-Neu5Ac) to a lactose acceptor substrate, the galactose moiety of lactose more specifically (Figure 1.5). α2,3-ST (EC 2.4.99.4) transfers Neu5Ac to C3-OH position, and α2,6-ST (EC 2.4.99.1) transfers Neu5Ac to C6-OH position, respectively. Especially, bacterial α2,3-STs and α2,6-STs are involved in GT42, GT52, GT80, and GT97 family.

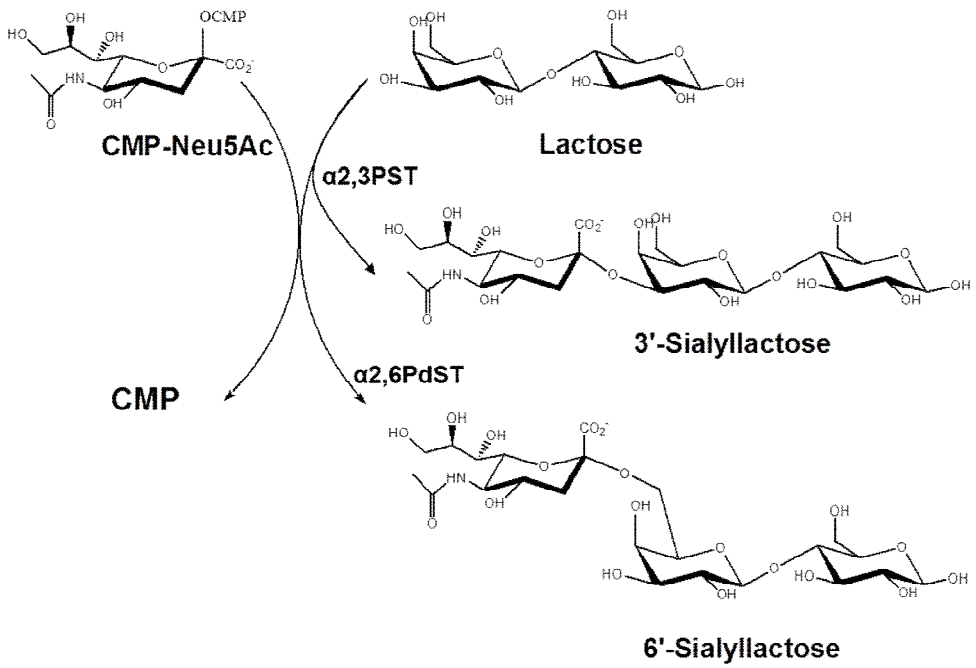


Figure 1.5 Synthesis of α2,3 and α2,6-sialyllactose by α2,3PST and α2,6PdST with CMP-Neu5Ac and lactose substrates

The biosynthesis of FLs (2′-FL/3-FL) involves the formation of Fucα1,2Galβ1,4Glc and Galβ1,4(Fucα1,3)Glc catalyzed by α1,2-FucT (EC 2.4.1.69) and α1,3-FucT (EC 2.4.1.152), respectively. The α1,2-FucT and α1,3-FucT catalyze the transfer of fucose from GDP-Fuc to the C2-OH of galactose and C3-OH position of glucose/GlcNAc moiety, respectively. In the case of α1,3-FucT, *N*-acetyllactosamine (Galβ1,4GlcNAc, LacNAc) is more favored than lactose (Galβ1,4Glc) as a acceptor substrate, resulting in generation of Lewis X (Le^x, Galβ1,4(Fucα1,3)GlcNAc) as a more preferred product than generation of 3-FL (Figure 1.6). The α1,3-FucTs from *H. pylori* belong to GT10 family showing inverting mechanism of GT-B fold. Whereas, the α1,2-FucTs from *H. pylori* are involved in GT11 family having inverting mechanism, but structural fold is not certain due to the lack of crystal structure.

1.3.2 Biological production of SLs and FLs

1.3.2.1 Synthesis of SLs

For the production of SLs using the cells, productions including a coupled microbial and single-cell method have been developed. In coupled microbial method, *two E. coli* stains responsible for CMP-Neu5Ac synthesis and sialic acid transfer reaction, and *Corynebacterium ammoniagenes* for UTP production were used by whole cell reaction. When *E. coli* cells overexpressing the α2,3-ST of *Neisseria gonorrhoeae* were put into the CMP-Neu5Ac production system after CMP-Neu5Ac was accumulated at 17 g/L, 33 g/L of 3′-SL was produced from orotic acid, Neu5Ac, and lactose (Endo et al. 2000).

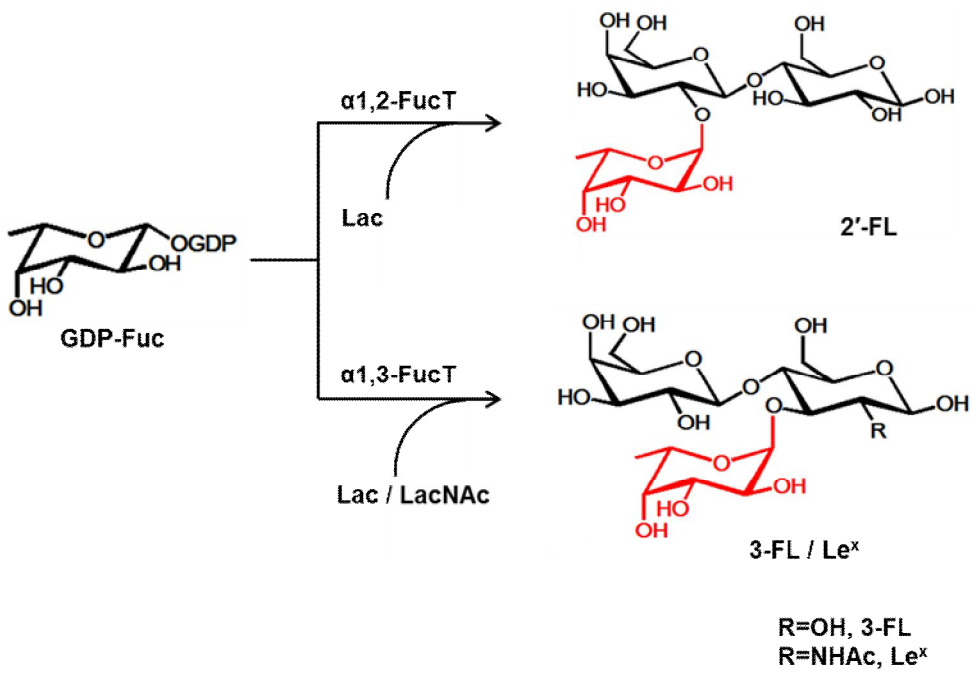


Figure 1.6 Synthesis of α1,2 and α1,3-fucosyllactose by α1,2-FucT and α1,3-FucT with GDP-Fuc and lactose/LacNAc substrates

However, the system was unsuitable for large-scale production, since the its high complexity and low productivity. Therefore, single cell method using the *E. coli* cells has been developed. In this system, NeuAc was produced via E. coli cells by expression of *N*-acetylglucosamine (GlcNAc) 2-epimerase and NeuAc synthetase. To prevent digestion of NeuAc, *E. coli* cell disrupted of NeuAc aldolase gene was used for production. After 22 h reaction with 120 g/L of GlcNAc in 5 L fermentor, 53 g/L of Neu5Ac was produced (Ishikawa and Koizumi 2010). SLs could be synthesized by *in vivo* production using living E. coli cells. Mutant *E. coli* strain devoid of genes for Neu5Ac degradation pathway was used with CMP-Neu5Ac synthesis enzymes (i.e. neuA, CMP-Neu5Ac synthase; neuB, sialic acid synthase; neuC, GlcNAc-6-phosphate-2-epimerase), which resulted in production of 25 g/L 3′-SL from glycerol and lactose (Fierfort and Samain 2008). However, it took a long time (70 h) to obtain the yield using this method, enzymatic *in vitro* synthesis of SLs has been considered therefore.

Previously, we successfully accomplished an enzymatic process for CMP-Neu5Ac synthesis using a five-enzyme system (GlcNAc-2-epimerase, acetate kinase, CMP kinase, Neu5Ac aldolase, and CMP-Neu5Ac synthetase), which could efficiently and cheaply supply the donor substrate CMP-Neu5Ac for ST reaction (J.S. Woo 2008; Lee et al. 2002) (Figure 1.7). As the next step, the large-scale synthesis of 3′-SL and 6′-SL has been studied using a combination of the enzymatic synthesis of CMP-Neu5Ac and ST reaction. For the STs, α2,3ST from *Pasteurella multocida* (α2,3PST) (Yu et al. 2005) and α2,6ST from *Photobacterium damselae* (α2,6PdST) (Cheng et al. 2010; Sun et al. 2008), which belong to GT family 80 (Yamamoto et al. 2008) with inverting mechanism of GT-B fold, were selected.

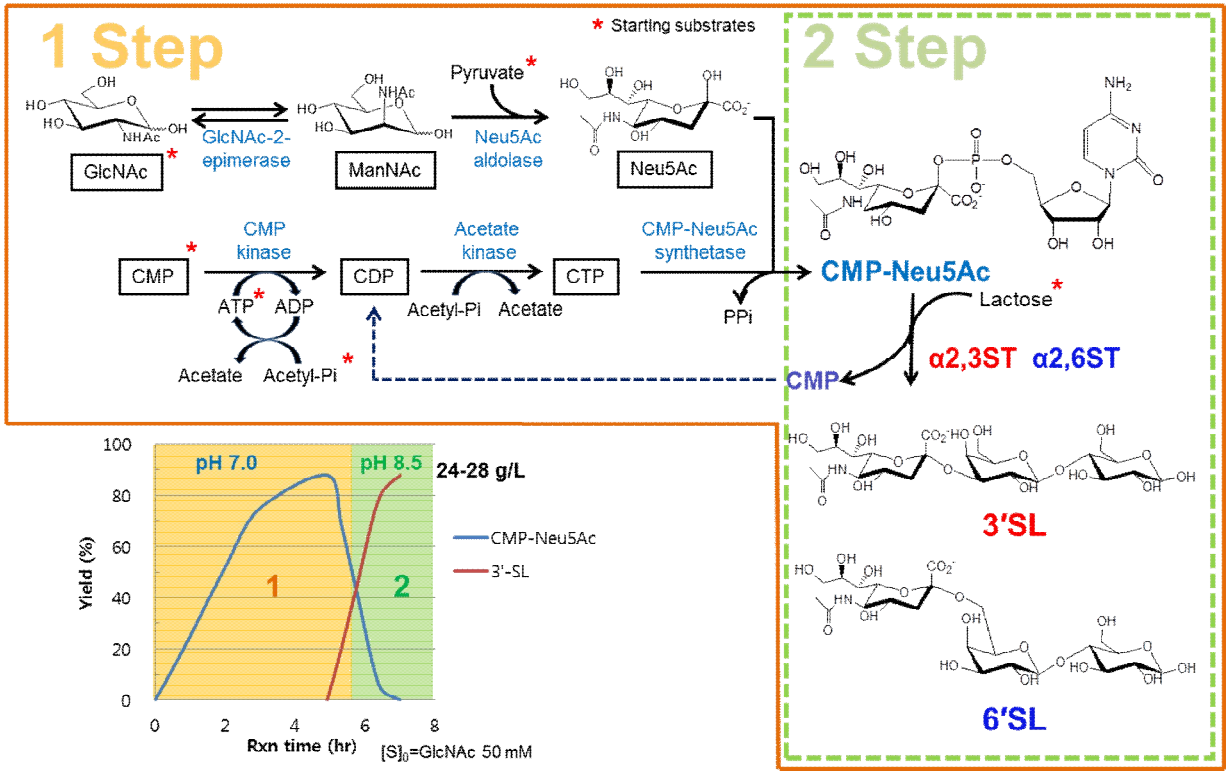


Figure 1.7 Enzymatic *in vitro* synthesis of 3′-SL and 6′-SL via 2-step reactions

Five enzymes are used to produce CMP-Neu5Ac from GlcNAc, CMP, ATP, pyruvate, and acetyl-phosphate. After synthesis of CMP-Neu5Ac, pH is sequentially adjusted to 8.5 for STs reaction to produce SLs

Each enzyme has its own advantages and disadvantages for the enzymatic synthesis of 3′-SL and 6′-SL using the STs. In general, α2,3PST has shown ca. 6 times higher specific activity than α2,6PdST, but it has pH-dependent multifunctional substrate regio-specificity, so that 6′-SL is produced as a by-product at below pH 7.0 (Yu et al. 2005). Since the optimum pH of the α2,3PST reaction is 8.5, the operation of the α2,3PST reaction under this condition can greatly reduce the generation of 6′-SL by-product (Sugiarto et al. ; Yu et al. 2005). However, since the optimum pH of multi-step enzyme reactions for CMP-Neu5Ac synthesis is 7.0, combining the two steps of the synthesis of CMP-Neu5Ac and ST reaction at around pH 7.0 is more desirable to make α2,3PST with reduced α2,6ST activity. In the case of α2,6PdST, its natural low activity hampers the large-scale mass production of 6′-SL (Sun et al. 2008), but no additional by-products caused by poor regio-selectivity are observed, unlike with α2,3PST.

1.3.2.2 Synthesis of GDP-fucose

FLs are produced from GDP-Fuc, which could be synthesized via salvage pathway from L-fucose or *de novo* pathway from glucose (Han et al. 2012). In the salvage pathway, after extracellular L-fucose is transferred into the cells, L-fucose is phosphorylated by L-fucokinase and converted into GDP-Fuc by L-fucose-1-phosphate guanylyltransferase using guanosine 5′-triphosphate (GTP) substrate. In *de novo* pathway, fructose-6-phosphate is metabolized into mannose-1-phosphate by mannose-6-phosphate isomerase (ManA) and phospho mannomutase (ManB). Mannose-1-phosphate is converted to GDP-mannose (GDP-Man) via mannose-1-phosphate guanylyltransferase (ManC) combined with GTP. And then, GDP-Fuc is

produced by GDP-Man 4,6-dehydratase (Gmd) and GDP-Fuc synthase (WcaG) through the GDP-4-keto-6-deoxymannose intermediate from GDP-Man (Figure 1.8). In the pathway, 1 mol of glucose is produced into 1 mol of GDP-Fuc at the expense of 1 mol of ATP, GTP, and NADPH.

For *in vivo* production of GDP-Fuc, *Saccharomyces cerevisiae* and *E. coli* cells were used. Mattila group constructed *S. cerevisiae* strain expressing *E. coli gmd* and *wcaG* genes to convert GDP-Man to GDP-Fuc. The strategy has an advantage of rich inherent cytosolic GDP-Man pool in yeast. As a result, 0.2 mg/L of GDP-Fuc was produced without addition of external GDP-Man (Mattila et al. 2000). However, since the GDP-Fuc yield was very low in the yeast system, *E. coli* harboring endogenous *manA*, *manB*, *manC* was used for GDP-Fuc production. Glucose-limited fed-batch fermentation of *E. coli* BL21 (DE3) expressing the *man B*, *manC*, *gmd* and *wcaG* genes was conducted, resulting in 170 mg/L of GDP-Fuc production (Lee et al. 2009). To improve the GDP-Fuc production by *de novo* pathway, guanosine nucleotides pathway was manipulated for increase in GTP that is used for ManC reaction (Figure 1.8). Using the fed-batch reaction, the overexpression of inosine 5′-monophosphate dehydrogenase (IMP), guanosine 5′-monophosphate synthetase (GuaB and GuaA), GMP reductase (GuaC), and guanosine-inosine kinase (Gsk) were investigated. Consequently, overexpression of Gsk showed the highest GDP-Fuc production, yielding 305 mg/L, which could be due to the increase of the intracellular level of GMP (Lee et al. 2012b).

GTP is the key substrate for the biosynthesis of GDP-Fuc in *de novo* pathway as well as in salvage pathway as illustrated in Figure 1.8. Since the FKP use GTP as substrate for fucose-1-phosphate guanylyltransferase reaction to convert

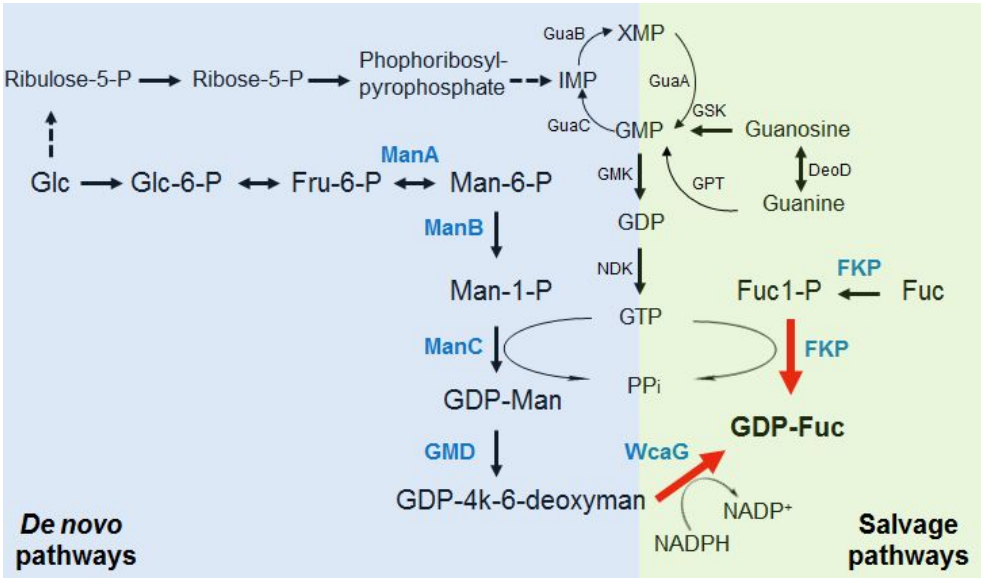


Figure 1.8 Enzymatic synthesis of GDP-Fuc by *de novo* and salvage pathway

For the production of FLs, donor substrate GDP-Fuc could be produced by *de novo* and salvage pathways. The enzymes are abbreviated as follows: ManA, mannose-6-phosphate isomerase; ManB, phosphomannomutase; ManC, mannose-1-phosphate guanylyltransferase; GMD, GDP-Man 4,6-dehydratase; WcaG, GDP-Fuc synthase; Gsk, guanosine-inosine kinase; DeoD, purine nucleotide phosphorylase; Gpt, guanine phosphorisyltransferase; Gmk, guanylate kinase; Ndk, nucleoside diphosphate kinase; GuaB and GuaA, guanine 5′-monophosphate (GMP) synthetase; IMP, inosine 5′-monophosphate dehydrogenase; FKP, bifunctional fucokinase/fucose-1-phosphate guanylyltransferase

L-fucose 1-phosphate into GDP-Fuc. In *E. coli*, GTP is produced from guanosine 5′-diphosphate (GDP) by nucleoside diphosphate kinase (Ndk, EC 2.7.4.6), and GDP is generated from guanosine 5′-monophosphate (GMP) via guanylate kinase (Gmk EC 2.7.4.8). There are two pathways to synthesize GMP, the *de novo* and salvage pathway. In *de novo* pathway, GMP is generated from ribulose 5-phosphate via serial steps and phosphoribosylpyrophosphate (PRPP) as an intermediate. On the other hand, in salvage pathway, GMP is generated from guanine catalyzed by guanine phosphoribosyltransferase (Gpt, EC 2.4.2.22) or from guanosine via Gsk (EC 2.7.1.73) and purine nucleoside phosphorylase (DeoD, EC 2.4.2.1). Recently, Zhai et al. constructed metabolic engineered *E. coli* stain for GDP-Fuc production by overexpressing the enzymes in salvage pathway considering the increase in GTP (Zhai et al. 2015). The shake flask fermentation could obtain the maximum GDP-Fuc content of 4.2 mg/g CDW via expression of FKP, Gpt, Gmk, and Ndk.

1.3.2.3 Synthesis of FLs

FL had been obtained via extraction from human milk (Anderson and Donald 1981) or chemical synthesis (MARTIN-LOMAS 1986), but the limited availability of human milk or the necessity of complex protection and deprotection of side group in chemical synthesis confined sufficient supply and cost efficiency. Therefore, enzymatic synthesis of FLs has been considered to be the most plausible means of the synthesis of FLs. For the production of FLs, although several FucTs have been cloned, identified and characterized from a mammalian or bacterial origin, most FucTs suffer from low expression in *E. coli* that prevents their use for the large scale synthesis of FLs. The low solubility of the α1,2-FucT and α1,3-FucT

prevents us from acquiring the crystal structure of FucTs and performing an efficient enzyme screening for protein engineering. Several attempts have been made to increase the solubility and stability of α1,2-FucT and α1,3-FucT, respectively.

In the case of FutC, which is α1,2-FucT from *H. Pylori* 26695, had been cloned by Wang et al. (1999). To inhibit the frame shift of FutC, 14 cytosine was replaced with 12 nucleotide encoding four proline (Drouillard et al. 2006). In addition, they tried to increase the solubility of FutC in *E. coli* by using strong promoter like tac or T7 and generating fusion protein with thioredoxine and glutathione S-transferase (GST). However, fusion protein linked FutC decreased the solubility by 70%. Since FutC has relatively high activity for the production of 2′-FL, the enzyme has been widely used for synthesis of 2′-FL although the solubility still remained at al low level. The *E. coli* strain (Col1F'ΔlacZ, DE3) harboring pET21a*futC* and pLNTR*rcsA* genes was used for 2′-FL production with *de novo* synthesis of GDP-Fuc. At high cell density culture with 15 g/L of lactose accomplished the 2′-FL yield for 11 g/L after 45 h (Drouillard et al. 2006). Using the same FutC, Baumgartner et al. (2013) utilized both *de novo* and salvage pathway from lactose and glycerol. To construct the strain harboring the two pathways, recombinant genes of the *de novo* (*ManB, ManC, Gmd, WcaG*) and salvage pathway (*fkp*) were integrated into the chromosome of *E. coli* JM109. It was shown that the strain carrying additional copies of the *futC* and *fkp* improved the 2′-FL production, resulting in 20 g/L (10 g/L in media and cell pellet, respectively) after 35.5 h using the fed-batch fermentation in 13.5 L scale. The strategy allows antibiotic-free fed-batch fermentation with improved production yield (Baumgartner et al. 2013). Recently, 2′-FL has been produced in engineered *E. coli*

BL21 star (DE3) by modulation of lactose metabolism and FutC (Chin et al. 2015). Chin et al. deleted whole endogenous lactose operon and introduced modified lactose operone lacZΔ15 to alleviate the metabolism of lactose and divert the lactose flux to the 2′-FL production. In addtion, three aspartate-fused FutC at its N-terminal was utilized to improve the protein solubility, which resulted in production of 6.4 g/L after 78 h by fed-batch culture (1 L) using glycerol as carbon source.

Like the FutC, most FutAs, α1,3-FucT from *H. pylori*, suffer from the low solubility in *E. coli*. Several trials have been conducted to increase the solubility and stability of α1,3-FucTs by deleting the α-helices containing hydrophobic, positively charged residues and partial heptad repeat region in FutAs (i.e., 8-10 repeats) from *H. pylori* NCTC 11639 (Lin et al. 2006) and UA948 (Ma et al. 2006). Despite the efforts to improve the solubility, the fraction of soluble protein of α1,3-FucTs has remained at a low level from 4.4 to 15 mg/L. In addition, to synthesize 3-FL, the highly stringent substrate specificity of α1,3-FucTs toward LacNAc (Galβ1,4GlcNAc) over lactose (Lin et al. 2006) hampers the production of 3-FL. Due to the bottleneck explained above, there has been little progress in 3-FL production, except for the construction of an *E. coli* strain for 3-FL synthesis by Hufner et al. (2010). They constructed the strain *E. coli* BW25113 (DE3) knocked out of the gene fucA, which encodes the fuculose-1-phosphate aldolase for the conversion of fuculoase-1-phosphate to L-lactaldehyde and dihyroxyacetone phosphate. The two plasmid carrying fkp, fucP (fucose transporter), and codon optimized futA were expressed in the strain, resulted in production of 3-FL in *E. coli*.

1.4 Strategies for protein engineering

1.4.1 Directed evolution and rational design of enzymes

Traditionally, directed evolution is used to describe entire range of molecular biology techniques that allow natural evolutionary process to be mimicked *in vitro* (Dalby 2011). Directed evolution can be used when the molecular basis for the property or detailed strutural information is inefficient for a target protein. The traditional directed evolution generally involves accumulation of mutations by iterative protocols such as random mutagenesis and *in vitro* recombination (Lutz 2010). As a *in vitro* random mutagenesis strategies, error-prone PCR (epPCR) (Chen and Arnold 1993), DNA shuffling (Stemmer 1994), and staggered extension process (StEP) (Zhao et al. 1998) have been used to introduce random point mutation in a population of DNA product and allow random recombination. Then mutants should be screened by high-throughput screening for the desired property of interest. Directed evolution shows high probability of success to find a valuable mutant since the beneficial mutations displaying improved phenotype could be accumulated by several rounds of generation. However, the approach can be problematic since the several rounds of evolution generate a high number of variants to be screened, which is time and labor consuming for screening assays (Lutz 2010; Steiner and Schwab 2012).

On the ohter hand, rational design usually focuses on the active site/substrate binding site or domain interfaces or hinge regions using the detailed structural information. The rapidly growing number of protein structure in the Protein Data Bank (PDB) and advances in homology modeling provide valuable clues for protein

engineering to identify the residues responsible for catalytic property and interactions with substrate or neighboring residue/domain. In pure rational design, mutations are introduced by site-specific mutagenesis using the detailed structural informations. Although site-specific mutagenesis is highly economical in screening capacity due to a significant reduction of the library size, the approach has relatively low possibility of success to generate a disired mutant. The rational design largely depends on 3D structure information and sufficient knowledge of the contribution of residues to enzyme properties, especially for detailed understandings about the binding pockets and active site region such as 3D structure with a bound transition state analogue (Bornscheuer 2013). The limitation of pure rational design can be arised from the incomplete understanding of flexibility of the protein and conformational changes upon substrate binding, sensitivity of the enzyme reaction to small changes in distances and geometry of the substrate, and adverse effect of some mutations on the expression and stability of the protein (Guo et al. 2004; Steiner and Schwab 2012).

1.4.2 Focused directed mutagenesis

Semi-rational design combines advantages of rational and random design to creat smaller but smarter libraries (Figure 1.9). The approach utilizes information of protein sequence, structure and function, as well as computational predictive algorithms to preselect promising target sites and limited amino acid diversity for protein engineering (Lutz 2010). The focus on specific region of protein can reduce library sizes while the consideration of evolutionary variability and 3D structural information, resulting in creation of libraries with high functional content. In the strategy, selection of fuctional residues responsible for the substrate binding,

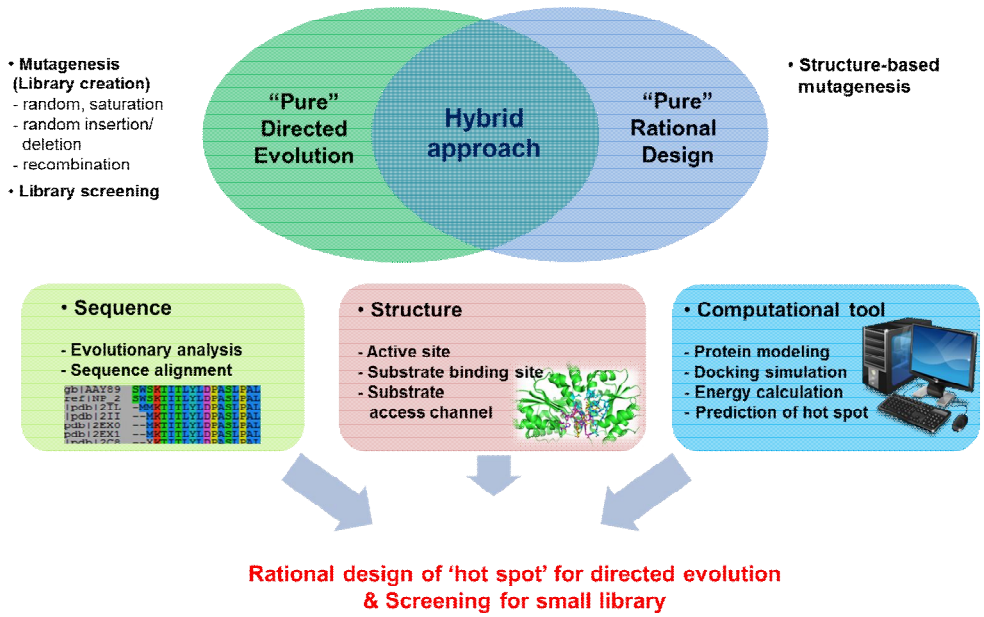


Figure 1.9 Protein engineering using hybrid approach

transition-state stabilization or producte release, is important for subsequent directed evolution. Targeted funtional residues are applied toward the directed evolution.

1.4.2.1 Selection of functional residues

To select the functional residues of the target region for directed evolution, structure and sequence-based design can be used combined with computational tools.

For sequence-based enzyme design, multiple sequence alignment and phylogenetic analyses have been used for the exploration of amino acid conservation and ancestral relationships among groups of homologous protein. These analyses enable protein engineer to identify functional hot spots by verifying evolutionary conservation (Lutz 2010).

For structure-based enzyme design, structural information can provide candidate residues located in target region such as active site, substrate binding site or domain interfaces or hinge regions. A small number of residues were targeted for saturation mutagenesis by examining residues within a given distance around the bound substrate/product in the crystal structure of P450 monooxygenase (Hoffmann et al. 2011) and nucleotidyltransferase (Moretti et al. 2011). In addition, structure modeling and docking simulation of substrate/analogue as well as crystal structure offer useful information to targeted region for protein engineering. Structure guided computational tool, for instance CAVER, B-factor, and SCHEMA have been developed to increase the enzyme's properties. The CAVER is a software tool used for the identification and characterization of tunnels dynamics in trajectories obtained by molecular dynamics simulations (Chovancova et al. 2012). The CAVER allows investigation of changes in the accessibility of the enzyme's tunnel and thereby on

the overall catalytic properties by mutation in tunnel (Biedermannova et al. 2012). For instance, Kong et al. (2014) eliminated the steric hindrance near the potential product-release site using the CAVER to improve the catalytic efficiency of epoxide hydrolase for bulky substrates.

B-factor in protein structure reflects the fluctuation of an atom about its average position (Yuan et al. 2005). Since the distribution of B-factors is regarded as flexibility and dynamics, the residues with a high B-factor were targeted for increasing of thermostability of enzymes (Blum et al. 2012; Floor et al. 2014). SCHEMA is enable to recombine multiple low identity proteins to obtain chimeric enzyme with improved thermostability (Heinzelman et al. 2010; Voigt et al. 2002).

In addition, sequence and structure database can be integrated for the identification of mutional range and hot spots. Among the large number of data analysis, HotSpot Wizard as a computational tool combines structure with sequence database for evolutionary analysis. The web server estimates the mutability of amino acids in the active site or lining the access channel (Pavelka et al. 2009) using the databases containing PDB, NCBI, BLAST, and CAVER etc. Pavlova et al. (2009) conducted molecular dynamics and HotSpot Wizard to estimate the mutability of hot spot residues located in access tunnel of *Rhodococcus rhodochrous* haloalkane dehalogenase. As a result, haloalkane dehalogenase mutant with up to 32-fold higher activity than wild-type toward 1,2,3-trichloropropane was screened from the libraries generated by saturation mutagenesis.

In addition, focused directed evolution could involve ‘neutral drifts’ that maintain the protein’s original function and structure by excluding deleterous mutations (Gupta and Tawfik 2008). With regard to neutral drift, small libraries for

directed evolution could be obtained, resulted in generation of properly folded and funtional variants with improved activity.

1.4.2.2 Directed evolution of functional residues

Targeted funtional residues are applied toward the directed evolution such as site-saturation mutagenesis (SSM), iterative saturation mutagenesis (ISM), and combinatorial active-site saturation test (CASTing). SSM allows the substitution of single amino acid against all 20 possible amino acids at once. To reduce the redundancy in the genetic code and frequency of premature stop codons, degenerate codons have been designed to have NNK/S codon or combinations thereof (Acevedo-Rocha et al. 2015; Kille et al. 2013).

ISM indicates iterative cycles of saturation mutagenesis at rationally chosen sites in an enzyme to improve catalytic properties such as catalytic activity, enantioselectivity, susbtrate specificity, or thermostability (Reetz and Carballeira 2007). Each of several sites is randomized by saturation mutagenesis to generate each different library, and then best hit in each library is identified and sequenced. Then the hit mutant is used as the template for another rounds of saturation mutagenesis. Such a convergent process maximizes the possibility of additive and/or cooperative effects of newly introduced mutation. ISM can be combined with CASTing, which was develped for creation of small libraries by simulataneosly randomizing a group consited of two or three amino acids based on 3D structure information. In this case, substrate binding pocket is targeted for mutagenesis. Using the iterative CASTing, enantioselectivity of epoxide hydrolase (Reetz et al. 2006b),

lipase (Reetz et al. 2010) and thermostability of lipase (Reetz et al. 2006a) were improved substantially.

1.4.3 Protein engineering of GTs

The random mutageneses with high-throughput screening have been carried out to increase the activity or to broaden the substrate specificity for α2,3-sialyltransferase (Aharoni et al. 2006), β1,3-galactosyltransferase (Yang et al. 2010), and oleandomycin GT (Williams et al. 2007), respectively. Otherwise, domain swapping based on the sequence information was carried out for the GT-B enzymes to expand the substrate specificity (Krauth et al. 2009; Park et al. 2009). When compared to directed evolution, the successful rational structure-guided mutageneses of GTs have been relatively few to increase the activity or to expand the substrate specificity.

1.5 Research objectives

The aim of this work is to produce high value added HMO, in particular SL and FL by enzymatic synthesis using glycosyltransferase. To increase the yield and productivity for the production of SLs and FLs, protein engineering was conducted to ST and FT.

At first, to meet the demands in the changes of ST properties, a hybrid approach combining alanine scanning based on a protein structure model with site-saturated mutagenesis was attempted. The hybrid approach called “semi-rational design” was developed to generate a smaller but higher-quality library to investigate a rather narrowly focused region by utilizing protein sequences,

structure information, and computational modeling (Chica et al. 2005; Lutz). Damborsky et al. carried out a small-library site-saturation mutagenesis near the substrate access tunnels in haloalkane dehalogenase, showing that the mutations restricting solvent water accessibility to the hydrophobic substrate-protein complex can increase its k_{cat}/K_m value by three times (Banas et al. 2006). Reetz et al. also focused on the interface of the FAD and NADP-binding domains of a monooxygenase, identified two amino acid residues as functional sites, and conducted saturation mutagenesis (Wu et al. 2010a).

We focused on the residues contacting with CMP-Neu5Ac and lactose within a distance of 5 Å from the boundary surface of the substrate molecules in the substrate binding pocket of each ST, and selected functional residues based on multiple sequence alignments and alanine scanning. Considering the conserved and functional residues deduced from multiple sequence alignment analysis (Prabhu et al. 2010; Yeom et al. 2008) and the neutral drift of some mutations, saturation mutagenesis was performed on the selected functional residues. Using this hybrid approach, herein we demonstrated that the ST activities of α2,3PST and α2,6PdST were increased by two-fold and five-fold, respectively, and in the case of α2,3PST, the 6′-SL by-product synthesis activity was greatly reduced at low pH (<pH 7.0). In addition, detailed characterizations of the screened mutants were performed, and the mutation effects on the activity of the STs were investigated at molecular levels based on a computer model. This study is the first report thus far showing that ST activity toward CMP-Neu5Ac and lactose can be improved by protein engineering for the synthesis of 3′-SL and 6′-SL. The mutants of α2,3PST and α2,6PdST can be

efficiently applied to one-pot reactions for the large-scale synthesis of 3′-SL and 6′-SL.

For the production of FLs, low solubility of FucTs in *E. coli* and highly stringent substrate specificity of α1,3-FucTs toward LacNAc (Galβ1,4GlcNAc) over lactose (Lin et al. 2006) in the case of α1,3-FucT, hampers the production of 2′-FL and 3-FL, and the bottlenecks highlights the need for protein engineering. Especially, there has been little progress in 3-FL production, relative to that for 2′-FL production (Baumgartner et al. 2013; Drouillard et al. 2006; Lee et al. 2012a). In addition, there have been no enzyme engineering studies of FucTs to improve the activity and/or expand the substrate specificity even though several glycosyltransferases (GTs) have been subjected to protein engineering.

In this study, FutA (α1,3-FucT, *futA* from *H. pylori* 26695), which is known to have a very low relative substrate specificity toward lactose over LacNAc (Dumon et al. 2004), was selected to synthesize 3-FL. Prior to protein engineering, codon optimization was followed by systemic truncation of the C-terminus to increase the expression level of the soluble enzyme. To further increase the specific activity of the truncated FutA for the lactose substrate, focused directed evolution as a semi-rational mutagenesis was performed by combining the structure-guided analysis of active site/substrate binding site and site-saturation mutagenesis (SSM), which generates a much smaller but higher-quality library (Davids et al. 2013; Goldsmith and Tawfik 2012). To select the functional residues of the target region for directed evolution, computational tools can be used for sequence-based (e.g., multiple sequence alignment and phylogenetic analysis) and structure-based (e.g., B-factors, CAVER, and SCHEMA) design (Davids et al. 2013; Lutz 2010). This study uses a

docking simulation and HotSpot Wizard as a computational tool for evolutionary analysis that is based on a structure and sequence database, to generate the mutability map (Pavelka et al. 2009). The functional residues were selected and were subsequently subjected to SSM followed by iterative saturation mutagenesis (ISM) (Reetz and Carballeira 2007). To reduce the screening effort of ISM, the target residues were clustered according to the structure information in order to select any efficient pathway.

As a result, a mutant was generated with quadruple mutations, and the effects that the mutations had on the α1,3-FucT activity were analyzed at the molecular level by using a computer model. The mutant of Δ52 FutA remarkably improved the yield and productivity of the *in vitro* production of 3-FL. Thus, this study is the first to thus far report that FucT's specific activity and substrate specificity for poor acceptor substrate can be significantly improved by conducting protein engineering.

The Δ52 FutA mutant could be efficiently applied to *in vivo* synthesis of 3-FL in *E. coli* system using salvage pathway. *E. coli* strain was constructed to be knocked out of the genes responsible for degradation of substrates, lactose and L-fucose. To increase the 3-FL production, media and culture condition were optimized. In addition, it was revealed that GDP-Fuc production is a rate limiting step in metabolic pathway for *in vivo* production of 3-FL. In order to solve the rate limiting step and increase the yield and productivity, several constructions of the multi-genes were overexpressed in *E. coli* BL21 (DE3) strain and fed-batch culture using glycerol was carried out.

Chapter 2.

Materials and methods

2.1 General method

2.1.1 Chemicals and materials

CMP-Neu5Ac sodium salt, GDP-L-Fuc disodium salt, 3′-SL, and 6′-SL were obtained from GeneChem Inc. (Daejeon, Korea). All of the other chemicals including 2′-FL, 3-FL and Le^x were purchased from Sigma-Aldrich, Inc. (St. Louis, MO) and Junsei Chemical Co. Ltd. (Tokyo, Japan).

Restriction enzymes, T_4 DNA ligase, DNA polymerase, and other DNA modifying enzymes were purchased from Thermo Fermentas (MA, USA), New England Biolabs (MA, USA), Takara (Takara bio, Shiga, Japan), Solgent (Korea), and Promega (Madison, WI, USA), and used as recommended by the manufacturers.

Escherichia coli DH5α, BW25113 (DE3) were used as hosts for expression of STs, FucTs, and FKP. For the culture of cells above, Difco Luria-Bertani (LB) (1% tryptone, 0.5% yeast extract, 1% sodium chloride) and Difco Terrific broth (TB) (1.2% pancreatic digest of casein, 2.4% yeast extract, 0.94% dipotassium phosphate, 0.22% monopotassium phosphate, and 0.4% glycerol) were used (Becton, Dickinson and company, USA).

2.1.2 Preparation of enzymes

2.1.2.1 Sialyltransferases (STs)

α2,3PST truncated with 24 amino acids at the N-terminus (Yu et al. 2005) was cloned in pET23a, and α2,6PdST truncated with 15 amino acids at the N-terminus

(Sun et al. 2008) was cloned in both pET15b and pET28a. The truncated α2,3PST and α2,6PdST were used as wild-type. Each plasmid was transformed into *E. coli* BW25113 (DE3), and the cloned ST was induced with 0.5 mM IPTG, when OD600 of the culture reached 0.5-0.8. The induced cells were further incubated in a rotary shaker at 200 rpm and 30 ºC for 5 h in the case of α2,3PST, and 18 ºC for 18 h in the case of α2,6PdST. The cells were harvested from 50 mL of the culture broth by centrifugation at 3,390 × g for 10 min at 4 ºC, suspended in 5 mL of 10 mM potassium phosphate buffer (pH 7.0) containing 2 mM EDTA, 1 mM phenylmethanesulfonylfluoride (PMSF), and 0.01% 2-mercaptoethanol, and lysed by sonication on ice. Cell lysate (soluble fraction) was obtained by centrifugation at 22,250 × g for 30 min at 4 ºC.

2.1.2.2 Fucosyltransferases (FucTs)

α1,2-FucT (FutC; GenBank accession number HP0094, (Drouillard et al. 2006)) and α1,3-FucT (FutA; GenBank accession number AAD07447.1, (Dumon et al. 2004)) were acquired from *H. pylori* 26695 based on carbohydrate-active enzymes database (CAZy) database (Lombard et al. 2014). In the case of FutC, 14C tract, which cause a translational frame shift, was replaced to 12C to make an active form of α1,2-FucT (Park 2012). *futC* cloned in pET24ma using BamHI and SalI was transformed into *E. coli* BW25113 (DE3), and the cells were incubated in LB broth with 50 $\mu g \cdot mL^{-1}$ of Kanamycin. And then the cells were induced with 0.5 mM IPTG, when OD_{600} of the culture reached 0.5-0.8. The induced cells were further incubated in a rotary shaker at 200 rpm and 18 ºC for 20 h. The cells were harvested from 50 mL of the culture broth by centrifugation at 3,390 × g for 10 min

at 4 ºC, suspended in 5 mL of 20 mM sodium phosphate buffer (pH 7.6), and lysed by sonication in ice. The supernatant of the cell lysate (soluble fraction) was obtained by centrifugation at 22,250 × g for 30 min at 4 ºC.

futA, α1,3-FucT from *H. pylori* 26695 was amplified using PCR as follows : initial denaturation at 95 ºC for 5 min, followed by 30 cycles of 95 ºC for 1 min, 55 ºC for 1 min, 72 ºC for 1 min 10 sec, and 72 ºC for 7 min for a final extension using 1X PCR buffer, 0.2 mM dNTP, 2.5 mM $MgCl_2$, and 100 pmol of primers (Forward: 5′-GACCATATGTTCCAACCCCTATTAG-3′/ Reverse: 5′-CCA CTCGAGCAACCCAATT-3′), 1X band doctor, 1 μL of genomic DNA of *H. pylori* 26695 and 2.5 U of Pfu DNA polymerase (Solgent, Korea). The PCR product was digested with NdeI/XhoI and ligated into the vector pET24ma using 5 U of T4 DNA ligase. The plasmid harboring *futA* was transformed into *E. coli* BW25113 (DE3), and the cells were incubated in LB broth with 50 $\mu g \cdot mL^{-1}$ of Kanamycin at 37 ºC, followed by induced with 0.05 mM IPTG, when OD_{600} of the culture reached 0.5-0.8. The induced cells were further incubated in a rotary shaker at 200 rpm and 18 ºC for 20 h. The cells were harvested and FutA was prepared as the same method mentioned above.

2.1.2.3 L-Fucokinase/GDP-fucose pyrophosphorylase (FKP)

L-Fucokinase/GDP-fucose pyrophosphorylase (FKP) from *Bacteroides fragilis* 9343 had been cloned into pET16b, and the plasmid was kindly obtained by Dr. Peng Wu (Albert Einstein College of Medicine, USA) (Wang et al. 2009). BW25113 (DE3) cells harboring the FKP-encoding plasmid were incubated in LB broth with Ampicillin (100 $\mu g \cdot mL^{-1}$) for shaking at 37 ºC until OD_{600} =0.9-1.0, at

which time FKP expression was induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). The temperature was lowered to 25 ºC. After 16 h, cells were harvested and resuspended in 20 mM of sodium phosphate buffer (pH 7.6) and lysed by sonication. The supernatant of the cell lysate (soluble fraction) was obtained by centrifugation at 22,250 × g for 30 min at 4 ºC.

2.1.3 *in vitro* assay and quantitative analysis

2.1.3.1 Sialyltransferases (STs) assay

ST activity assays were performed in a total volume of 100 μL of 50 mM Tris-HCl buffer at pH 8.5 for α2,3PST (Yu et al. 2005), and at pH 8.0 (Yamamoto et al. 2007) for α2,6PdST. The reaction mixtures contained 10 mM CMP-Neu5Ac, 5 mM lactose, and appropriate amounts of a specified enzyme. The reactions took place for 1 h for α2,3PST and 2 h for α2,6PdST at room temperature. The reactions were stopped by heating at 95 ºC for 40 sec, after which the reaction mixture was centrifuged, and its supernatant was analyzed quantitatively using a Bio-LC DX-300 (Dionex Co., Sunnyvale, CA). The supernatant of the reaction mixture was diluted with distilled water, and injected into a Bio-LC DX-300 HPAEC with a pulsed amperometric detector (Dionex Co., Sunnyvale, CA). To separate the lactose, 3′-SL, and 6′-SL in the reaction mixture using a CarboPac PA-100 column (Dionex Co., Sunnyvale, CA), 100 mM NaOH solution was used at a flow rate of 1 mL/min. ST activities were also quantitatively determined by a colorimetric method using cresol-red (pKa=8.32) for α2,3PST (Park et al. 2009) and phenol-red (pKa=7.5) for α2,6PdST as pH indicators. The method is based upon the color and

absorbance changes of the pH indicators according to the proton release during the glycosidic bond formation of the products (Figure 2.1A). The ST activity assays were conducted in a 96-well plate with 100 μL of 5 mM Tris-HCl buffer solution (pH 8.5 for α2,3PST, and pH 8.0 for α2,6PdST) containing 4 mM CMP-Neu5Ac, 4 mM lactose, 0.2 mM pH indicator, and a specified enzyme. The increase in absorbance at 405 nm for cresol-red and the decrease in absorbance at 560 nm for phenol-red were recorded for each sample at 30 sec or 1-min intervals for 5-10 min using a UV spectrophotometer (Spectrostar nano, BMG LABTECH, Germany) (Figure 2.1B). The reactions were started by adding the substrate mixtures to the enzyme reaction mixtures, and then the initial reaction rate was calculated by measuring the changes in absorbance in the reaction mixture. The control for each sample containing the same reaction mixture without lactose substrate was used, and ΔOD was calculated between the control and reaction samples. The released proton concentration during the reaction was calculated based on an HCl calibration curve. To generate a calibration curve with either a donor or acceptor substrate, 4 mM CMP-Neu5Ac or 4 mM lactose, with 0.2 mM pH indicator and appropriate amounts of enzyme (10% v/v) were mixed into the same 100 μL of 5 mM Tris-HCl reaction buffer containing 0-1.2 mM HCl. The enzyme activity was defined as the amount of enzyme required to produce 1 μmol of proton released during glycosidic bond formation per minute at room temperature. Correlation coefficient between the colorimetric and Bio-LC assays was 0.95 (Figure 2.2).

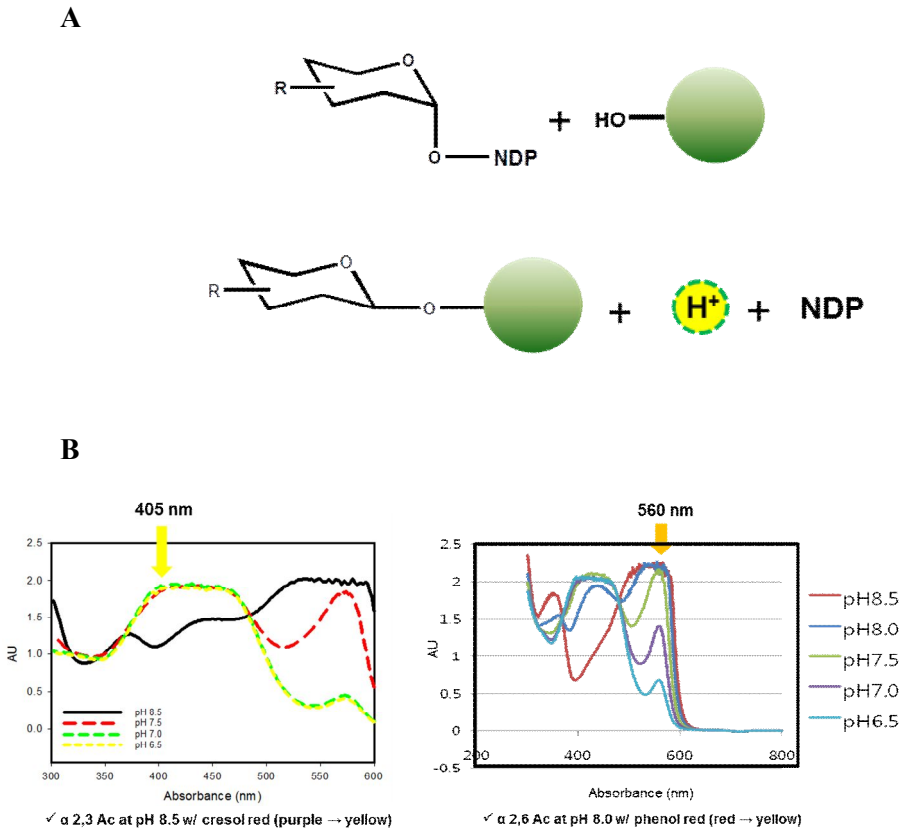


Figure 2.1 ST assay using a colorimetric method

ST assays were conducted by colorimetric method according to the release of proton during the glycosidic bond formation (A). α2,3PST and α2,6PdST assays were carried out at pH 8.5 and pH 8.0, respectively (B).

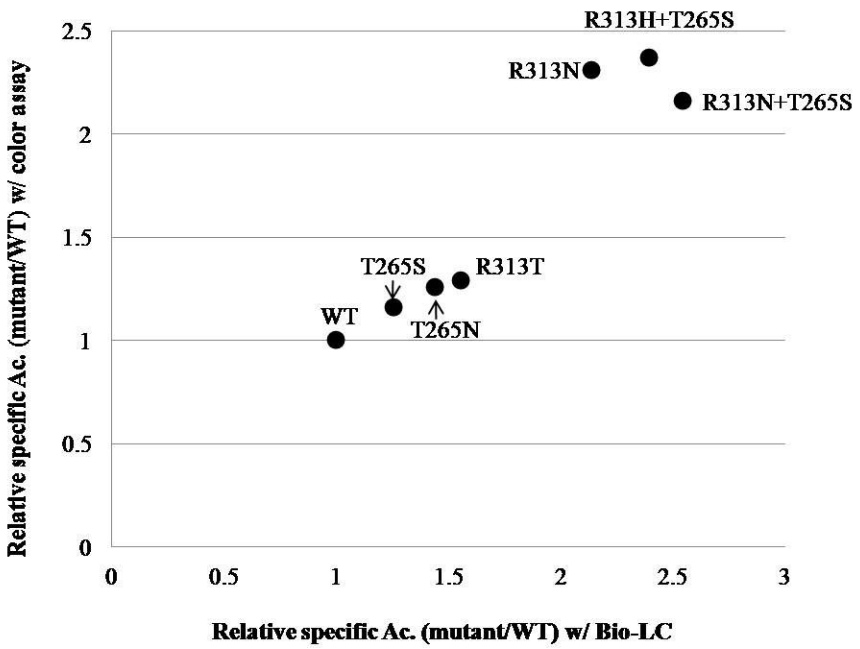


Figure 2.2 Correlation coefficient of relative specific activity between Bio-LC and colorimetric assay

Relative specific activities (mutant/WT) of α2,3PST mutants were compared between the Bio-LC and colorimetric assay with the purified enzyme. Bio-LC method detected the produced 3′-SL from the reaction with 5 mM CMP-Neu5Ac, 2.5 mM Lac, and 0.1 μg of enzyme in 50 mM Tris-HCl buffer pH 8.5. Specific activities of α2,3PST WT were determined as 24.75 U/mg and 22.71 U/mg by Bio-LC and colorimetric method, respectively. Correlation coefficient between the two methods was also determined as 0.95.

2.1.3.2 Fucosyltransferase (FucT) assay

FucT activity assays were performed in 50 mM sodium phosphate buffer (pH 7.6) containing 5 mM GDP-L-Fuc, 15 mM lactose (or 5 mM LacNAc), 5 mM $MgCl_2$, and appropriate amounts of a specified enzyme. The reactions were carried out for 3 h at 37 ºC, and stopped by heating at 95 ºC for 40 sec. The reaction mixture was then centrifuged at 22,250 × g for 10 min, and its supernatant was analyzed quantitatively using a Bio-LC DX-300 (Dionex Co., Sunnyvale, CA). The supernatant of the reaction mixture was diluted with distilled water, and injected into a Bio-LC DX-300 HPAEC with a pulsed amperometric detector (Dionex Co., Sunnyvale, CA). To separate the 2′-FL, 3-FL, and lactose in the reaction mixture using a CarboPac PA-1 column (Dionex Co., Sunnyvale, CA), the column was equilibrated with 100 mM NaOH for 7 min, and eluted with linear gradient from 0-90 mM of sodium acetate in 100 mM NaOH for 7-24 min at a flow rate of 0.7 mL/min. Meanwhile, LacNAc and Le^x were separated using CarboPac PA-100 column (Dionex Co., Sunnyvale, CA) with linear gradient from 60-100 mM NaOH and 50-100 mM sodium acetate for 18 min at a flow rate of 0.7 mL/min.

One-pot reactions with GDP-Fuc synthesis using FKP were conducted in 50 mM sodium phosphate buffer (pH 7.6) containing 5 mM L-Fuc, 5 mM GTP, 5 mM ATP, 5 mM $MnSO_4$, 5 mM $MgCl_2$, 15 mM lactose, FKP (10% v/v), and appropriate amounts of a specified FucT enzyme. The reactions were carried out at 37 ºC, and stopped by heating at 95 ºC for 40 sec. The reaction mixtures were treated as mentioned above, and analyzed by Bio-LC DX-300 for quantification of produced 2′-FL and 3-FL.

2.1.3.3 FKP assay and quantification of GDP-Fuc

For the *in vitro* synthesis of GDP-Fuc, FKP (10% v/v) was used with 5 mM L-Fuc, 5 mM GTP, 5 mM ATP, and 5 mM $MnSO_4$ in 50 mM sodium phosphate buffer (pH 7.6) at 37 ºC. The reactions was stopped by heating at 95 ºC for 40 sec, and centrifuged at 22,250 × g for 10 min. After centrifugation, supernatant was analyzed quantitatively using a HPLC (Shimadzu Corp., Japan) with UV detector (260 nm). To separate GDP-Fuc in the reaction mixture using a Hypersil GOLD aQ column (150 mm × 4.6 mm, 5 μm) (Thermo, CA), the column was equilibrated with buffer A (5 mM tetrabutylammonium hydrogen sulfate and 2% v/v of acetonitrile in 30 mM potassium phosphate buffer, pH 6.0), and eluted with linear gradient from 100% to 60% of buffer A with 0-40% of acetonitrile for 60 min at a flow rate of 1 mL/min.

2.2 Protein engineering of STs

2.2.1 Homology modeling and sequence alignment

The 3D homology modeling for α2,6PdST was done with MODELLER program (Sali and Blundell 1993) using 2Z4T (PDB accession code of α2,6-ST from *Photobacterium* sp. JT-ISH-224) as a template sequence (Kakuta et al. 2008), which has 76% protein sequence similarity with α2,6PdST. The alignment between α2,6PdST and 2Z4T was done using ClustalW1.8. The residues within 5 Å distance from the boundary surface of the substrate molecules in the binding pocket were selected using Swiss PDB Viewer v.3.7. PSI-BLAST (Altschul et al. 1997) searches for α2,3PST and α2,6PdST were done with E-value better than the

threshold (E-value <0.005), and then the called protein sequences were retrieved from the NCBI database. The multiple sequence alignments of α2,3PST and α2,6PdST were performed using clustalW2 (**http://www.ebi.ac.uk/Tools/msa/** clustalw2/). The alignment was examined using SeaView (ver. 4.3.0) to identify the conserved and non-conserved sequences.

2.2.2 Alanine scanning and saturation mutagenesis

For alanine scanning, all the non-conserved amino acid residues in the substrate binding pocket of the target STs were individually mutated to alanine using GCG codon. Soluble fractions of the enzymes were obtained after cell sonication, and then ST activities were measured with the cell-extracts. The levels of the soluble protein expression for alanine-substituted enzymes relative to the wild-type were calculated by the image analysis of the soluble fraction bands on SDS-PAGE using a GeneTools analysis software equipped with G:Box Chemi (Syngene, UK).

Saturation mutagenesis was carried out to generate various single mutants for the selected residues from alanine scanning. The α2,3PST/pET23a and α2,6PdST/pET15b plasmids were used as DNA templates, and NNK codon was used for saturation mutagenesis. The PCR for ST gene amplification was conducted with an initial denaturation at 95 ºC for 5 min, followed by 15 cycles of 95 ºC for 30 sec, 55 ºC for 1 min, 68 ºC for 12 min, and 68 ºC for 12 min for a final extension. The PCR product was subsequently treated with DpnI for 1 h at 37 ºC. 10 μL of the PCR product was transformed into *E. coli* DH5α, and the harvested plasmids from the cells were transformed into *E. coli* BW25113 (DE3).

2.2.3 Mutants screening

In order to screen all the single amino acid-exchange mutants, a sufficient number of colonies (>200 colonies at one site) were examined to satisfy the condition of $P_i > 0.99$, where $P_i = 1 - (1 - F_i)^T$, where F_i represents the frequency at which sequence *i* is present in the NNK library (i.e. since NNK contains 32 codons, F_i becomes 1/32), and P_i is the probability such that the sequence *i* is detected among the T transformants generated from the library (where T is the number of colonies examined). Individual colonies were inoculated into 96-deep-well microplates, with each well containing 500 μL of LB medium supplemented with 100 $\mu g \cdot mL^{-1}$ ampicillin. After sealing the plates with Axyseal sealing film (Axygen Scientific, Inc., Union city, CA), cells were grown in a shaking incubator at 500 rpm and 37 ºC for 18 h. 25 μL of each culture was transferred to a fresh deep-well plate containing 500 μL of LB medium supplemented with 100 $\mu g \cdot mL^{-1}$ ampicillin and 0.5 mM IPTG. The freshly inoculated plates were incubated for 18 h at 30 ºC for α2,3PST and 40 h at 18 ºC for α2,6PdST. The cells were harvested by centrifugation at 3390 ×g for 10 min at 4 ºC. In the case of α2,3PST, cell pellets were thoroughly resuspended in 100 μL of 5 mM Tris-HCl buffer (pH 8.5), and then 10 μL of the whole cells were used to screen mutants. On the other hand, in the case of α2,6PdST, to break out the harvested cells, the cell pellets were thoroughly resuspended with 50 μL of BugBuster protein extraction reagent (Novagen, San Diego, CA), followed by incubation for 20 min at room temperature. The cell debris was subsequently removed by centrifugation. Positive variants were identified from 96-well plates based on their colors and corresponding absorbance

changes. For the colorimetric assay, 10 μL of whole cells or clear supernatants were transferred to a new microplate containing 90 μL of substrate solution (5 mM Tris-HCl, pH 8.5 or pH 8.0, 4 mM lactose, 4 mM CMP-Neu5Ac, and 0.2 mM pH indicator), and then the absorbance changes were measured at several points in time. The positive mutants were cultured once again in a 250 mL shake flask (working volume 50 mL), and the initial ST activity (U/mL) of the mutants was confirmed by the same colorimetric assay.

2.2.4 Purification and kinetic assays

Wild-type α2,3PST, α2,6PdST, and the screened mutants expressed in *E. coli* BW25113 (DE3) cells were harvested and purified by Ni-NTA column (QIAGEN Inc., Valencia, CA). First, a protein mixture was loaded onto a column equilibrated with a 50 mM Tris-HCl buffer (pH 8.0) containing 5 mM imidazole and 300 mM NaCl. The ST fraction was eluted from the column by applying a 50-250 mM imidazole buffer containing 300 mM NaCl. The active fractions were then collected and desalted by VIVASPIN (Sartorius Stedim Biotech GmbH, Germany). The protein concentration was determined using a Bradford protein assay kit (Bio-Rad Laboratories Inc., Hercules, CA). Kinetic analysis of ST variants for acceptor and donor substrates was carried out using the same colorimetric method. The kinetic parameters of k_{cat} and K_m were obtained by nonlinear regression analysis of the Michaelis-Menten equation using SigmaPlot 10.0.

2.3 Increase of soluble expression of FucTs

2.3.1 Soluble expression of α1,2-FucT

2.3.1.1 Codon optimization and co-expression with chaperone

The *futC*, α1,2-FucT gene was newly synthesized for codon optimization with consideration for GC codon %, minimization of repeat sequence and stabilization of mRNA secondary structure using "POMBE" program (Bioneer, Korea). The synthesized gene subcloned into T-vector was amplified using the primers (forward: 5′-TCGACTCATATGGCCTTTAAGGTG-3′, reverse: 5′-TCGACTCTCGAGATC TAAAGCGTTATACTT) under the PCR condition: initial denaturation at 95 ºC for 5 min, followed by 30 cycles of 95 ºC for 30 sec, 55 ºC for 1 min, 72 ºC for 1 min 10 sec, and 72 ºC for 7 min for a final extension using 1X PCR buffer, 0.2 mM dNTP, 2.5 mM $MgCl_2$, and 100 pmol of primers, 50-100 ng of template DNA, and 2.5 U of Pfu DNA polymerase (Solgent, Korea). The PCR product was restricted using NdeI/XhoI and ligated into pET24ma treated the same restriction enzymes.

Codon-optimized FutC was co-expressed with pBAD vector harboring GroEL/GroES chaperone in *E. coli* BW25113 (DE3). The *E. coli* cells harboring the *futC* and GroEL/GroES genes were cultured in LB media supplemented with Kanamycin (50 μg mL^{-1}), Amplicillin (100 μg mL^{-1}), and 1 mM arabinose at 37 ºC. When OD600 was reached at 0.5-0.7, the cells were induced with various concentration of IPTG (0.005-0.5 mM) and further incubated at 18 ºC for 20 h.

2.3.1.2 Introduction of fusion protein

For the increase of soluble expression of FutC α1,2-FucT, various fusion proteins were introduced at N-terminus of codon-optimized FutC (cFutC). For the construction of fusion protein linked cFutC, various fusion proteins were cloned into pET24ma using NdeI and SacI restriction enzymes. Among the fusion proteins, gene sequences of F-ePGK (full length of *E. coli* phospho-glycerate kinase), N-ePGK (N-domain of *E. coli* phosphoglycerate kinase, Figure 2.3A) (Song et al. 2012), ArsC (*E. coli* arsenate reductase) (Song et al. 2011), GMPK (*E. coli* guanylate kinase), and ACK (*E. coli* acetate kinase) were retrieved from NCBI database. In the case of 30Kc19 originated from silkworm *Bombyx mori* (Yang et al. 2011), pET28a harboring the gene was used as a template for amplification. According to the sequence information, primers were generated (Table 2.1) to include NdeI and SacI restriction sites for the cloning into pET24ma. Each F-ePGK, N-ePGK, ArsC, GMPK, and ACK gene was amplified by PCR from genomic DNA of *E. coli* K12, and ligated into the NdeI-SacI site of pET24ma plasmid under T7 promoter. To construct the fusion expression vector with cFutC, cFutC was subsequently ligated into the pET24ma vector using the primers as shown in Table 2.1 after each of fusion proteins. To insert the spacer sequence between the fusion protein and cFutC, the primer for amplification of cFutC was designed to have enterokinase cleavage sequence (D4K) between fusion protein and cFutC. As shown in Figure 2.3B, in addition, linker sequence containing $(Gly_4Ser)_2$ was followed by cFutC (target gene) for efficient further purification of the cFutC gene.

Table 2.1 Primers used to construct fusion protein linked FutC

Fusion protein	Primers
F_ePGK	F: 5′-TCGACTCATATGTCTGTAATTAAGATG-3′
	R: 5′-TCGACTGAGCTCCTTCTTAGCGCGCTC-3′
N_ePGK	F: 5′-TCGACTCATATGTCTGTAATTAAGATG-3′
	R: 5′-TGCTTTACCCAGCGCGTC-3′
	F': 5′-GGTAAAGCA GGCGGT GAAGGTAAAGTACTG-3′
	R': 5′-TCGACTGAGCTCCTTCTTAGCGCGCTC-3′
ArsC	F: 5′-TCGACTCATATGAGCAACATTACCATT-3′
	R: 5′-TCGACTGAGCTCTTTCAGGCGCTTACCCGC-3′
GMPK	F: 5′- TCGACTCATATGGCTCAAGGCACGCTTTAT-3′
	R: 5′-TCGACTGAGCTCGTCTGCCAACAATTTGCT-3′
ACK	F: 5′- TCGACTCATATGTCGAGTAAGTTAGT-3′
	R: 5′-TCGACTGAGCTCGGCAGTCAGGCGGCTCGC-3′
α1,2-FucT	**Primers**
FutC[1]	F: 5′-GCATGAGCTCGACGATGACGATAAAATGGCCTTTAAGGTG-3′
	R: 5′-ATATCTCGAGAGAGCCACCTCCGCCTGAACCGCCTCCACCAT CTAAAGCGTTATACTTCTG-3′
FutC[2]	F: 5′-TCGACTCCCGGGATGGCCTTTAAGGTGGTG-3′
	R: 5′-TCGACTCTCGAGCTAATCTAAAGCGTTATACTT-3′

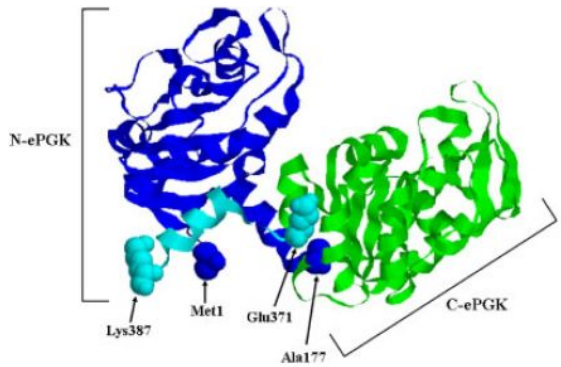
Primers for fusion proteins were used for cloning of fusion proteins into pET24ma using NdeI and SacI.

FutC[1] was used to generate the fusion construction of FutC at C-terminus of F_ePGK, N_ePFK, ArsC, GMPK, ACK, and 30Kc19.

FutC[2] was used to generate the fusion construction of FutC at C-terminus of GST harbored in pGEX 4T-1.

Underline presents restriction enzymes used for cloning.

A



B

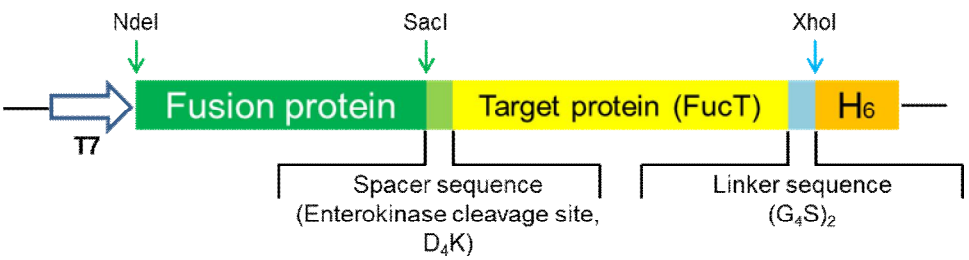


Figure 2.3 Construction of fusion protein-linked vectors used for soluble expression of target protein, cFutC

(A) Crystal structure of *E. coli* PGK (PDB ID: 1ZMR). The fusion partner N-ePGK includes the segments from Met1 to Ala177 and from Glu371 to Lys378 that were linked by glycine linker (-Gly-Gly), that is, N-(Met1 to Ala177)-Gly-Gly-(Glu371 to Lys387)-C. (B) Fusion expression vector is designed for fusion protein, spacer sequence, target protein, liker sequence, and His tag in order.

In the case of GST (glutathione-S-transferase) (Shimada et al. 2005) fusion, pGEX 4T-1 vector harboring GST was used for construction of fusion vector of cFutC. The cFutC gene was PCR amplified using the primers including SmaI/XhoI sites (Table 2.1), and ligated into pGEX 4T-1 vector to be followed by GST.

Each generated fusion expression vector was transformed into *E. coli* BW25113 (DE3), and the cells were incubated in LB broth with 50 $\mu g \cdot mL^{-1}$ of Kanamycin or 100 $\mu g \cdot mL^{-1}$ of Ampicillin. And then the cells were induced with 0.01 mM IPTG, when OD_{600} of the culture reached 0.5-0.8. The induced cells were further incubated in a rotary shaker at 200 rpm and 18 ºC for 20 h. The cells were harvested from 50 mL of the culture broth by centrifugation at 3,390 × g for 10 min at 4 ºC, suspended in 5 mL of 20 mM sodium phosphate buffer (pH 7.6), and lysed by sonication in ice. The supernatant of the cell lysate (soluble fraction) was obtained by centrifugation at 22,250 × g for 30 min at 4 ºC.

2.3.2 Soluble expression of α1,3-FucT

2.3.2.1 Systemic truncation of C-terminus

To increase the expression of the soluble fraction of FutA, C-terminus consisting of α-helix and heptad repeat regions was systemically truncated using the primers as shown in Table 2.2. Each plasmid harboring *futA* differentially truncated at C-terminus was transformed into *E. coli* BW25113 (DE3), and the cells were induced with 0.05 mM IPTG, when OD_{600} of the culture reached 0.5-0.8 in LB media. The induced cells were further incubated in a rotary shaker at 200 rpm and 18 ºC for 20 h. The cells were harvested from 50 mL of the culture broth

Table 2.2 Primers used to construct C-terminal truncated FutA

Truncated form	Cloning primers
Δ9	F: 5′- TCGACTCATATGTTCCAACCCCTATTAG-3’ R: 5′- TCGACTCTCGAGCACCGCGCGCAACAAAGG-3’
Δ45	F: 5′- TCGACTCATATGTTCCAACCCCTATTAG-3’ R: 5′- TCGACTCTCGAGATAATTAACCCTCAAATCATCATAATTA-3’
Δ52	F: 5′- TCGACTCATATGTTCCAACCCCTATTAG-3’ R: 5′- TCGACTCTCGAGATAATTAACCCTCAAATCATCAATGGAT-3’
Δ59	F: 5′- TCGACTCATATGTTCCAACCCCTATTAG-3’ R: 5′- TCGACTCTCGAGAATGGATACTAACGGCTT-3’

Each truncated form was cloned into pET24ma using NdeI and XhoI restriction enzymes.

by centrifugation at 3,390 × g for 10 min at 4 ºC, suspended in 5 mL of 20 mM sodium phosphate buffer (pH 7.6), and lysed by sonication in ice. The supernatant of the cell lysate (soluble fraction) was obtained by centrifugation at 22,250 × g for 30 min at 4 ºC.

2.3.2.2 Codon optimization and expression optimization

The truncated FutA (Δ52 FutA) gene was newly synthesized with codon optimization to maintain high level of charged tRNA (DNA2.0, Menlo Park, CA) (Welch et al. 2009), and then cloned into pJExpress401 containing T5 promoter. Codon-optimized Δ52 FutA was transformed into *E. coli* BW25113 (DE3), induced with 1 mM IPTG in TB media, and the cells were further incubated at 30 ºC for 12 h. The preparation of the enzyme was proceeded following the same method as above.

2.4 Protein engineering of FucTs

2.4.1 Homology modeling and docking simulation of lactose

The 3D homology modeling for Δ52 FutA was done with MODELLER program (Sali and Blundell 1993) using 2NZY (PDB accession code of α1,3-FucT from H. pylori NCTC11639) as a template (Sun et al. 2007), which has 89% protein sequence identities with Δ52 FutA. Putative catalytic key amino acid of Δ52 FutA was selected from the superimposed structure with 2NZY and the substrate binding site from the model structure. In addition, the coordinates of lactose substrate were generated manually and energetically optimized using the

MM2 force field using Chem3D Ultra 8.0. AutoDock (version 3.00) was applied to the docking of lactose into the homology model of Δ52 FutA by using Oε2 of the key residue as a grid center. Among the 200 docking poses generated from the docking simulations, the one with minimum docking energy value E/δ (<-0.97 kcal/(mol·Å)) was selected.

2.4.2 Selection of functional residues for the focused directed evolution

Residues within 12 Å distance from the Cα of Glu96 were picked out using PyMOL after identification of key catalytic residue of Δ52 FutA. HotSpot Wizard 1.7 (http://loschmidt.chemi.muni.cz/hotspotwizard/) was used to call the residues with high mutability, and the overlapping residues among the above selected residues were sorted out for SSM. To make an expanded mutational range considering the binding sites of the two substrates (GDP-Fuc donor and lactose acceptor), putative functional residues were identified using the mentioned docking simulation. The residues within 8 Å distance from the boundary surface of substrates were picked out following the same method as above. Using the HotSpot Wizard, the residues with high mutability among the selected residues were also sorted out in the same manner.

2.4.3 Saturation mutagenesis and library creation

For the site-directed mutagenesis, E96 chosen as a key catalytic key residue from the model structure of Δ52 FutA was substituted for alanine using GCG codon.

Saturation mutagenesis was carried out to search 'best hit' single mutation site among the selected residues within 12 Å from Cα of E96 at 1st generation mutagenesis. The pJExpress401 plasmid carrying the codon optimized *futA* gene truncated at C-terminus (Δ52 *futA*) was used as a DNA template, and NNK (N = G or A or T or C, K = G or T nucleotide) codon was used to generate the library of the each residue. The PCR for amplification of the plasmid including Δ52 *futA* gene was conducted with an initial denaturation at 95 ºC for 5 min, followed by 15 cycles of 95 ºC for 30 sec, 55 ºC for 1 min, 68 ºC for 12 min, and 68 ºC for 12 min for a final extension. The PCR product was subsequently treated with DpnI for 1 h at 37 ºC. The PCR product (10 μL) was transformed into *E. coli* DH5α, and the harvested plasmids from the cells were transformed into *E. coli* BW25113 (DE3). For ISM at 2nd generation mutagenesis, the best mutant from the 1st generation mutagenesis was used as a DNA template for the next saturation mutagenesis, and NNK codon was applied to other sites to change. PCR amplification and library construction were conducted as explained above.

2.4.4 Mutants screening

In order to screen all the single amino acid-changed mutants, sufficient numbers of colonies (>200 colonies at one site) were examined to satisfy the condition of 99% probability (Choi et al. 2014). Individual colonies were inoculated into 96-deep-well microplates, where each well contained 500 μL of TB medium supplemented with 50 μg mL^{-1} kanamycin. After sealing the plates with Axyseal sealing film (Axygen Scientific, Inc., Union city, CA), cells were grown in a shaking incubator at 500 rpm and 37 ºC for 18 h. Each culture (25 μL) was

transferred to a fresh deep-well plate containing 500 μL of TB medium supplemented with 50 μg mL-1 kanamycin and 0.5 mM IPTG. The freshly inoculated plates were incubated for 18 h at 30 ºC. The cells were harvested by centrifugation at 3390 ×g for 10 min at 4 ºC. Cell pellets were thoroughly resuspended with 50 μL of BugBuster protein extraction reagent (Novagen, San Diego, CA), followed by incubation for 20 min at room temperature. The cell debris was subsequently removed by centrifugation. Positive variants were identified from 96-well plates based on their color changes and corresponding absorbance changes according to the proton release during the reaction (Choi et al. 2014; Park et al. 2009). For the colorimetric assay, 20 μL of clear supernatant was transferred to a new microplate containing 80 μL of substrate solution (10 mM Tris-HCl, pH 8.0, 10 mM lactose, 4 mM GDP-Fuc, and 0.4 mM phenol-red), and then the absorbance changes were measured at 560 nm as a function of time. The control well used to calculate ΔOD for the color changes of each well contained the same reaction mixture except for lactose substrate. The positive mutants showing higher activities than that of wild-type were cultured once again in a 250-mL shake flask (working volume 50 mL), and the initial activity (U/mL) and yield for the synthesis of 3-FL were measured by Bio-LC.

2.4.5 Determination of specific activity and kinetic parameters

Wild-type FutA (i.e. Δ52 FutA), and the screened mutants expressed in E. coli BW25113 (DE3) cells were harvested and purified by Ni-NTA column (QIAGEN Inc., Valencia, CA). First, a protein mixture was loaded onto a column equilibrated with a 50 mM Tris-HCl buffer (pH 8.0) containing 5 mM imidazole and 300 mM

NaCl. The enzyme fraction was eluted from the column by applying a 50-250 mM imidazole buffer containing 300 mM NaCl. The active fractions were then collected and desalted by VIVASPIN (Sartorius Stedim Biotech GmbH, Germany). The protein concentration was determined using a Bradford protein assay kit (Bio-Rad Laboratories Inc., Hercules, CA). Specific activities (U/mg) were determined with 5 mM GDP-Fuc, 15 mM lactose, and 20 μg of purified FutA in 50 mM sodium phosphate buffer (pH 7.6) at 37 ºC. Produced 3-FL was quantified using Bio-LC and enzyme activity was defined as the amount of enzyme required to produce 1 μmol of 3-FL per minute at 37 ºC.

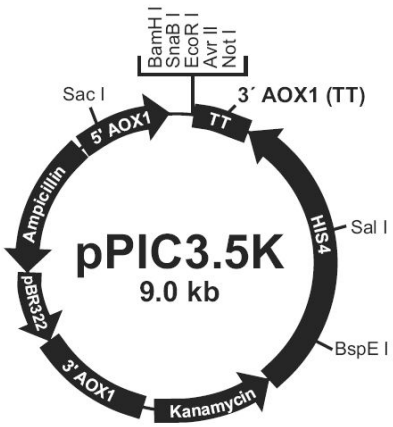
Kinetic analyses of Δ52 FutA variants for acceptor and donor substrates were carried out using a luminescent GDP-Glo assay kit (Promega Corp., Madison, WI) by the quantification of released GDP. The kinetic parameters of k_{cat} and K_m were obtained by nonlinear regression analysis of the Michaelis-Menten equation using SigmaPlot 10.0.

2.5 *in vivo* production of 3-FL

2.5.1 Construction of yeast strain for GDP-Fuc production

The GMD and GFS (WcaG) genes, amplified by PCR from *E. coli* K12 genomic DNA, were cloned into the pPIC3.5K vector using EcoRI/NotI and BglII/NotI restriction enzymes, respectively, to construct a vector system for multiple integration into the *Pichia pastoris* genome. As shown in Figure 2.4, GMD and GFS gene were integrated with P_{AOX1} (alcohol oxidase promoter) in a form of 'P_{AOX1}-Kozak seq-GFS-TT-P_{AOX1}-Kozak seq-GMD-TT'.

A



B

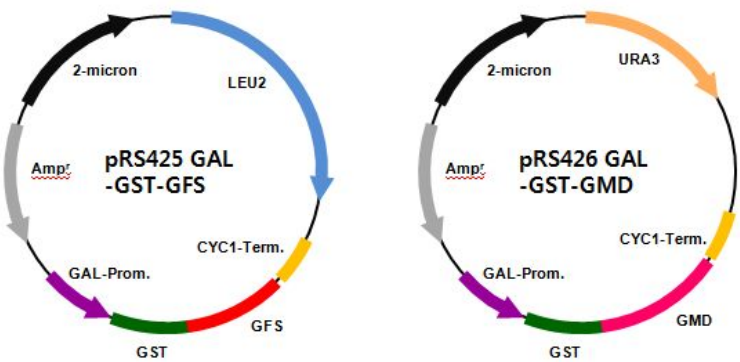


Figure 2.4 Vector construction for expression of GMD and GFS in yeasts

(A) pPIC3.5K vector for *in vivo* multiple integration of GMD and GFS in the *P. pastoris* genome. (B) pRS425GAL-GST and pRS426GAL-GST vector for expression in *S. cerevisiae*

Whereas, for the construction of expression system in *Saccharomyces cerevisiae*, the GMD and GFS genes were cloned into BamHI and XhoI sites of pRS426GAL-GST and pRS425GAL-GST, generating pRS425GAL-GST-GFS and pRS426GAL-

GST-GMD, respectively (Figure 2.4). Primers used for plasmid construction are shown in Table 2.3.

For the production of GDP-Fuc in *P. pastoris*, single colony was cultured in Buffered glycerol-complex medium (BMGY) at 30 ºC until the OD_{600} was reached at 3-4 of log phase. Cells grown in BMGY was harvested and resuspended in 100 mL of Buffered methanol-complex medium (BMMY) to make 1 of OD_{600}. 0.5% of methanol was added every 24 h for 4-5 days to induce GMD and GFS genes. After cultivation for 4-5 days, cells were harvested, lysed by vortex with glass beads in a lysis buffer.

On the other hand, overnight cultured *S. cerevisiae* cells harboring the pRS425GAL-GST-GFS and pRS426GAL-GST-GMD were grown from an OD_{600} of 0.3 to 0.7 in SC-Ura-Leu medium containing 2% dextrose, washed twice with distilled water and transferred into SC-Ura-Leu medium containing 4% galactose for induction. After cultivation for 24 h, cells were harvested, lysed by vortex with glass beads in a lysis buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol, and 0.5% NP-40] containing 0.1% protease inhibitor cocktail (Calbiochem). Produced GDP-Fuc was quantitatively analyzed as mentioned at section 2.1.3.3.

Table 2.3 Primers used for cloning of GMD and GFS for construction of yeast expression system

GMD/ GFS	Primers
GMD_P	F: 5′-ATATGAATTCACCATGGGAAAAGAAAAAATCGCTTTAATC-3′
	R: 5′-ATATGCGGCCGCTCAGTGGTGGTGGTGGTGGTGTTCATAAAA ATTCCTTAAAA-3′
GFS_P	F': 5′-ATATATAGATCTACCATGGGAAATGAGATTATT-3′
	R': 5′-ATATGCGGCCGCTCAGTGGTGGTGGTGGTGGTGAACCTCTA AAAGCTTCAAATA-3′
GMD_S	F: 5′-TCGACTGGATCCATGAAAGAAAAAATCGCTTTA-3′
	R: 5′-GTATGTCTCGAGTTCATAAAAATTCCTTAAAAT-3′
GFS_S	F: 5′-TCGACTGGATCCATGAATGAGATTATTTTAATC-3′
	R: 5′-GTATGTCTCGAGAACCTCATTAAGCTTCAAATA-3′

GMD_P and GFS_P were used for cloning into pPIC3.5K for *P. pastoris* culture

GMD_S and GFS_S were used for cloning into pRS425/426-GST for *S. cerevisiae* culture.

Underlines display the restriction enzymes used for cloning.

2.5.2 Construction of *E. coli* strain for 3-FL productionAll strains, plasmid used in this study are listed in Table 2.4. *E. coli* DH5α (Invitrogen, USA) was used for plasmid cloning and propagation, while *E. coli* BW25113 (DE3) and *E. coli* BL21 (DE3) were used for construction of stain to produce 3-FL. pJExpress401 (DNA2.0, Menlo Park, CA) and pCDFm (modified by our lab from pCDFDuet-1 (Novagen, USA)) vectors were used for cloning and expression of desired genes in *E. coli* BW25113 (DE3) or BL21 (DE3).

Chromosomal DNA from *E. coli* K12 was prepared using G-spin (TM) for Bacteria Genomic DNA Extraction kit (iNtRON, Seungnam, South Korea). Plasmids from *E. coli* strains were prepared using the GeneAll DNA Purification System (Geneall Biotechnology, Seoul, South Korea). Primers used in this study were commercially synthesized by Cosmo Bioscience (Cosmo, Seoul, South Korea).

To construct the salvage pathway system from extracellular L-fucose, fucose isomerase (*fucI*) and fuculose kinase (*fucK*) genes responsible for L-fucose degradation were deleted from the chromosome of BW25113 (DE3), resulting in strain BW ΔfucIK (Table 2.4). Deletion of *fucIK* gene locus was also conducted in BL21 (DE3) strain. In company with deletion of *fucIK*, β-galactosidase (*lacZ*) on chromosome of BL21 (DE3) was knocked out to eliminate the lactose degradation pathway, resulting in strain BL21 ΔlacZ ΔfucIK (Table 2.4). Genetic deletion and integration were carried out using the Red recombination system, which is a one-step inactivation system developed by Datsenko and Wanner (2000). Briefly, transformant carrying a Red helper plasmid was grown in 5-ml SOB cultures with ampicillin and L-arabinose at 30 °C until OD_{600} was reached at 0.6.

Table 2.4 Bacterial strains and plasmids for construction of *E. coli* strain for the production of 3-FL

Strains	**Description**
E. coli DH5α	F- , ϕ80d, lacZΔM15, endA1, recA1, hsdR17(rK - mK -), supE44, thi-1, gyrA96, relA1, Δ(lacZYA-argF)U169
E. coli BW25113 (DE3)	F-, DE(araD-araB)567, lacZ4787(del)::rrnB-3, LAM-, rph-1, DE(rhaD-rhaB)568, hsdR514
BW ΔfucIK	*E. coli* BW15113 (DE3) Δ*fucIK*
E. coli BL21 (DE3)	F– *ompT* gal dcm lon *hsdSB*(rB- mB-) λ(DE3)
BL21 ΔlacZ ΔfucIK	*E. coli* BL21 (DE3) Δ*lacZ* Δ*fucIK*
Plasmids	**Description**
pCDFm	T7 promoter, CloDF13 ori, Sm^r modified for multi-cloning
pJExpress401	T5 promoter, ColE1 ori, Kan^r
pC-F	pCDFm - *fkp* (NcoI/SalI)
pC-FG	pCDFm - *fkp* (NcoI/SalI) + T7_*gsk* (BamHI/SpeI)
pC-FGL-1	pCDFm - *fkp* (NcoI/SalI) + T7_*gsk* (BamHI/SpeI) + T7_*lacY* (SpeI/PacI)
pC-FGL-2	pCDFm - *fkp* (NcoI/SalI) + T7_*gsk* (BamHI/SpeI) + rbs_*lacY* (XbaI/PacI)
pC-FGN-1	pCDFm - *fkp* (NcoI/SalI) + T7_*gsk* (BamHI/SpeI) + T7_*ndk* (SpeI/PacI)
pC-FGN-2	pCDFm - *fkp* (NcoI/SalI) + T7_*gsk* (BamHI/PacI) + rbs_*ndk* (XbaI/PacI)
pC-GpGmNF	pCDFm - P_{T7}_*gpt* (BamHI/PacI) + rbs_*gmk* (BamHI/PacI) + rbs_*ndk* (BamHI/PacI) + *fkp* (BamHI/PacI)
pC-BCGW	pCDFm - *manB-C* (NcoI/HindIII) + P_{T7}_*Gmd-WcaG* (NdeI/EcoRV)
pC-BCGW+G	pCDFm - *manB-C* (NcoI/HindIII) + P_{T7}_*Gmd-WcaG* (NdeI/EcoRV) + rbs_*gsk* (SpeI/PacI)
pC-BCGW+G+F	pCDFm - *ManB-C* (NcoI/HindIII) + P_{T7}_*Gmd-WcaG* (NdeI/EcoRV) + rbs_*gsk* (SpeI/PacI) + rbs_*fkp* (BamHI/PacI)
pJ401-FutA	pJExpress401 - Δ52 FutA quadruple mutant

And then, cells were washed three times with ice-cold 10% glycerol and concentrated by 100-fold to make electro-competent cells. PCR products were gel-purified, digested with DpnI, and eluted in water. Electroporation was performed by using a Cell-Porator with a voltage booster and 0.15-mm chambers according to the manufacturer's instructions (GIBCO BRL) by using 25 μL of cells and 10–100 ng of PCR product. The 1 ml of SOC was added in shocked cells, and then one-half of the cells were spread onto agar plate to select an appropriate antibiotics (chloramphenicol or apramycine) resistant transformant after incubation for 1 h at 37 °C. After primary selection, mutants were kept in a medium without corresponding antibiotics. They were grown non-selectively and colonies were purified once at 37 °C; they were then tested for ampicillin sensitivity to test for the loss of the helper plasmid. All primer information of the deletion and integration targets is listed in the Table 2.5.

To construct the plasmid for 3-FL production by salvage pathway, *fkp* gene cloned in pET16b was amplified by PCR with the set of FKP-F/FKP-R primers (Table 2.6). Amplified *fkp* gene was cloned into pCDFm vector using NcoI/SalI enzymes to yield pC-F (Table 2.4). To construct gene cluster at downstream of *fkp*, target gene was subcloned into expression vector carrying T7 promoter such as pET24ma. Expression vector harboring the target gene was then amplified by PCR to allow carrying T7 promoter or rbs region for cloning into pCDFm vector. *gsk* gene of salvage pathway was amplified from *E. coli* K-12 and cloned into pET24ma. pET24ma carrying *gsk* gene was amplified by PCR using T7-F (BamHI) and Gsk-R primers, and digested with BamHI/SpeI and cloned into pCDFm vector treated with BglII and SpeI.

Table 2.5 Primers used in the constrction of deletion strain

Deletion	Pairs of primer
fucI-F	5′-TAGCAG AGCGCCAATA AAGCGAGATG AGGAATCCTG ATG TCATCGCAGTACTGTTG-3′
fucK-R	5′-AGGTTCAGTTTGCGGGTAGAAATAACGGTACTGATAATGTAG GAACTTCGGCGCGCCT-3′
lacZ-F	5′-TTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATG GTGTAGGCTGGAGCTGCTTC-3′
lacZ-R	5′-GCGAAATACGGGCAGACATGGCCTGCCCGGTTATTATTA CATATGAATATCCTCCTTA-3′
Insertion	**Pairs of primer**
malEFG-hindsacF	5′-ATAT AAGCTT GAGCTC GATTAAGGGA GATAACAAAA-3′
malEFG-xbaR	5′- AAT TCTAGA ACAATTCCACACATTATACGAGCCGATGATTA ATTGTCAAAGGCGCTCCGTGGAGCG-3′
TacP-bamF	5′-ATAT GGATCC TTGACAATTAATCATCGGCTC-3′
Rrnb-malfuR	5′-CGATTCAGATTGCAGACGCGCAAAAAGGCCATCCGTC-3′
Rrnb-malfuF	5′-GACGGATGGCCTTTTTGCGCGTCTGCAATCTGAATCG-3′
MalEFG-ecoR	5′- TAT GAATTC GCAGGGATACATTCCGAGG-3′

Table 2.6 Primers used in the construction of vectors

Cloning	Pairs of primer
FKP-F	5′- TCGACTCCATGGGCCAAAAACTACTATCTTTA-3′
FKP-R	5′- TCGACTGTCGACTTATGATCGTGATACTTGGAA-3′
T7-F (BamHI)	5′-TCGACTGGATCCGAAATTAATACGACTCACTATAGGGGAATTGTG-3′
T7-F (SpeI)	5′- TCGACGACTAGTGAAATTAATACGACTCACTATAGG-3′
RBS-F (XbaI)	5′- TCGACGTCTAGAAATAATTTTGTTTAACTT-3′
RBS-F (BamHI)	5′- TCGACTGGATCCAATAATTTTGTTTAACTTTAA-3′
Gsk-R	5′- TCGACTACTAGTTTAACGATCCCAGTAAGACTC-3′
LacY-R	5′- TCGACTTTAATTAAAGATCTTTAAGCGACTTCATTCAC-3′
Ndk-R	5′- TCGACTTTAATTAAAGATCTTTAACGGGTGCGCGGGCA-3′
Gpt-R	5′- TCGACTTTAATTAAAGATCTTTAGCGACCGGAGATTGG-3′
Gmk-R	5′- TCGACGTTAATTAAAGATCTTCAGTCTGCCAACAATTT-3′
FKP-R1	5′- TCGACTTTAATTAAAGATCTTTATGATCGTGATACTTG-3′
BC-F	5′-ATATCCATGGGCATGGCGCAGTCGAAACTC-3′
BC-R	5′-AATTAAGCTTTTACTCGTTCAGCAACGTC-3′
WG-F	5′- ATATCATATGTCAAAAGTCGCTCTC-3′
WG-R	5′- AATGATATCTTACCCCCGAAAGCGGTC-3′
Subcloning	**Pairs of primer**
Gsk-F_24	5′- TCGACTCATATGAAATTTCCCGGTAAA-3′
Gsk-R_24	5′- TCGACTCTCGAGACGATCCCAGTAAGACTC-3′
LacY-F_24	5′- TCGACTGGATCCATGTACTATTTAAAAAAC-3′
LacY-R_24	5′- TCGACTCTCGAGTTAAGCGACTTCATTCAC-3′
Ndk-F_24	5′- TCGACTCATATGGCTATTGAACGTACT-3′
Ndk-R_24	5′- TCGACTCTCGAGTTAACGGGTGCGCGGGCA-3′
Gpt-F_24	5′- TCGACTCATATGAGCGAAAAATACATC-3′
Gpt-R_24	5′- TCGACTCTCGAGTTAGCGACCGGAGATTGG-3′
Gmk-F_24	5′- TCGACGCATATGGCTCAAGGCACGCTT-3′
Gmk-R_24	5′- TCGACGCTCGAGTCAGTCTGCCAACAATTT-3′

To add *lacY* gene at downstream of of *gsk*, pET24ma carrying *lacY* was amplified by PCR using T7-F (SpeI)/LacY-R primer pairs for cloning of P_{T7}_lacY, or using RBS-F (XbaI)/LacY-R primer pairs for cloning of rbs_lacY. Amplified genes are digested with SpeI/PacI for cloning of P_{T7}_lacY or XbaI/PacI for cloning of rbs_lacY, and cloned into pCDFm vector treated with SpeI and PacI. Whereas, to add *ndk* gene at downstream of of *gsk*, pET24ma carrying *ndk* was amplified by PCR using T7-F (SpeI)/Ndk-R primer pairs for cloning of P_{T7}_Ndk, or using RBS-F (XbaI)/ Ndk-R primer pairs for cloning of rbs_Ndk. Amplified genes are digested with SpeI/PacI for cloning of P_{T7}_ndk and XbaI/PacI for cloning of rbs_ndk, and cloned into pCDFm vector treated with SpeI/PacI. To combine *fkp* with *gpt*, *gmk*, and *ndk* gene in pCDFm vector, *gpt*, *gmk*, and *ndk* genes subcloned into pET24ma were amplified using T7-F (BamHI) and Gpt-R or Gmk-R or Ndk-R primers for cloning under T7 promoter. For cloning under rbs sequence, *gmk* and *ndk* genes were amplified using RBS-F (BamHI) and Gmk-R or Ndk-R primers. Amplified genes are digested with BamHI/PacI and cloned into pCDFm vector in order.

All primers used in the cloning for construction of vectors are listed in the Table 2.6.

2.5.3 Media optimization

Minimal media was used for biotransformation from glycerol, L-fucose, and lactose to 3-FL production. The media was prepared with 3 g/L KH_2PO_4, 12 g/L K_2HPO_4, 5 g/L $(NH_4)_2SO_4$, 0.1 g/L NaCl, 0.3 g/L $MgSO_4 \cdot 7H_2O$, 0.015 g/L $CaCl_2 \cdot 2H_2O$, 7.5 g/L $FeSO_4 \cdot 7H_2O$ in 100 g/L sodium citrate, 7.5 µg/L thiamine, and trace element [100X trace element, 5 g/L EDTA, 0.83 g/L $FeCl_3 \cdot 6H_2O$, 84

mg/L $ZnCl_2$, 13 mg/L $CuCl_2 \cdot 2H_2O$, 10 mg/L $CoCl_2 \cdot 2H_2O$, 10 mg/L H_3BO_3, 1.6 mg/L $MnCl_2 \cdot 4H_2O$, pH 7.5]. To increase the solubility of α1,3-FucT, yeast extract was added in minimal media according to the composition of yeast extract in TB broth. And then, the concentration of yeast extract in minimal media was optimized for the solubility of α1,3-FucT and 3-FL production.

2.5.4 Culture condition

2.5.4.1 Batch culture in flask

For the production of 3-FL using co-expression system of two plasmids, pCDFm and pJExpress401, the vectors harboring the corresponding genes were transformed into engineered *E. coli* cells by heatshock method developed by Hanahan (1983). Transformant, which is selected on Kan and Sm antibiotics, was cultured in selective LB medium. Seed culture grown in LB medium for overnight was transferred to a 250 mL baffled flask to make 1% inoculation volume (v/v) for main culture. 250 mL flask contained 50 mL of minimal or optimized media supplemented with 50 μg/mL of Kan/Sm and 5 g/L glycerol as carbon source. Cells were cultured at 37 °C with 200 rpm until cells had reached OD_{600} at 0.9-1.0, and then 0.1 mM of IPTG was added to the media. At time of induction, 10 mM L-fucose and 20 or 40 mL lactose were added, and cells were further cultured at 30 °C for protein expression and 3-FL production.

2.5.4.2 Fed-batch culture in flask

Fed-batch cultivation was performed using a 250 mL baffled flask with 50 mL of working volume of minimal or optimized media as the same method mentioned above. In the case of optimized media, 1% yeast extract was added to minimal media along with 5 g/L of glycerol. Cells were cultured at 37 °C with 200 rpm until cells had reached OD_{600} at 0.9-1.0, and then 0.1 mM of IPTG was added to the media. After addition of substrates with induction, cultivation was proceeded at 30 °C with 200 rpm at a pH controlled to 7.0 using 25% (v/v) ammonium hydroxide solution. After glycerol added initially was completely consumed, 50% (v/v) glycerol solution dissolved in minimal media with antibiotics was added to the culture media for making 1.5 g/L of glycerol concentration. During the cultivation, glycerol was fed when the glycerol concentration remained below 1 g/L. Glycerol concentration was determined by chemical assay using chromotropic acid (Sturgeon et al. 1979).

2.5.5 Quantification of intracellular and extracellular 3-FL

During the cultivation, 1 mL of culture was sampled and centrifuged (22,250 × g, 3 min) to separate the media and cells. Media was heated at 95 °C for 40 sec, and centrifuged at 22,250 × g for 20 min to get clear supernatant for the detection of 3-FL secreted in media. Cell pellets were washed with 1 mL of deionized water and resuspended in 400 μL 10 mM sodium phosphate buffer (pH 7.6). Cells were disrupted by sonication, heated at 95 °C for 40 sec, centrifuged at 22,250 × g for 20 min to get supernatant for the detection of 3-FL in cytosol. Produced 3-FL was quantified by Bio-LC as mentioned in 2.1.3.2 section.

Chapter 3.

Protein engineering of α2,3/2,6-sialyltransferase for *in vitro* sialyllactose synthesis

3.1 Selection of target region and alanine scanning

3.1.1 Selection of target region

Prior to the selection of target residues for alanine scanning, the 2IHZ crystal structure of α2,3PST (Ni et al. 2007) was used to identify substrate binding pocket regions. For the homology modeling of α2,6PdST, 2Z4T was used as a template, whose structural data was acquired with a protein-ligand complex of CMP and lactose (Kakuta et al. 2008). When the model structure of α2,6PdST was superimposed onto the 2IHZ structure of α2,3PST, its structure was very similar to 2IHZ due to the 58% amino acid similarity. Both α2,3PST and α2,6PdST have GT-B fold structures consisting of two separate Rossmann-fold domains. According to the 2IHZ α2,3PST crystal structure, the catalytic key residue that functions as a general base to deprotonate from the reactive hydroxyl group of lactose acceptor was identified as Asp141 (Kim et al. 2008; Ni et al. 2007), which corresponded to Asp215 of α2,6PdST. The active site of the α2,3PST and α2,6PdST is located in the deep cleft between the two Rossmann domains. In the case of α2,3PST, upon the binding of CMP-Neu5Ac to the active site, its closed conformation (i.e. active form) is induced by the movement of α12a/α12b helices in the C-terminal domain, which is known to help define the subsequent lactose binding site (Ni et al. 2006).

According to the 2IHZ α2,3PST crystal structure, 76 amino acid residues were identified within a sphere of 12 Å radius from the C_α of Asp141. Among them, when the two substrate binding pocket regions were examined in detail following the catalytic mechanism, 10 and 30 residues were identified to interact with the two substrates within 5 Å distance from the boundary surface of lactose (specifically, 5

Å distance from the oxygen atoms of hydroxyl groups) and CMP-Neu5Ac (specifically, 5 Å distance from the oxygen/nitrogen atoms of hydroxyl groups/amino groups), respectively. In the case of the α2,6PdST homology model, 78 residues were identified within a sphere of 12 Å radius from the C_α of Asp215, and similarly, 10 and 22 residues were identified in the binding sites of the lactose and CMP-Neu5Ac, respectively (Figure 3.1).

3.1.2 Multiple sequence alignment with GT family 80

To minimize the experiments for alanine scanning, conserved sequences among STs were analyzed and eliminated. Multiple sequence alignments for α2,3PST and α2,6PdST were done with the sequences of the GT family 80 retrieved by PSI-BLAST, such *as Pasteurella multocida* gi 90109232, *Pasteurella dagmatis* gi 260914426, *Photobacterium leiognathi* gi 267844791, *Photobacterium damselae* gi 269102115, *Photobacterium phosphoreum* gi 148537333, *Photobacterium* sp. JT-ISH-224 gi 159147903, *Vibrionaceae photobacterium* gi 178847432, *Vibrio* sp. JT-FAJ-16 gi 158455298, *Haemophilus ducreyi* gi 33151326, and *Shewanella piezotolerans* gi 212637736 (Yamamoto et al. 2008). A total of 27 conserved amino acid residues including catalytic key residues such as Asp141, H311, S355, and S356 in α2,3PST (Asp215, H387, S431, and S432 in α2,6PdST) were found (Figure 3.2). Considering that the conserved residues in a protein structure have high possibilities of showing functional or structural importance for

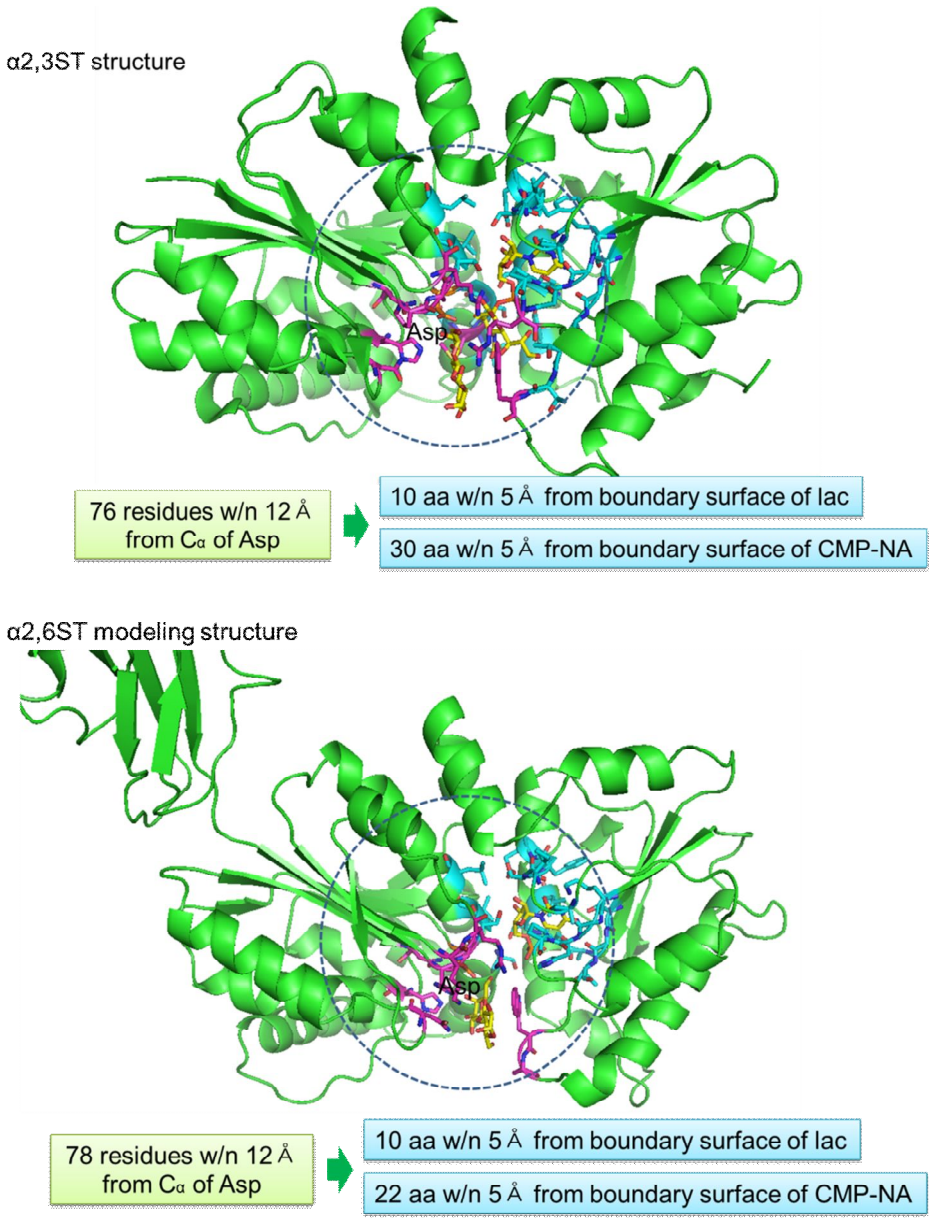


Figure 3.1 Selection of residues from substrate binding pocket

The residues within 5 Å distance from the boundary surface of lactose and CMP-NeuAc were selected for α2,3PST and α2,6PdST, respectively.

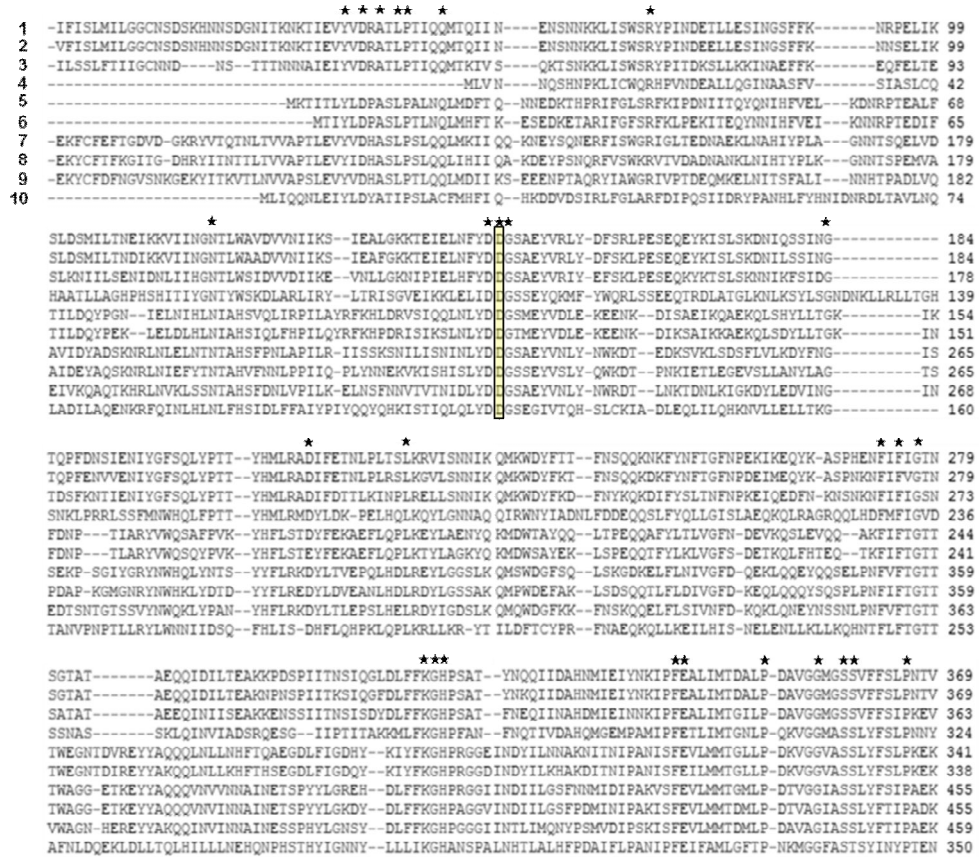


Figure 3.2 Multiple alignments of the amino acid sequences of GT family 80

The amino acid sequences from the GT family 80 were aligned by ClustalW2.1. The catalytic key residues Asp is boxed in black and conserved residues are labeled using the code.

its catalytic activity, they were primarily eliminated in the alanine scanning experiments. For example, H311, S355, and S356 in α2,3PST are conserved for the stabilization of CMP-Neu5Ac by holding the phosphate oxygen with hydrogen bonds (Kim et al. 2008; Ni et al. 2007). Similarly, the corresponding H387, S431, and S432 in α2,6PdST were identified as conserved amino acid residues that have the same function. Among the 10 and 30 residues interacting with the lactose and CMP-Neu5Ac in the binding pocket of α2,3PST, 5 (P34, N85, H112, M144, and R313 for lactose) and 12 (L40, S143, V147, T265, T267, T268, T269, P312, I335, S336, L357, and Y388 for CMP-Neu5Ac) non-conserved residues were selected for alanine scanning, respectively. Similarly, 5 (H106, K135, R136, N157, and W347 for lactose) and 6 (S108, L112, T344, I411, S412, and L433 for CMP-Neu5Ac) respective non-conserved residues in α2,6PdST were selected for alanine scanning.

3.1.3 Alanine scanning to determine the functional residues

Alanine scanning for all the selected residues was performed to determine the functional residues for site-saturation mutagenesis as the next step. The relative ST activity was compared using the pH-based colorimetric method mentioned in the material and methods section. Since the pK_a of cresol red and phenol red are 8.32 and 7.5, respectively, cresol red was better suited for α2,3PST assay, whose optimal pH was 8.5, and phenol red for α2,6PdST assay, whose optimal pH was 8.0. The relative activities (U/mL) and relative specific activities (i.e. relative activity/level of soluble protein expression relative to the wild-type) were determined.

3.1.3.1 Alanine scanning for α2,3PST

The relative activities (U/mL) and relative specific activities (i.e. relative activity/level of soluble protein expression relative to the wild-type, (Table 3.1 and Figure 3.3) for Ala-substituted α2,3PST mutants are shown in Table 3.1(A).

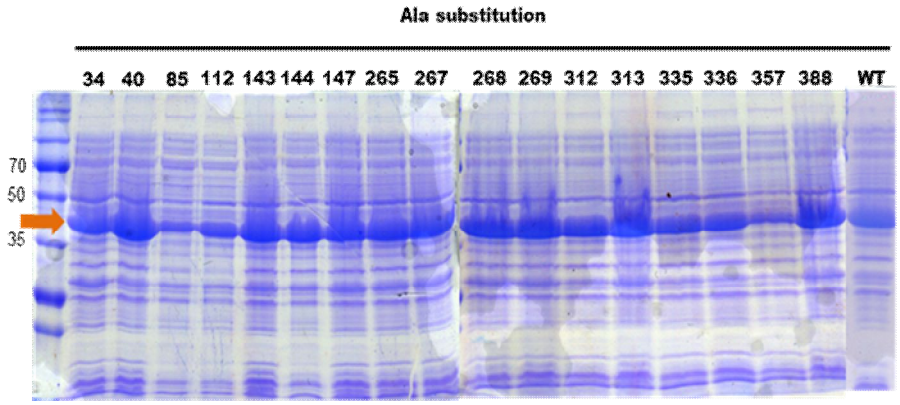
The soluble enzyme expression levels of the most α2,3PST mutants were about the same as that of wild-type α2,3PST, except for N85A and L357A mutants, which showed less than 30% of the wild-type levels (Table 3.2 and Figure 3.3). The N85A mutant appeared to produce improperly folded enzymes, and lost most of the α2,3 ST activity. In addition, despite having a reasonable expression level, H112A almost lost its activity, yielding 6% of the activity of the wild-type enzyme. Such remarkable losses in α2,3 ST activity suggest that N85 is an important key residue for proper protein folding and substrate binding, and H112 is a key residue involved in substrate binding and/or significantly modulating the catalytic efficiency of the enzyme. According to the structure of the α2,3PST-CMP-3F(a)Neu5Ac-lactose ternary complex, N85 appears to make water-mediated H-bonding interactions with the O1' and O5′ of the galactose moiety of the lactose, and Nε2 of His112 is likely to generate H-bonds with O3' and O4' of the galactose moiety. Thus, N85 and H112 were eliminated in the following mutation study. The L357 site was also excluded from the candidates of the study owing to a 29% of low soluble protein expression level compared to the wild-type despite its relative specific activity of 78%. In result, only the mutation sites yielding proper soluble expression with high relative specific activities were selected. Among the remaining mutation sites, the sites showing over 60% of the relative specific activity compared to the wild-type were chosen from the perspective of neutral

Table 3.1 Levels of soluble protein expression for α2,3PST/α2,6PdST alanine-substituted mutants relative to the wild-type

α2,3PST	Relative soluble protein level	α2,6PdST	Relative soluble protein level
WT	1.00	WT	1.00
P34A	0.84	H106A	0.24
L40A	1.43	S108A	0.10
N85A	0.27	L112A	1.45
H112A	0.69	K135A	0.54
S143A	1.02	R136A	0.34
M144A	0.82	N157A	1.63
V147A	0.84	T344A	2.84
T265A	0.78	W347A	0.28
T267A	1.31	I411A	0.97
T268A	1.46	S412A	1.06
T269A	1.06	L433A	1.20
P312A	1.03		
R313A	0.94		
I335A	0.81		
S336A	0.83		
L357A	0.29		
Y388A	1.11		

Levels of soluble protein expression were obtained by gel image analysis of the soluble fraction bands on SDS-PAGE presented in Figure 3.3.

A



B

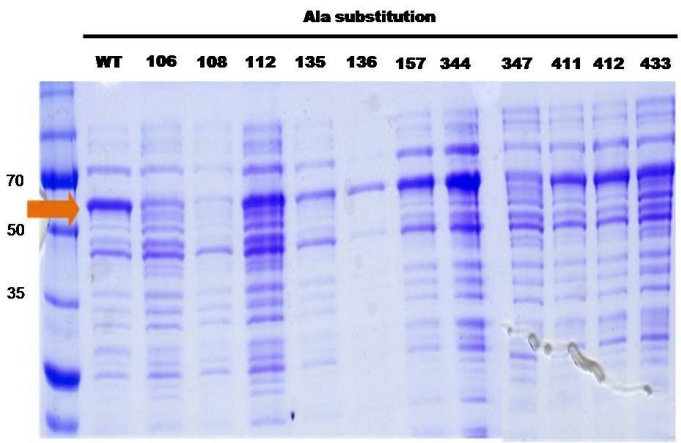


Figure 3.3 Levels of soluble protein expression for alanine-substituted mutants and wild-type of α2,3PST (A)/ α2,6PdST (B) by SDS-PAGE

The soluble protein of alanine-substituted enzymes was obtained by centrifugation at 22,250 × g for 30 min at 4 ºC after cell lysis. After the gel was stained by Coomassie Brilliant Blue R-250, the levels of the soluble protein expression for the mutants relative to the wild-type were calculated by the image analysis of the soluble fraction bands on SDS-PAGE.

Table 3.2 Relative activities and relative specific activities of alanine-substituted mutant enzymes for α2,3PST /α2,6PdST

(A) α2,3PST mutants	Relative activity (%)[a]	Relative specific activity (%)[b]	(B) α2,6PdST mutants	Relative activity (%)[a]	Relative specific activity (%)[b]
WT	100	100	WT	100	100
P34A	61	73	H106A	n.d.	n.d.
L40A	77	54	S108A	n.d.	n.d.
N85A	n.d.	n.d.	L112A	61	42
H112A	4	6	K135A	8	15
S143A	39	38	R136A	10	31
M144A	12	14	N157A	32	19
V147A	49	56	T344A	38	13
T265A	48	61	W347A	16	57
T267A	26	20	I411A	169	173
T268A	102	70	S412A	38	36
T269A	48	46	L433A	114	95
P312A	48	47			
R313A	107	113			
I335A	43	53			
S336A	17	21			
L357A	23	78			
Y388A	30	27			

Relative activity[a] for alanine-substituted mutants was determined by U/mL, and relative specific activity (%)[b] was defined as (U/mL of mutant/level of soluble protein expression of mutant relative to the WT)/(U/mL of WT/relative level of soluble protein expression of WT). The mean values are shown from duplicate experiments. n.d.: not detected.

drift, with P34, T265, T268, and R313 remaining for the sites for saturation mutagenesis.

3.1.3.2 Alanine scanning for α2,6PdST

Likewise, among the alanine-scanned α2,6PdST mutants, H106A, S108A, and W347A mutants showed less than 30% soluble expression levels compared to the wild-type α2,6PdST, and H106A and S108A showed no activities (Table 3.2(B)). The H106 residue is situated within an H-bonding distance from O4' and O5′ of the galactose moiety of the lactose, and appears to contribute to the binding of the lactose. The relative specific activities of L112A, N157A, and T344A mutants were less than 50%, even though their soluble protein expression levels were quite comparable or higher than that of the wild-type α2,6PdST. Among them, the N157 appears to make an H-bond with the O1 of the glucose moiety of the lactose, which corresponds to N85 in α2,3PST interacting with the galactose moiety of the lactose. As a result, I411 and L433 were chosen for the sites for site-saturation mutagenesis, since I411A and L433A mutants showed 173% and 95% of its relative specific activities, respectively, without changing their soluble protein expression levels.

3.2 Saturation mutagenesis for functional residues

3.2.1 Screening of α2,3PST variants

P34, T265, T268, and R313 residues in α2,3PST were individually subjected to saturation mutagenesis, and the mutants showing high activities were screened at pH 8.5 using the same colorimetric assay. In the case of R313 located on the loop

in the binding site of the glucose moiety of lactose, R313A and R313G replaced with small amino acids appeared to show essentially the same ST activity as the wild-type, but the mutants replaced with hydrophilic residues like Ser, Thr, Tyr, Asp, His, and Asn showed at least 1.5-times higher activities than the wild-type (Table 3.3). The relative specific activities of the R313T, R313Y, R313D, R313H, and R313N mutants were 129, 125, 108, 146, and 231% compared to the wild-type, respectively, indicating that the R313N mutant showed the highest specific activity (Table 3.4 (A)). In the case of T265 located on the loop at a 4.7 Å distance from the N atom of NH2 of cytidine in CMP-Neu5Ac, the T265G, T265S, and T265N mutants were selected (Table S2), and their relative specific activities were 94, 116, and 126%, respectively (Table 3.4 (A)). Despite a slightly low specific activity of T265G, the mutant showed the highest overall activity (i.e. 168%) due to its higher expression level. On the other hand, no mutant at P34 and T268 sites showed higher specific activities than its wild-type (below 90%), even though the protein expression levels of the mutants were similar to that of the wild-type (data not shown).

3.2.2 Combinations of the single mutations of α2,3PST

To identify any synergistic combinations of the single mutations, a total of 7 double mutants were constructed by site-directed mutagenesis based on the high specific activities of R313N, R313H, R313Y, and R313T, as well as T265N and T265S variants (the R313T/T265N double mutant was not constructed, but its activity would be lower than that of the R313T/T265S mutant according to the data shown in Table 3.5).

Table 3.3 Relative activities of R313 and T265 single mutants of α2,3PST

α2,3PST mutants	Relative activity (%)[a]		Relative activity (%)[a]
WT	100		
R313 variants		T265 variants	
G	134	G	168
S	167	V	78
T	171	S	125
Y	179	N	86
N	150		
D	166		
H	233		
P	94		
I	52		

[a] Relative activities of R313 and T265 mutants for α2,3PST were determined in U/mL and compared to the wild-type. Reactions were carried out using colorimetric method with 4 mM CMP-Neu5Ac, 4 mM lactose, 0.2 mM dye (cresol-red) and 10 μL of enzyme of cell-extract in 5 mM Tris-HCl pH 8.5 buffer.

Table 3.4 Relative specific activities for various single and double mutants of α2,3PST and α2,6PdST

(A)	Enzymes (α2,3PST)	Relative Specific Ac (%)[a]	(B)	Enzymes (α2,6PdST)	Relative Specific Ac (%)[a]
	WT	100 ± 8		WT	100 ± 4
	R313N	231 ± 29		I411T	198 ± 4
	R313H	146 ± 9		L433S	289 ± 8
	R313T	129 ± 21		L433T	296 ± 1
	R313Y	125 ± 17		I411T/L433S	194 ± 1
	R313D	108 ± 13		I411T/L433T	510 ± 15
	T265N	126 ± 11			
	T265S	116 ± 2			
	T265G	94 ± 3			
	R313N/T265S	216 ± 10			
	R313H/T265S	237 ± 7			

[a] Relative specific activities for α2,3PST and α2,6PdST variants were compared with those of parent enzymes corresponding to 22.71 U/mg and 3.67 U/mg, respectively. Reactions were carried out using colorimetric method with 4 mM CMP-Neu5Ac, 4 mM lactose, 0.2 mM dye (cresol-red/phenol-red) and 0.45 μg (α2,3PST) / 2.5 μg (α2,6PdST) of purified enzyme in 5 mM Tris-HCl (pH 8.5/pH 8.0) buffer. Values are means with standard deviations of results from three experiments.

Table 3.5 Relative activities of R313/T265 double mutants of α2,3PST

α2,3PST mutants	Relative activity (%)[a]
WT	100
R313/T265 variants	
R313H/T265N	75
R313H/T265S	173
R313Y/T265N	57
R313Y/T265S	162
R313N/T265N	156
R313N/T265S	247
R313T/T265S	105

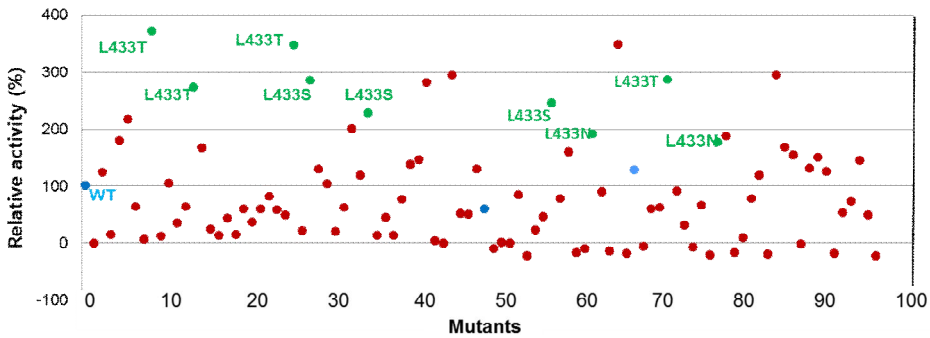
[a] Relative activities of R313/T265 double mutants for α2,3PST were determined in U/mL and compared to wild-type. Reactions were carried out using colorimetric method with 4 mM CMP-Neu5Ac, 4 mM lactose, 0.2 mM dye (cresol-red) and 10 μL of enzyme of cell-extract in 5 mM Tris-HCl pH 8.5 buffer.

The protein expression levels of all the double mutants were quite similar to that of the wild-type, and notably, only the relative activities of R313H/T265S and R313N/T265S were higher than those of the other double mutants (Table 3.5). The specific activities of R313H/T265S and R313N/T265S mutants were 237% and 216%, respectively, compared to that of the wild-type (Table 3.4).

3.2.3 Screening of α2,6PdST variants

In the case of α2,6PdST, I411 and L433 were individually subjected to saturation mutagenesis. According to the modeling structure of α2,6PdST, I411 is at an H-bonding distance of 3.4 Å from the O2 of the cytosine ring, and L433 is located in a deep cleft region for the CMP binding site by making a H-bond with an O3' atom of the ribose ring of CMP. Among the L433 variants, L433S and L433T mutants were screened to have higher activity than wild-type (Figure 3.4). It was revealed that L433S and L433T mutants showed ca. 3-fold improvement in the specific activities (Table 3.4 (B)). Notably, among the I411 variants screened, only I411T mutant was screened (Figure 3.4) to have the highest activity. The mutant showed increase in the activity, but also an enhanced expression level of its soluble protein compared to the wild-type α2,6PdST, which was also confirmed by SDS-PAGE (Figure 3.5). In addition, when the expression vector was changed into pET28a, expression levels of wild-type and I411T were enhanced. As a result, I411T mutant showed 2-fold improvement in the specific activity (Table 3.4).

A



B

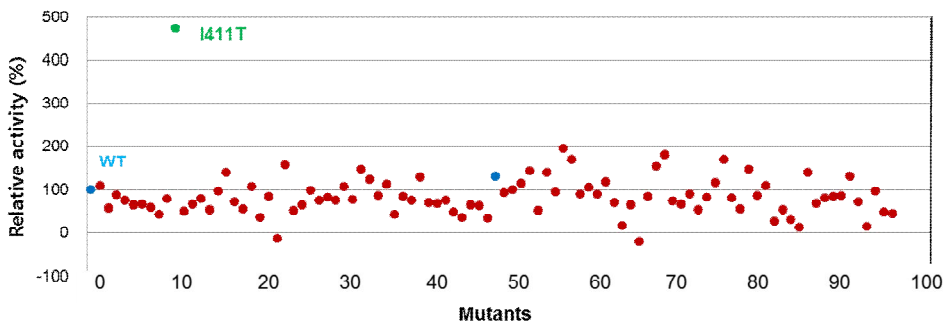


Figure 3.4 Screening of L433 and I411 variants generated by saturation mutagenesis of α2,6PdST

(A) L433S and L433T showed 2-3-fold higher activity than that of wild-type (blue dot). (B) I411T showed the highest activity, that is more than 4-fold increase in activity than wild-type.

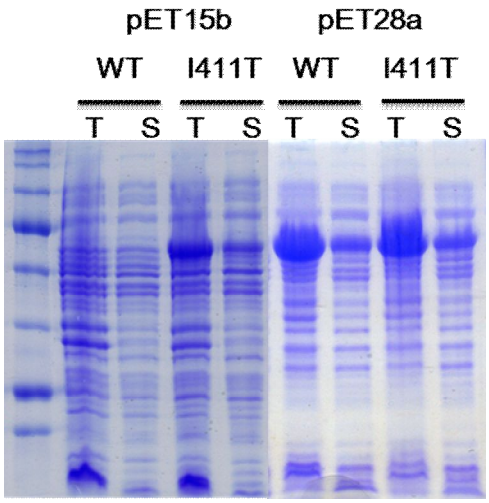


Figure 3.5 Protein expression of wild-type and I411T mutant of α2,6PdST according to the expression vector

In the case of pET15b harboring wild-type and I411T, protein expression level of I411T was much higher than wild-type. Protein expression levels of wild-type and I411I were increased by changing the expression vector into pET28a.

3.2.4 Combinations of the single mutations of α2,6PdST

To identify any synergistic combinations of the single mutations, two double mutants, I411T/L433S and I411T/L433T, were constructed. As a result, I411T/ L433S and I411T/L433T mutants showed 1.9- and 5.1-fold increase in the specific activity compared with the wild-type (Table 3.4 (B)). The result showed that the double mutation I411T/L433T showed an additive effect.

3.3 Characterization of α2,3PST mutants

3.3.1 Determination of kinetic parameters

To understand underlying mechanisms of the individual mutations of α2,3PST to the changes in specific activities, the kinetic parameters of the single mutants and double mutants were further evaluated using the purified enzymes (Table 3.6 (A) and Figure 3.6). R313N and R313H single mutations enhanced k_{cat} values for CMP-Neu5Ac (ca. 2-fold increase) and lactose (ca. 40-70% increase), whereas the mutations slightly decreased their substrate binding affinities for CMP-Neu5Ac (ca. 40-80% K_m value increase) and lactose (ca. 10% K_m increase). In the results, the k_{cat}/K_m values for CMP-Neu5Ac of the R313N and R313H mutations were increased by ca. 32% and 20%, and those for lactose were enhanced by 55% and 33%, respectively. In the case of T265S mutation, the k_{cat}/K_m value was increased by ca. 30% for CMP-Neu5Ac due to the increase in the k_{cat} value, but no significant changes were shown in k_{cat} and K_m values for lactose.

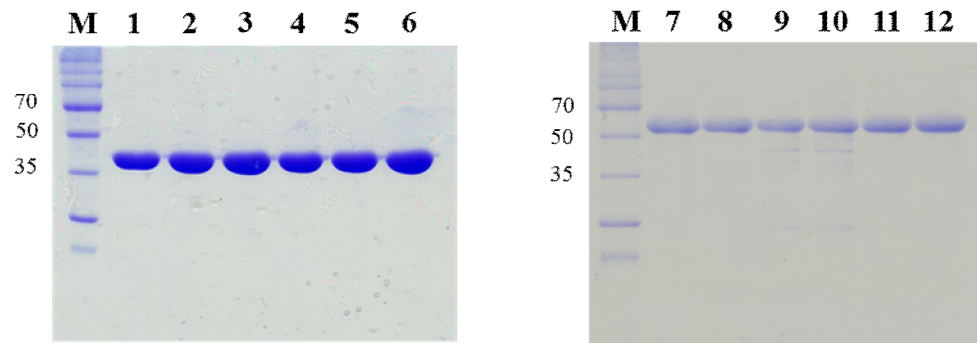


Figure 3.6 SDS-PAGE of purified wild-type and mutant enzymes of α2,3PST and α2,6PdST

M, marker; 1-6, α2,3PST wild-type and variants; 1, wild-type; 2, R313N; 3, R313H; 4, T265S; 5, R313N/T265S; 6, R313H/T265S; 7-12, α2,6PdST wild-type and variants; 7, wild-type; 8, I411T; 9, L433S; 10, L433T; 11, I411T/L433S; 12, I411T/L433T

Table 3.6 Kinetic parameters of wild-type and mutants for α2,3PST

Mutants of α2,3PST	CMP-Neu5Ac[a] (donor)			Lac[b] (acceptor)		
	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
WT	12.6±0.35	1.73±0.10	7.27±0.22	53.2±4.25	2.36±0.05	22.6±2.28
R313N	23.3±0.60	2.42±0.09	9.63±0.11	89.7±0.20	2.55±0.01	35.1±0.08
R313H	26.6±0.20	3.14±0.11	8.49±0.22	75.4±5.15	2.51±0.18	30.1±0.05
T265S	22.2±0.05	2.32±0.09	9.56±0.35	58.6±4.65	2.48±0.05	23.7±2.35
R313N/T265S	36.0±0.90	2.91±0.42	12.6±1.51	82.1±0.35	2.14±0.07	38.4±1.09
R313H/T265S	44.6±1.15	4.09±0.15	10.9±0.11	73.9±0.01	1.64±0.01	45.0±0.28

a Kinetic measurements at 10 mM lactose and variable CMP-Neu5Ac concentrations (0.25-20 mM).

b Kinetic measurements at 20 mM CMP-Neu5Ac and variable lactose concentrations (0.03-150 mM).

The kinetic assays were performed in duplicate, and the kinetic parameters were obtained by nonlinear regression analysis of the Michaelis-Menten equation.

The kinetic parameters of T265N single mutant were not determined, because the mutation showed relatively lower activity than T265S when combined with R313H /Y/N mutations in the double mutants as shown in Table 3.5.

In general, both double mutants (R313H/T265S and R313N/T265S) had significantly improved k_{cat} values for both CMP-Neu5Ac (ca. 3 to 3.5-fold increase) and lactose (ca. 50% increase). However, their K_m values for CMP-Neu5Ac and lactose were somewhat increased and decreased, respectively, indicating the reduced binding affinity for CMP-Neu5Ac donor, but the enhanced binding affinity for lactose acceptor. Overall, by varying the concentration of CMP-Neu5Ac or lactose, the k_{cat}/K_m values of both the R313N/T265S and R313H/T265S mutants were identified as increased by a factor of 1.7 and 1.5 for CMP-Neu5Ac, whereas 1.7- and 2-fold increases were observed for lactose compared to that of the wild-type, respectively.

3.3.2 Confirmation of side-reaction activity for 6′-SL synthesis

The 6′-SL synthesis activities of α2,3PST for R313 single and double mutants were compared to that of the wild-type at below pH 7.0 (Figure 3.7). Interestingly, the R313 single mutants (R313N, R313T, R313Y, R313H, and R313D) and the double mutants R313N/T265S and R313H/T265S showed very low α2,6 activity at pH 6.0 and pH 6.5, although wild-type α2,3PST was known to produce 3-4% 6′-SL among the overall products. When the reactions were conducted with the purified enzymes (2.25 μg, 0.05 U) for 30 min at pH 6.0, the production yields of 6′-SL of the R313 single and double mutants decreased by a factor of 4 to 30 compared to that of the wild-type. Notably, all the R313 single and double mutants did not

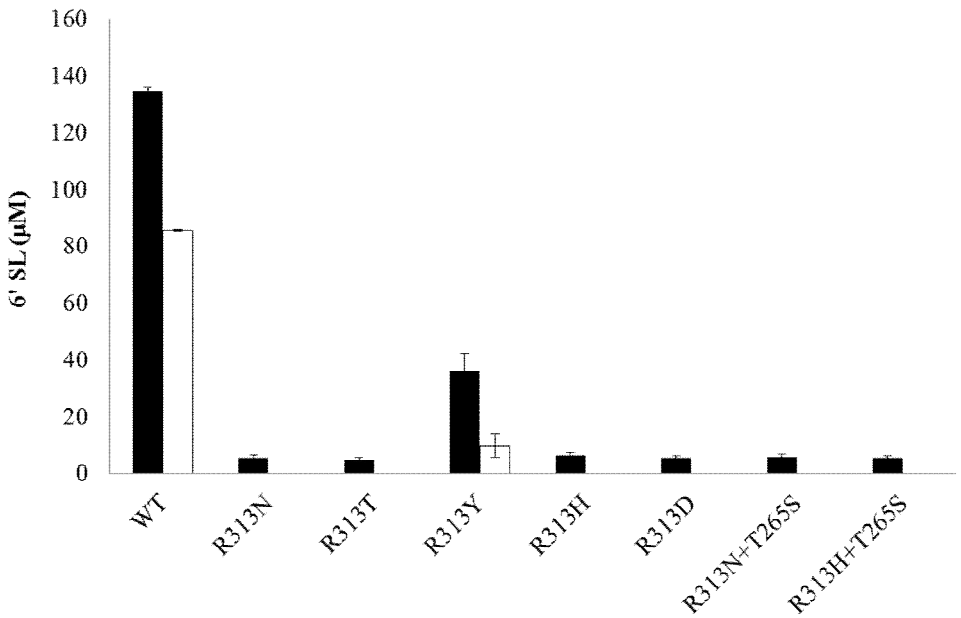


Figure 3.7 SDS-PAGE of purified wild-type and mutant enzymes of α2,3PST and α2,6PdST

Reactions were performed with 10 mM CMP-Neu5Ac and 5 mM lactose in 50 mM MES buffer pH 6.0 and pH 6.5 at room temperature using 2.25 μg of purified enzymes. After 30 min, the 3′-SL and 6′-SL produced were quantified by Bio-LC. Under this condition, 3′-SL were produced for saturated state (ca. 95% conversion yield). Black and blank bars indicate the 6′-SL (μM) produced at pH 6.0 and pH 6.5, respectively. Assays were performed in duplicate.

produce 6′-SL at all at pH 6.5, except for R313Y. The result demonstrates that R313 is a key amino acid residue for reducing the α2,6 ST side-reaction activity of α2,3PST.

3.3.3 Confirmation of sialidase activity

Interestingly, the R313Y mutant was recently shown to decrease the sialidase activity of α2,3PST at pH 5.5 (Sugiarto et al.). Additionally, we have measured the α2,3 sialidase activities of R313 single and of double mutants for 3′-SL with the excess amount of the purified enzymes (45 μg, which is 100-fold higher concentration of the enzyme used in the α2,3 ST reaction) at pH 5.5 (50 mM MES buffer). Under this condition, α2,3 sialidase activities of all the mutants decreased compared to that of the wild-type. For examples, R313N, R313H, R313Y, R313D, R313T, R313N+T265S, and R313H+T265S mutants produced 13.1%, 7.8%, 3.3%, 3.6%, 13.5%, 1.4%, and 1.9% of sialic acid from 1 mM of 3′-SL in 1h, respectively, whereas 19.6% of sialic acid was produced by the wild-type (Figure 3.8). This result indicates that α2,3 sialidase activity of the mutants decreased significantly, and 7 to 69% of their original sialidase activities remained intact in the mutants. In the case of R313N+T265S mutant, 93% of its α2,3 sialidase activity decreased, suggesting that the decrease in the sialidase activity would be one of the major causes of high ST specific activity (2-fold, shown at Table II) of the mutant.

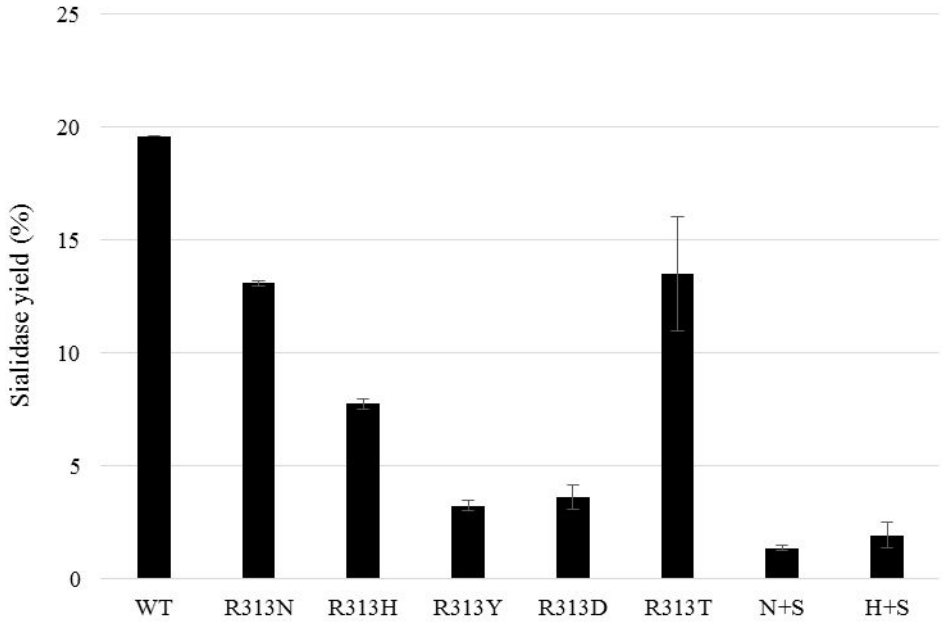


Figure 3.8 Sialidase activities of α2,3PST variants

Reactions were performed with 1 mM 3-SL in 50 mM MES buffer pH 5.5 at room temperature using 45 μg of purified enzymes. After 1 h, sialic acids produced by sialidase activity were quantified by Bio-LC. Assays were performed in duplicate.

3.3.4 Analysis of mutants model structure

In the case of α2,3PST, only the R313, which is present in the loop region making an H-bond with glucose in the 2IHZ structure (α2,3PST-CMP-3F(*a*)Neu5Ac-lactose) (Ni et al. 2007), was able to generate interesting and highly active mutants. Its replacements with hydrophilic residues such as T/Y/D/H/N rather than hydrophobic or small amino acid residues displayed somewhat higher ST activity. The 2ILV structure (α2,3PST-CMP-lactose ternary structure) superimposed with 2IHK (α2,3PST-CMP-3F(*e*)Neu5Ac binary structure) gives useful information about Neu5Ac and lactose coordinates for the α2,3 ST reaction, such as indicating a very close positioning of the galactose O3' of the lactose to the anomeric carbon (C2) of sialic acid (Figure 2A) (Ni et al. 2007). According to the structure, $N\eta_1$ and $N\eta_2$ of R63 located at 3 Å from the carboxylic acid of sialic acid helps to anchor the Neu5Ac during its distortion along the reaction coordinate. While the $N\eta_1$ of R313 is 4.9 Å away from the $N\eta_2$ of R63, R313N mutation decreases the distance (corresponding to the distance from $N\delta_2$ of N313 to $N\eta_2$ of R63) to 2.63 Å. In the R313H mutant, $N\varepsilon_2$ of H313 could also make H-bonds with the $N\eta_1$/$N\eta_2$ atoms of R63. This calculation suggests that a more stable structure of Neu5Ac is formed by R313N and R313H mutations by generating more reactive conformation of α2,3PST through anchoring the Neu5Ac in close proximity to the lactose. In addition, the reoriented hydroxyl group generated by T265S mutation is likely to result in H-bonding (2.87 Å) to the N4 atom of the cytosine ring of CMP, whereas the distance between $O\gamma_1$ of T265 and CMP is 4.87 Å in the 2ILV structure of closed conformation (Figure 3.9A).

A **B**

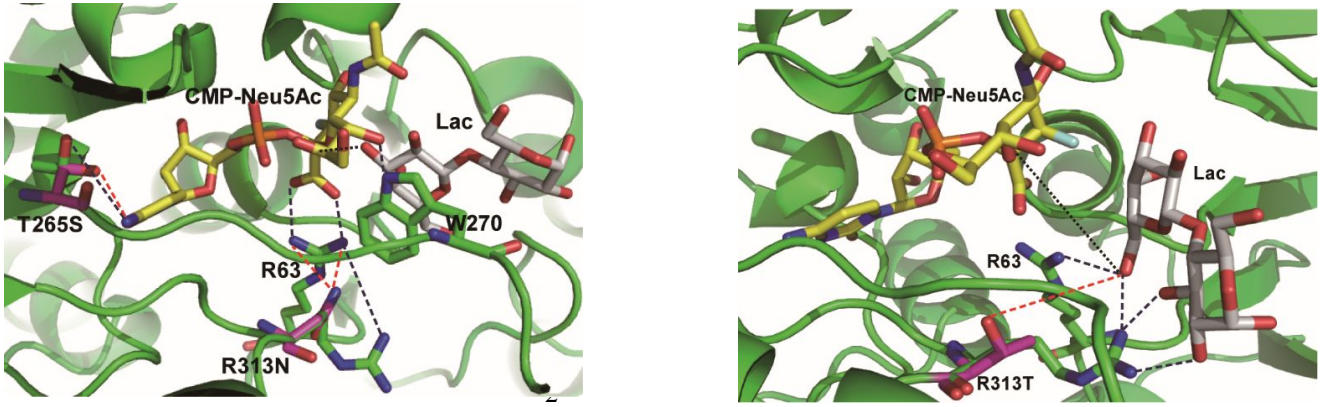


Figure 3.9 α2,3PST structure of probable productive conformation for the α2,3 ST/α2,6 ST activity

(A) The structure of a probable productive conformation for the α2,3 ST activity of α2,3PST is shown with 2ILV (α2,3PST-CMP-lactose) and superimposed 2IHK (α2,3PST-CMP-3F(e)Neu5Ac). Mutations for R313N and T265S are presented by magenta color with original residues (green). (B) The structure of a probable productive conformation for the α2,6 ST activity of α2,3PST is shown with 2IHZ (α2,3PST-CMP-3F(a)Neu5Ac-lactose). R313T showing a lowered α2,6 ST activity is presented by magenta color with original residue (green). CMP-Neu5Ac and lactose are shown according to the crystal structures. The distances of glycosidic bond are drawn with black close dashes. And the distance observed from original residues are drawn in blue dashes, whereas the distances generated from mutated residues are drawn in red dashes

This interaction might help to stabilize the CMP in more active conformation, and trigger the movement of the loop near Neu5Ac like a movement through interaction between G266 and N4 of CMP (Ni et al. 2007). The movement could help to define the acceptor binding pocket by W270 located on the loop. This additive T265S mutation to R313N/H significantly increased the k_{cat} values for CMP-Neu5Ac, and slightly increased the binding affinity of lactose (Table **3.6(**A)).

In terms of the α2,6 ST side reaction activities of α2,3PST, all the R313 single variants (N, T, Y, H, D) and double variants (R313N/T265S, R313H/T265S) showed significantly low levels of α2,6 ST activities compared to that of the wild-type (Figure 3.7). Notably, none of the mutants showed α2,6 ST side reaction activity at pH 6.5, except for R313Y. The 2IHZ structure (α2,3PST-CMP-3F(*a*)Neu5Ac-lactose ternary structure) indicates alternative lactose binding conformation for α2,6 ST activity at low pH (Ni et al. 2007). In the structure, R313 can make an H-bond to the O6' of galactose with a distance of 2.84 Å (Figure 2B), suggesting that the H-bonding would help the O6' atom of the galactose to situate toward the C2 of sialic acid. When the R313N/T/Y/H/D variants were compared with the wild-type α2,3PST, the distances from O6' of the galactose to the replaced amino acids were much longer, and the longest distance (6.4 Å) was found with R313T, which shows the lowest α2,6ST activity. R313 residue also appears to bind the O2 and O3 of the glucose moiety of the lactose in the 2IHZ structure, leading the O6' of the galactose moiety of lactose to come closer to the anomeric carbon (C2) of Neu5Ac (Figure 3.9B). In the case of other R313 mutants, the distance between the replaced residue at the 313 site and glucose becomes much longer, suggesting that R313 is the key determinant residue to generate the side reaction of

α2,6 ST activity. When α2,3PST and α2,6PdST were superimposed, their overall structures were quite similar, with minor conformational differences, including the lactose orientation. The R313 in α2,3PST appeared to correspond to the R136 in α2,6PdST. In the results on α2,6PdST, the relative specific activity of the R136A mutant was only 30% that of its wild-type (Table 3.2(B)), indicating again that the residue might contribute to the orientation of the lactose acceptor required for its α2,6 ST activity.

3.4 Characterization of α2,6PdST mutants

3.4.1 Determination of kinetic parameters

The kinetic parameters of the single mutants I411T, L433S, and L433T of α2,6PdST were determined, as shown in Table 3.7 (B). All the single mutants showed significantly enhanced k_{cat} values for CMP-Neu5Ac (ca. 2.5 to 4.7-fold increase) and lactose (ca. 5.9 to 20-fold increase), whereas the mutations decreased their substrate binding affinities for CMP-Neu5Ac (ca. 3-40% increase in K_m value) and lactose (ca. 3.6 to 8-fold increase in K_m value). However, unlike others, only the L433T mutation increased the substrate binding affinity for CMP-Neu5Ac by about 2.5-fold. In the results, the k_{cat}/K_m values for CMP-Neu5Ac of the I411T, L433S, and L433T mutants were increased by 2.4, 3.5, and 6.4-fold, and those for lactose were also enhanced by 1.6, 2.5, and 2.2-fold, respectively. To identify any synergistic combinations of the single mutations, two double mutants, I411T/L433S and I411T/L433T, were constructed.

Table 3.7 Kinetic parameters of wild-type and mutants for α2,6PdST

Mutants of α2,6PdST	CMP-Neu5Ac[a] (donor)			Lac[b] (acceptor)		
	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
WT	4.32±0.33	5.87±0.40	0.74±0.01	4.30±0.16	7.62±1.41	0.58±0.09
I411T	10.9±0.25	6.06±0.30	1.80±0.05	25.4±0.35	27.3±1.15	0.93±0.03
L433S	20.3±2.00	8.15±1.39	2.61±0.69	85.0±0.45	59.7±0.55	1.43±0.03
L433T	10.7±0.25	2.27±0.16	4.70±0.20	76.5±3.25	61.1±4.35	1.26±0.04
I411T/L433S	16.9±0.55	10.7±1.16	1.58±0.11	102±1.00	75.7±2.85	1.35±0.04
I411T/L433T	18.8±0.65	3.37±0.28	5.60±0.26	110±2.00	62.0±4.95	1.78±0.11

[a] Kinetic measurements at 10 mM lactose and variable CMP-Neu5Ac concentrations (0.25-20 mM).

[b] Kinetic measurements at 20 mM CMP-Neu5Ac and variable lactose concentrations (0.03-150 mM).

Both of the double mutants had increased k_{cat} values for CMP-Neu5Ac by 3.9 to 4.4-fold, whereas the CMP-Neu5Ac binding affinity of the I411T/L433S mutation decreased by 1.8-fold, but the I411T/L433T mutation had increased CMP-Neu5Ac binding affinity by about 57%. In all, I411T/L433S and I411T/L433T had increased k_{cat}/K_m values for CMP-Neu5Ac by 2- and 7.5-fold, respectively. Surprisingly, the k_{cat} values for lactose of the I411T/L433S and I411T/L433T mutations were 24 and 26-fold higher than that of the wild-type, while their K_m values for lactose increased 9.9 and 8.1-fold, respectively. The k_{cat}/K_m values for lactose of I411T/L433S and I411T/L433T increased 2.3 and 3.1-fold relative to that of the wild-type. The I411T/L433T double mutation displayed further improvement in activity than single mutations mainly by increasing the k_{cat} values, which resulted in an increase (ca. 20-40%) of the k_{cat}/K_m values for both substrates compared to its single mutant L433T.

3.4.2 Analysis of mutants model structure

In the case of α2,6PdST, the I411 located on a bent loop toward CMP is within H-bonding distance (3.4 Å) of the N4 atom of the cytosine ring of CMP (Figure 3. 10A). When I411 was replaced by modeling with Thr, a hydrophilic amino acid, the H-bond distance between the carbonyl oxygen of T411 and the cytosine ring was not changed, but a possible H-bond was newly generated between P408 and T411. Therefore, I411T mutation may reduce the flexibility of the bent loop containing P408 and T411 by H-bonding, and bring A409/K410 onto the rigid loop and cytosine ring closer, which allows less room for CMP-Neu5Ac to move around in the active site, and perhaps pushes CMP-Neu5Ac towards the lactose, resulting

A B

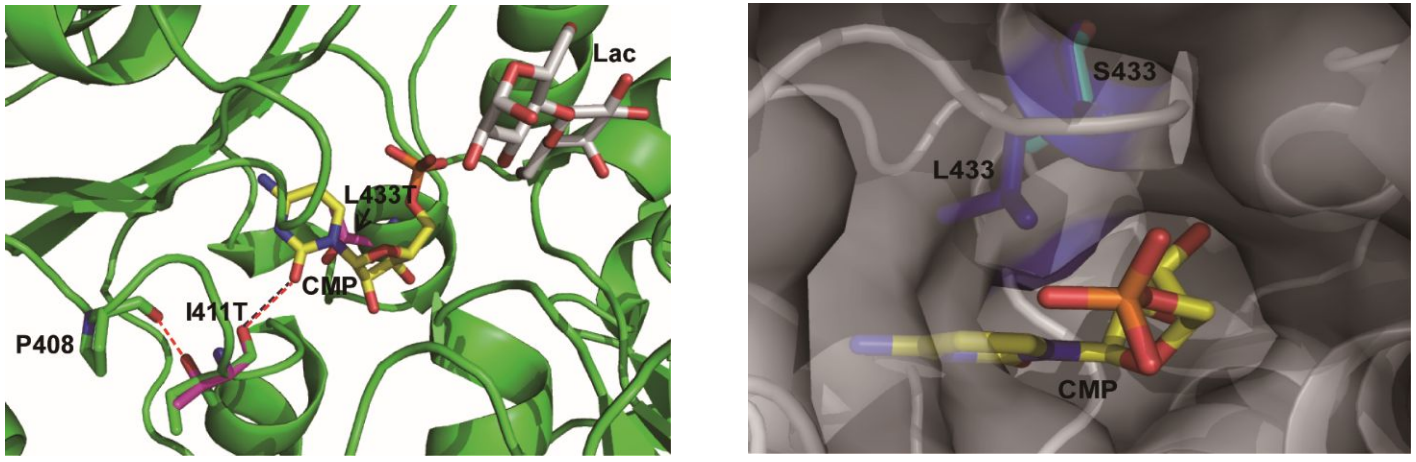


Figure 3.10 Modeling structure of α2,6PdST superimposed with mutated enzyme

(A) The modeling structure of I411T/L433T mutant of α2,6PdST is shown with CMP and lactose. Mutations for I411T and L433T are presented by magenta color with original residue (green). (B) The modeling structure of L433S mutant of α2,6PdST is shown as a boundary surface form with bound CMP. S433 is presented by cyan color with original wild-type residue (L433) (blue). The distance observed from original residues are drawn in blue dashes, whereas the distances generated from mutated residues are drawn in red dashes.

in a two-fold increase in the k_{cat}/K_m of the enzyme.

L433 of α2,6PdST is located on a short alpha-helix very close to the CMP-Neu5Ac binding site in the deep cleft of the active site and generates an H-bond (3.77 Å) between the main chain of L433 and O3' of CMP. When L433 was replaced with S or T, the distance of the H-bond was not changed much, but the enzyme surface for the CMP binding site was changed and the substrate access channel became more open (Figure 3.10 B). When the L433 residue and CMP-Neu5Ac were represented using a sphere form, structural collisions occurred between L433 and the cytidine ring. Whereas, L433S or L433T mutation showed no collision with CMP-Neu5Ac, and the enlarged space for the CMP binding would allow for easy access of CMP-Neu5Ac, resulting in an increased k_{cat} value (ca. 2.5 to 4.7-fold) for CMP-Neu5Ac. Therefore, the double mutation I411T/L433T showed an additive effect by enhancing the k_{cat}/K_m value by 7.5-fold relative to the wild-type for CMP-Neu5Ac.

According to the current model, however, both I411 and L433 are situated inside the substrate binding pocket to make more direct contact with CMP-Neu5Ac rather than lactose, so that the two residues appear not to affect the direct lactose binding. Therefore, in the case of lactose, it is quite hard to explain the mechanisms using the current model of the 25-fold increase in the k_{cat} value for I411T/L433S and I411T/L433T, and the dramatic increases in K_m (ca. 3.5-fold and 8-fold in I411T and L433S/T mutants, respectively).

3.5 *in vitro* production of SLs using the STs mutants

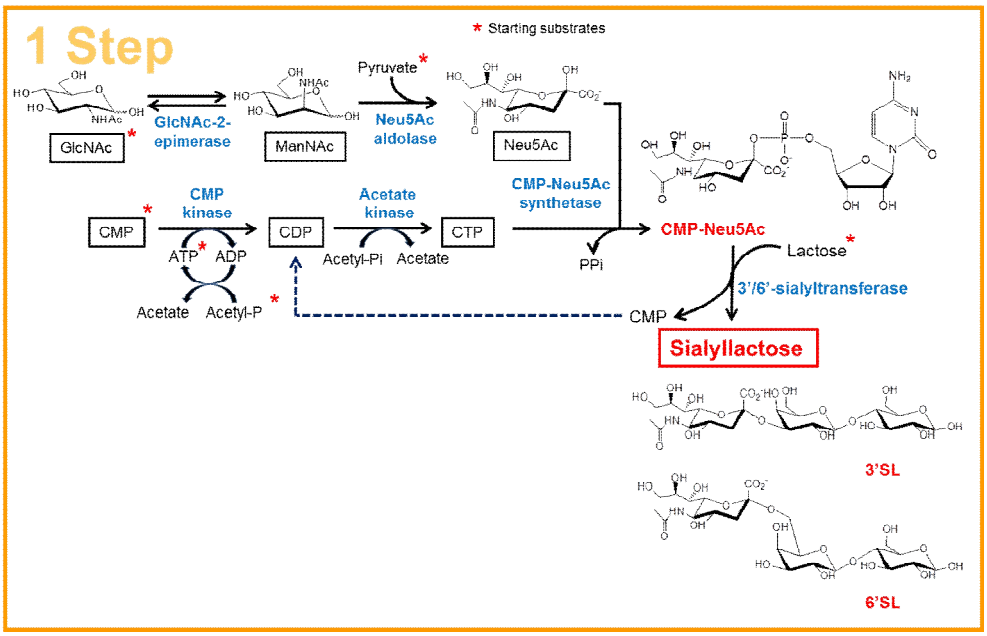
3.5.1 3′-SL production

When the α2,3PST R313N mutant with enhanced α2,3 ST activity and eliminated α2,6 ST side activity was used for 3′-SL production, one-pot reaction combining CMP-Neu5Ac and 3′-SL synthesis was accomplished at pH 7.0 (Figure 3.11). Using the substrates such as *N*-acetyl glucosamine, pyruvate, CMP, ATP, Acetyl-phosphate, and lactose, 40 g/L of 3′-SL was produced in 12 h at 7 L scale (Woo J.S. 2013), that is considerable increase in 3′-SL yield when compared with 23-28 g/L of 3′-SL production by the wild-type α2,3PST. In addition, one-pot reaction allowed to recycle of CMP for the CDP production, which resulted in only 20% of CMP relative to original amount, could be used for the reaction. As the amounts of substrates and enzyme used for the reaction were reduced, the cost of materials was reduced for 50%. Purification costs of the product were also reduced by suppressing the generation of 6′-SL by-product through utilization of α2,3PST mutant.

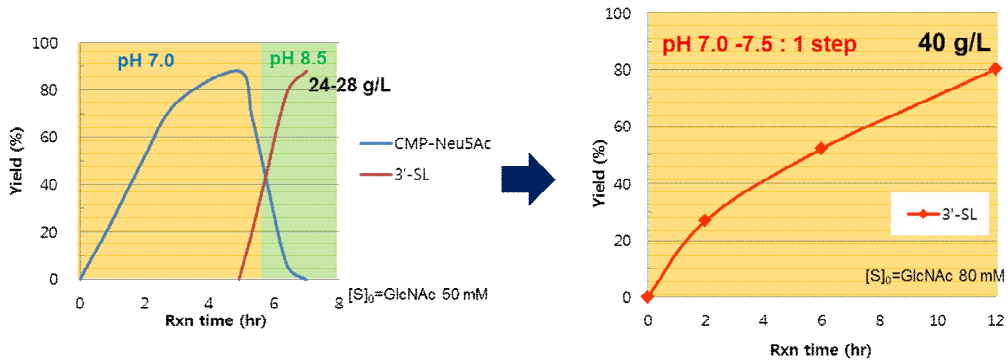
3.5.2 6′-SL production

6′-SL production was proceeded by one-pot reaction combining CMP-Neu5Ac and 6′-SL synthesis under the same reaction condition mentioned above. Using the L433S mutant of α2,6PdST with improved catalytic activity, 6′-SL production yield was improved for 40 g/L in 10 h (Woo J.S. 2013). As the 3′-SL production, CMP was recycled and the cost of substrates and enzyme was also reduced.

A


1 Step
* Starting substrates
GlcNAc
GlcNAc-2-epimerase
ManNAc
Pyruvate
Neu5Ac aldolase
Neu5Ac
CMP
CMP kinase
ATP
ADP
CDP
Acetate kinase
Acetyl-Pi
Acetate
CTP
CMP-Neu5Ac synthetase
PPi
CMP-Neu5Ac
Lactose
3'/6'-sialyltransferase
CMP
Acetate
Acetyl-P
Sialyllactose
3'SL
6'SL

B


pH 7.0
pH 8.5
24-28 g/L
CMP-Neu5Ac
3'-SL
Yield (%)
Rxn time (hr)
[S]0=GlcNAc 50 mM
pH 7.0 -7.5 : 1 step
40 g/L
3'-SL
[S]0=GlcNAc 80 mM

C

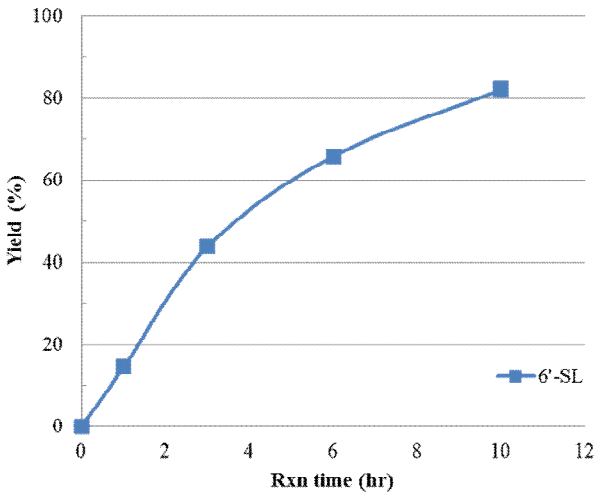


Figure 3.11 Improvement of production process and SLs yield

(A) One-pot reaction of 3'-/6′-SL production using GlcNAc, CMP, ATP, Acetyl-P, pyruvate, and lactose substrates through the utilization of α2,3PST/α2,6PdST variant. (B) 3′-SL yield increased from 24 g/L to 40 g/L in 12 h, coinciding with the reduction of costs for materials and product purification. (C) 6′-SL yield increased up to 40 g/L in 10 h.

3.6 Discussion

We have attempted to engineer two STs, α2,3PST and α2,6PdST, using a hybrid approach to increase catalytic activity for the production of 3'-and 6′-SL. Functional residues contributing to the catalytic activity and the expression level of soluble protein of the two STs were selected by combining computer modeling of the ST structures and alanine scanning experiments, after excluding the conserved sequences extracted from multiple alignments. Through the alanine scanning, the mutations displaying complete loss of ST activity or low ST activity were eliminated, and the functional candidate residues for site-saturation mutagenesis were identified.

Saturation mutagenesis of the functional residues could generate the mutants with enhanced activity and reduced side activity. In the case of α2,3PST, mutation of R313 was able to generate interesting and highly active mutants. Its replacements with hydrophilic residues such as T/Y/D/H/N rather than hydrophobic or small amino acid residues displayed somewhat higher ST activity. Especially, it is revealed that the R313 is a key residue to control the multifunction of α2,3PST. According to the results, we identified the R313 residue could contribute to the orientation of the lactose acceptor required for its α2,6 ST activity as well as sialidase activity. In the case of R313N+T265S mutant, 93% of its α2,3 sialidase activity decreased, suggesting that the decrease in the sialidase activity would be one of the major causes of high ST specific activity of the mutant. However, the measurement of the reduced α2,3 sialidase activities was performed at a different reaction condition compared with α2,3 ST reaction. The enzyme concentration was 100-fold higher and the reaction time was 6 times longer than

those of α2,3 ST reaction (In the case of α2,3 ST activity measurement, reaction was proceeded with ca. 0.45 μg of the purified enzyme in 5 mM Tris-HCl buffer at pH 8.5 for 10 min.). All together, the α2,3 sialidase activity of R313N+T265S mutant was ca. 600 times lower than its α2,3 ST activity. Therefore, the effect of the decrease in α2,3 sialidase activity of the mutants on the increase in α2,3 ST activity was negligible as we have expected. Nevertheless, our experimental results showed that R313 residue is also somewhat involved in controlling the α2,3 sialidase activity of wild-type α2,3PST, suggesting that R313 is a crucial residue to control the multifunction of α2,3PST.

All of the high-performing R313 mutants of α2,3PST, such as R313N, R313N/T265S, and R313H/T265S, showed no changes in their optimum pH of the α2,3 ST activity, displaying 50-60% relative activity at pH 7.0 compared to those at pH 8.5. The same fold changes in their α2,3ST activities at pH 7.0 remained in all the mutants, resulting in two-fold higher activity for the synthesis of 3′-SL than that of wild-type α2,3PST without 6′-SL by-product formation.

In the case of α2,6PdST, I411 and L433 sites, which are situated inside the substrate binding pocket to make more direct contact with CMP-Neu5Ac rather than lactose, could generate mutants with enhanced catalytic activity. The mutants of I411T, L433S, and L433T enhanced k_{cat} values for CMP-Neu5Ac substrate by making the CMP-Neu5Ac move toward lactose or deleting the steric hindrance. Although these mutants decreased binding affinity of lactose, the mutants significantly increased k_{cat} values for lactose. It is quite hard to explain the exact mechanism, but increase in k_{cat} for lactose may due to the stabilization of the

transition-state complex and fast conformational dynamics caused by improvement of turnover rate for CMP-Neu5Ac.

In summary, this study provides insight into the protein engineering strategy to improve the activity of STs belonging to GT-B fold enzymes, which can be utilized for the production of 3'- and 6′-SL. Our approach is a hybrid approach that combines rational design based on computer modeling, alanine scanning, and saturation mutagenesis, and it can generally be used with any other proteins. The results have demonstrated that the catalytic activity of ST can be efficiently improved by the mutagenesis mainly focused on its active site, on the surrounding regions of the binding sites of the two substrates, i.e. glycan donor (CMP-Neu5Ac) and glycan acceptor (lactose). The alanine scanning and saturation mutation could generate dramatic changes in the ST expression level as a soluble form, and in its specific activities. Although most of them are still quite speculative and unknown, the detailed mechanisms to make such changes in protein solubility and activity could be explained to a certain extent by computer modeling and measurement of the kinetic parameters. In the case of α2,3PST/α2,6PdST, our results suggested that the introduction of a few H-bonds might be able to move the anomeric carbon of the CMP-Neu5Ac donor toward the hydroxyl group of the acceptor lactose, and/or to stabilize the binding of the two substrates in the active site, and that some mutations to reduce the steric hindrance to the access of substrates are keys to the improvement of their catalytic activities. In addition, since the 2,6 ST side reaction activity of α2,3PST appears to be caused by erroneous positioning of the two substrates in the active sites (especially the lactose acceptor), its abolishment was quite effective with even only one site-directed mutation, such as the mutation at

R313, yielding quite fascinating results. Another interesting observation in the mutation study was that one or two amino acid changes such as L433S and I411T/L433T of α2,6PdST could make dramatic changes in the k_{cat} value for lactose.

For further studies, one-pot reaction combining multiple enzyme steps for the large-scale synthesis of CMP-Neu5Ac and the SLs could be carried out with enhanced yield and productivity without generating the by-product using the improved ST mutants. Furthermore, the costs of the synthesis of 3'- and 6′-SL as well as their purification process originating from by-product formation (6′-SL in the case of 3′-SL synthesis) were greatly reduced, and the mutants could be widely used in the synthesis of various other 3'-and 6'-sialyloligosaccharides.

Chapter 4.

Solubilization of α1,2- and α1,3-fucosyltransferase and focused directed mutagenesis to increase in catalytic efficiency for fucosyllactose synthesis

4.1 Construction of *in vitro* production of fucosyllactose (FLs)

4.1.1 Production of GDP-fucose by salvage pathways

For *in vitro* production of GDP-Fuc that is used for donor substrate of fucosyloligosaccharide, L-fucokinase/GDP-fucose pyrophosphorylase (FKP) was used from L-fucose substrate. Since the biofuntional FKP converts L-fucose into GDP-Fuc via a fucose-1-phosphate intermediate (Figure 4.2), L-fucose, ATP, GTP, and $MnSO_4$ were used as starting substrates. In previous study, it was revealed that 1:1 ratio of Mn to L-fucose is required to produce high yield of GDP-Fuc (Kim 2013). For the large scale synthesis of GDP-Fuc, 1.5 L of reaction mixture containing 5 mM L-fucose, 5 mM ATP, 5 mM GTP, and 5 mM $MnSO_4$ in sodium phosphate buffer (pH7.6) was used with 20% v/v of FKP cell extract. The reaction was proceeded at 37 ºC with 200 rpm of shaking for 3 h. As shown at Figure 4.1, 64% yield of GDP-Fuc (1.24 g/L, total 1.86 g) was produced after 3 h, and then purified as 85% purity.

4.1.2 One-pot reaction for the production of FLs

For the economical production of FLs, one-pot reaction system was constructed combining the FKP and FucT responsible for synthesis of GDP-Fuc and FLs, respectively (Figure 4.2). Under the reaction condition containing 5 mM L-fucose, 5 mM ATP, 5 mM GTP, 5 mM $MnSO_4$, and 5 mM lactose acceptor, 2′-FL and 3-FL production were compared according to the enzyme unit ratio of FKP and each FucT. As shown at Figure 4.3, yield of 2′-FL and 3-FL increased as the ratio of FKP to FucT increased.

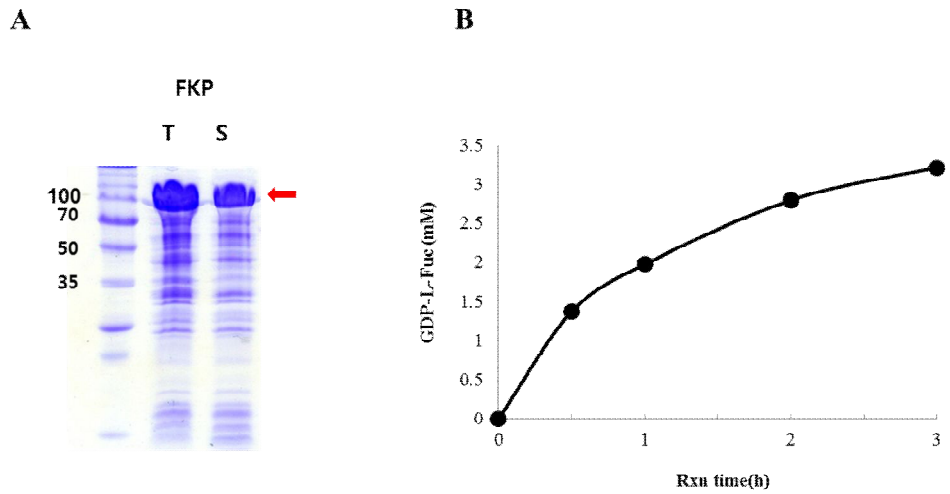


Figure 4.1 Protein expression of FKP and *in* vitro production of GDP-Fuc

(A) FKP was expressed as a soluble form with 105 kDa. (B) Large scale (1.5 L) synthesis of GDP-Fuc was conducted for 3 h, resulting in production of 1.24 g/L of GDP-Fuc.

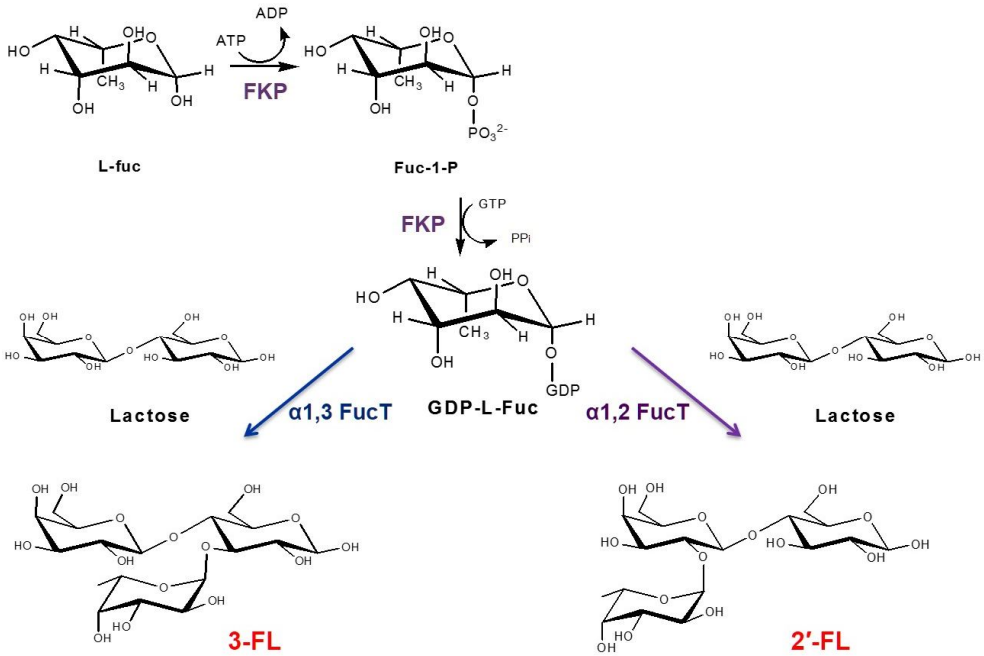
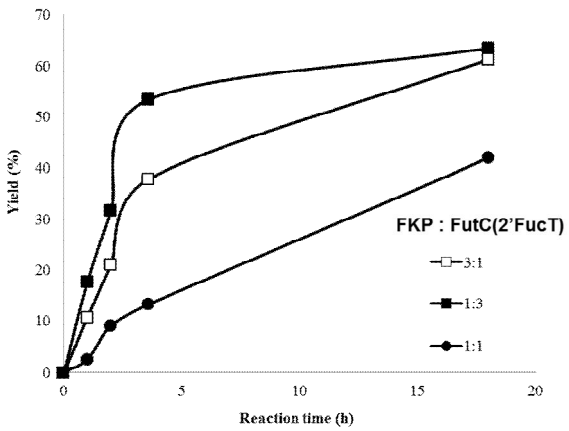


Figure 4.2 One-pot reaction for the production of FLs

2′-FL and 3-FL could be produced by one-pot reaction combining FKP with α1,2-FucT or α1,3-FucT, respectively, from L-fucose, ATP, GTP and lactose substrate.

A



B

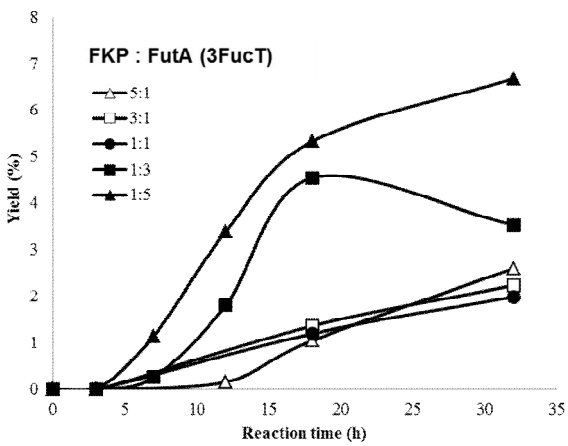


Figure 4.3 Production of 2′-FL and 3-FL according to the enzyme ration of FKP and FucT

The result elucidates that FucT is rate limiting step in one-pot reaction with FKP, indicating that need for engineering of FucT to increase the yield and productivity. It has been found that the rate limitation arised from FucT was due to the low protein solublity and low catalytic activity of FucTs.

In the case of one-pot reaction, in addition, Tris-HCl buffer hampered the FLs production, because amine of Tris-HCl generates complex with $MgATP^{2-}$, resulting in lack of ATP substrate. In one-pot reaction, yield increased by 1.9-fold when reaction buffer was changed from Tris-HCl to sodium phosphate buffer.

4.2 Soluble expression of fucosyltransferases (FucTs) in *E. coli*

4.2.1 Soluble expression of α1,2-FucT

The nucleotide sequence of FutC was codon optimized (cFutC), which resulted in 87% homology with its original sequence (Figure 4.4). As shown at Figure 4.5 A, total protein expression (33 kDa) was increased more than 5-fold than that of original FutC, but soluble protein expression was remained at a low level. To increase the soluble expression of cFutC by assisting the protein folding from aggregation form, co-expression of GroEL/ES chaperone was carried out according to the IPTG concentration. As a result, induction of 0.01 mM IPTG was the most effective to increase the soluble protein of cFutC, resulting in 5.6-fold improvement in 2′-FL yield.

For further increase in soluble expression of cFutC, various fusion proteins were introduced at N-terminus of cFutC. Fusion proteins used in this study are known to be as a water-soluble and active form.

Figure 4.4 Sequence alignment of original FutC and codon optimized FutC

Sequence alignment was done with original FutC and codon optimized FutC (cFutC). Sequence identity of the two sequences is 87%.

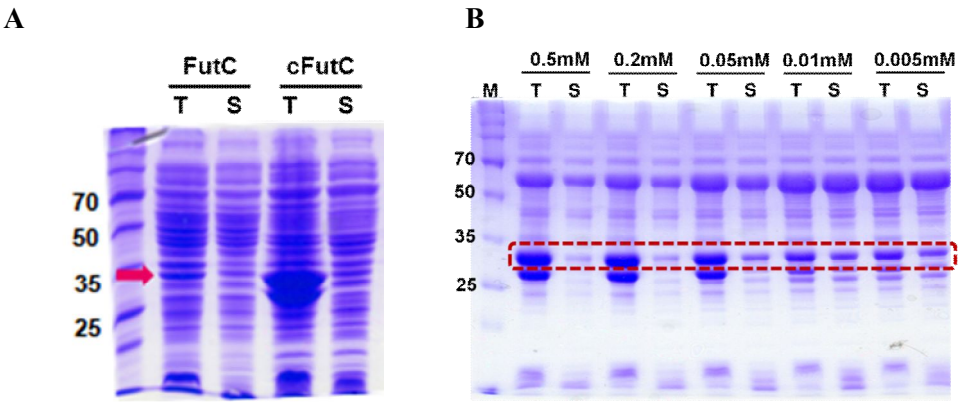


Figure 4.5 Protein expression of original FutC and codon optimized FutC (cFutC)

(A) Total (T, soluble and insoluble) and soluble (S) proteins of FutC and cFutC were analyzed by SDS-PAGE. (B) Co-expression with chaperones (GroEL/ES) for cFutC was analyzed according to the IPTG concentrations (0.005-0.5 mM). The protein size of GroEL and GroES are 60 and 10 kDa, respectively.

Especially, ArsC, F-ePGK, and N-ePGK from *E. coli* are recently studied as a very stable structure even under the protein denaturating conditions (Han et al. 2008). In addition, 30Kc19 originated from silkworm has been found to be capable of enhancing the solubility of aggregation-prone protein (Park et al. 2015). We constructed the fusion-cFutC expression system by introducing the fusion proteins (ArsC, F-ePGK, N-ePGK, GMPK, ACK, 30Kc19, and GST) at N-terminus of cFutC. The spacer including enterokinase cleavage sequence was inserted after fusion protein for reactive conformation of cFutC by making a space between the fusion partner and cFutC and further purification of cFutC protein. After protein expression, protein size combining fusion partners and cFutC was confirmed by analyzing the crude cell lysate. As shown in Figure 4.6, soluble expressions of fusion partner-fused cFutC were increased excepting for GST and 30Kc19 than that of cFutC with co-expression of chaperone. Especially, cFutC with the fusion of F-ePGK, N-ePGK, ArsC, and ACK dramatically increased the protein solubility (Kim et al. 2014). When the crude extracts of cFutC with the fusion proteins were used for 2′-FL production with GDP-Fuc, cFutC with the fusion of F-ePGK, N-ePGK, ArsC, and ACK showed higher activity than that of cFutC. In particular, F-ePGK fused cFutC exhibited the highest activity, showing 12.7- and 2.3-fold increase in 2′-FL production yield relative to those of cFutC and cFutC with co-expression of chaperone GroEL/ES, respectively (Table 4.1). The fold improvement of α1,2-FucT activity for 2′-FL production was not proportional to the expressed soluble protein, since the amount of soluble protein of F-ePGK fused cFutC was much more than 2.3 times than that of cFutC with co-expression with chaperon.

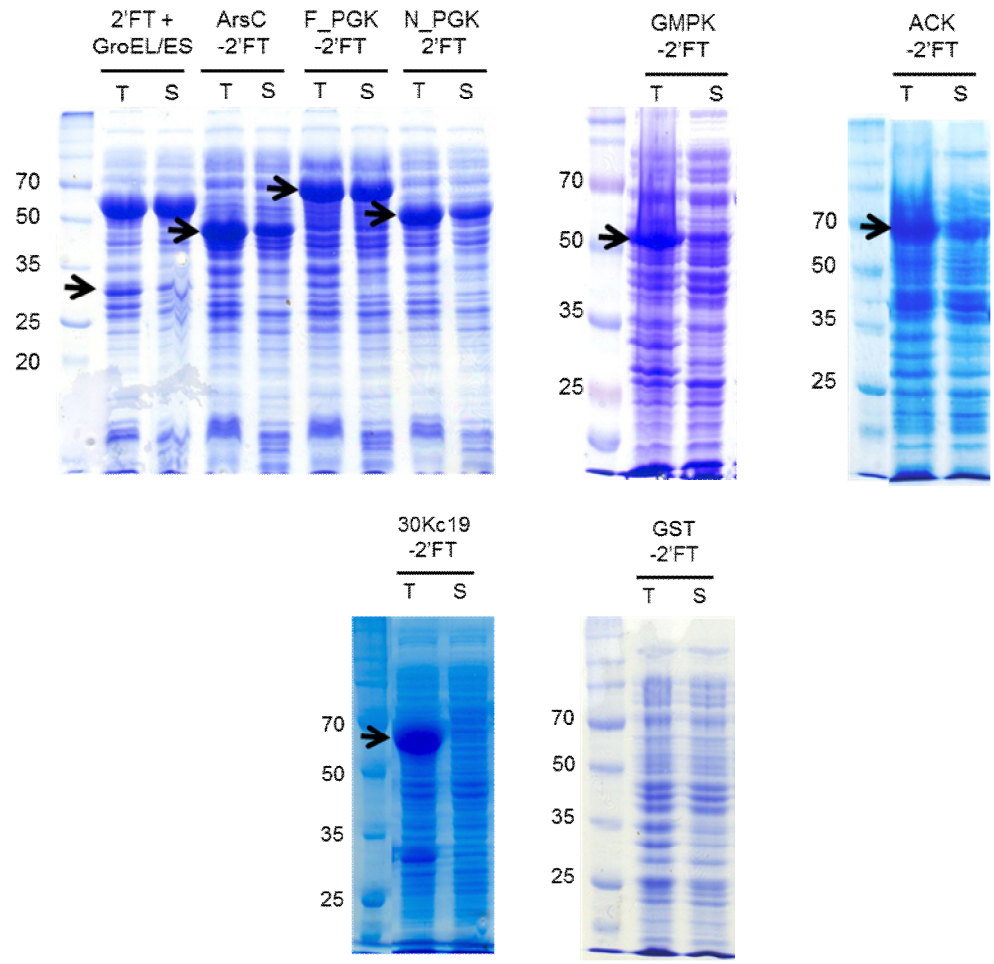


Figure 4.6 Protein expression of fusion protein-fused cFutC

Molecular weights of ArsC (16 kDa), F-ePGK (40 kDa), N-ePGK (21 kDa), GMPK (23 kDa), ACK (44 kDa), and 30Kc19 (30 kDa) -fused cFutC are 49, 73, 52, 56, 77, 63 kDa, respectively. GST (25 kDa)-fused cFutC was not expressed at all.

Table 4.1 Relative activities of fusion partner fused cFutC

	cFutC	cFutC + GroEL/ES	ArsC-cFutC	N-ePGK-cFutC	F-ePGK-cFutC	ACK-cFutC
Rel.Ac (%)	7.9	44	61	82	100	34

Reactions were carried out in 50 mM sodium phosphate pH 7.6 containing 5 mM GDP-Fuc, 2.5 mM lactose, 2.5 mM $MgCl_2$, and 20% (v/v) of crude cell extract of fusion protein fused cFutC at 37 ºC.

The result indicate that fusion partner F-ePGK might affect the α1,2-FucT activity partly.

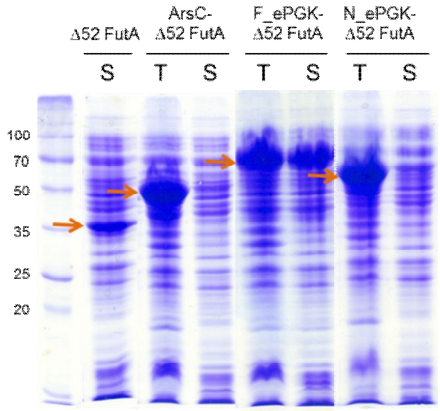
Nevertheless, total activity (U/mL) of F-ePGK fused cFutC as a fused form for direct expression was higher than those of cFutC or cFutC with co-expression of chaperone.

On the other hand, fusion protein expression system was applied to truncated α1,3-FucT (Δ52 FutA) using the same method described above. Three fusion partners such as F-ePGK, N-ePGK, and ArsC that were effective for soluble expression of cFutC, were used for construction of Δ52 FutA expression system. As a result, only F-ePGK fused Δ52 FutA showed increased solubility, but α1,3-FucT activity was almost abolished (Figure 4.7) for all the fusion expression system. As illustrated in Figure 4.8, there is little hindrance around N-terminus of FutC, whereas there is a structural hindrance around N-terminus of FutA. Although space sequence is inserted between F-ePGK and FutA, the hindrance could make collision with fusion protein. In addition, FutA have been known to form a dimer structure by C-terminus. The dimerization of the FutA might cause collisions between two fusion proteins or fusion protein and FutA, resulting in abolishment of activity. The significant reduction of α1,3-FucT activity might be due to restriction of substrates access or product release.

4.2.2 *in vitro* synthesis of 2′-FL

To produce 2′-FL by *in vitro* system, 20% v/v of FutC was used for the reaction with 5 mM GDP-Fuc, 5 mM lactose, and 5 mM $MgCl_2$ in 50 mM sodium phosphate buffer pH 7.6.

A



B

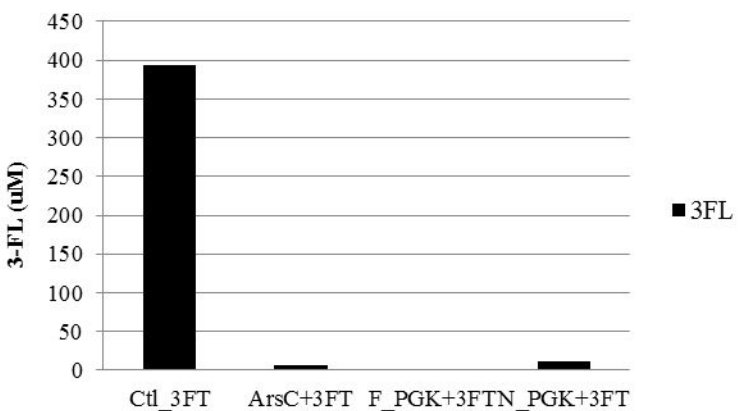


Figure 4.7 Protein expression of fusion protein-fused Δ52 FutA

(A) Fusion protein-fused Δ52 FutA were expressed with molecular weights of ArsC-fused Δ52 FutA (53 kDa), F-ePGK-fused Δ52 FutA (77 kDa), and N-ePGK-fused Δ52 FutA (58 kDa), respectively. (B) 3-FL production was carried out in 50 mM sodium phosphate pH 7.6 containing 5 mM GDP-Fuc, 2.5 mM lactose, 2.5 mM $MgCl_2$, and 20% (v/v) of crude cell extract of fusion protein-fused Δ52 FutA at 37 ºC.

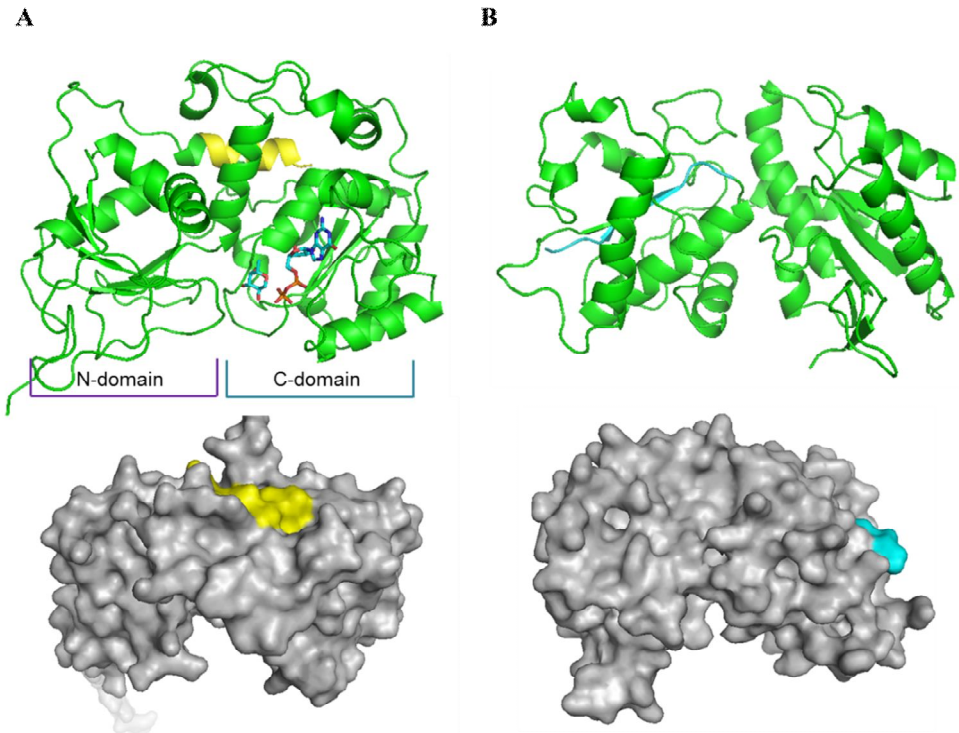


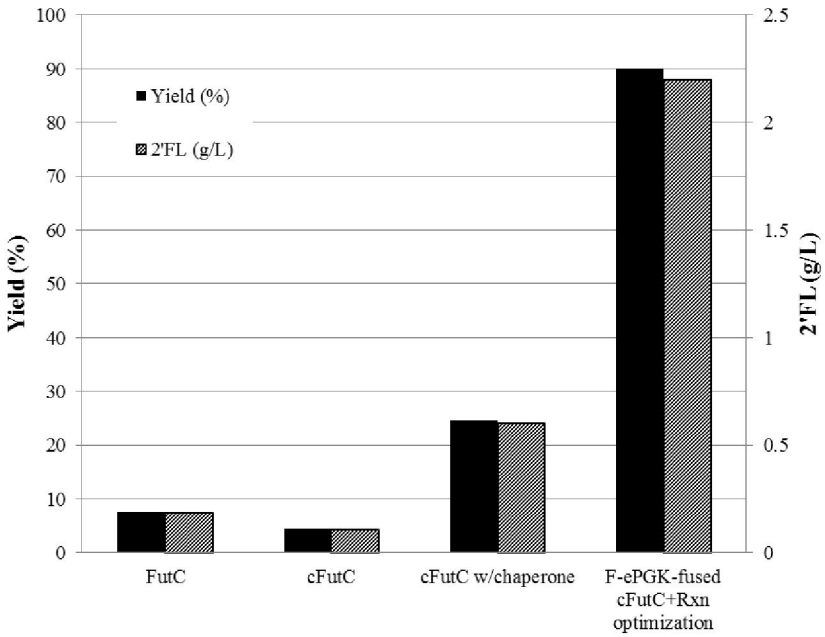
Figure 4.8 Modeling structure of Δ52 FutA (A) and FutC (B)

FutA shows GT-B topology consisting of two Rossmann domains that face each other and are linked flexibility. FutC is assumed to have GT-B fold due to the template structure of 2HHC (PDB accession code of FucT NodZ from *Bradyrhizobium*) used for modeling. N-terminus of Δ52 FutA and FutC are displayed in yellow and cyan, respectively.

As shown in Figure 4.9, 2′-FL yields were improved according to the increase in soluble protein expression of FutC. cFutC did not increase the 2′-FL yield, but cFutC with co-expression of chaperone showed 5.6-fold increase in yield relative to that of cFutC, resulting in achievement of 25% yield. With the F-ePGK fused cFutC, yield increased to 48%, which resulted in 1.9-fold increase in 2′-FL yield compared to cFutC with co-expression of chaperone. At the same time, 2′-FL reaction condition was re-optimized. In order to increase the reaction rate, relative cheap substrate, lactose, was able to be added up to 15 mM without substrate inhibition. At the optimized condition, using the F-ePGK-fused cFutC, over 90% yield was achieved after 3 h, based on 5 mM GDP-Fuc. As a result, yield and productivity (g/L/h) significantly increased by a factor of 12 and 47 respectively, compared to those of original FutC.

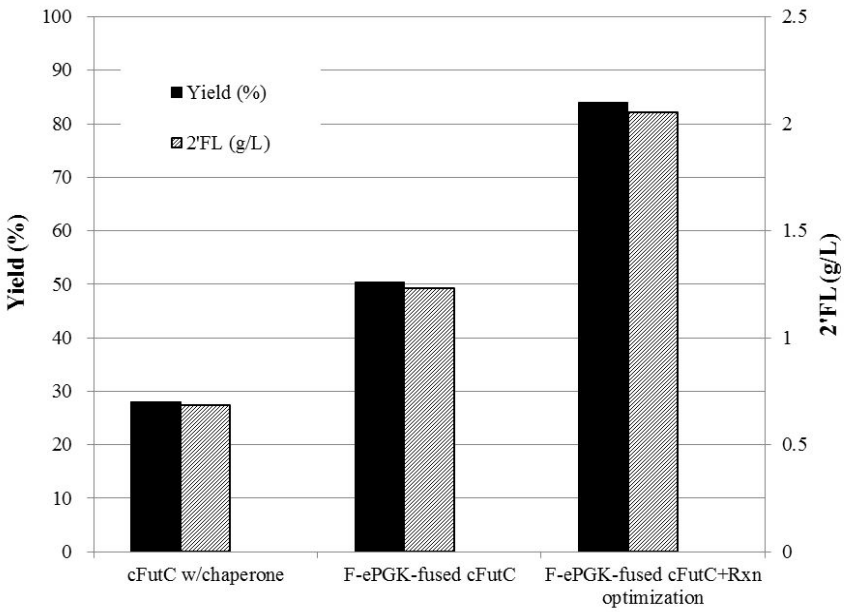
When F-ePGK-fused cFutC was used for one-pot reaction combining synthesis of GDP-Fuc and 2′-FL, 84% 2′-FL yield was achieved after 8 h based on 5 mM L-fucose, indicating 3-and 4.5-fold increase in 2′-FL yield and productivity respectively, compared to those of cFutC with co-expression of chaperon (Figure 4.9).

A



	2'FL (g/L)	Reaction (h)	Productivity (g/L/h)	Rxn Cond'n
Non-optimized 2'FT (FutC)	**0.19**	**12**	**0.016**	5 mM GDP-Fuc + 5 mM Lac
Optimized 2'FT (cFutC)	**0.11**	**12**	**0.009**	5 mM GDP-Fuc + 5 mM Lac
Optimized 2'FT w/chaperone (cFutC w/chaperone)	**0.60**	**12**	**0.050**	5 mM GDP-Fuc + 5 mM Lac
F-ePGK-fused cFutC+ Rxn optimization	**2.20**	**3**	**0.733**	5 mM GDP-Fuc + **15 mM Lac**

B



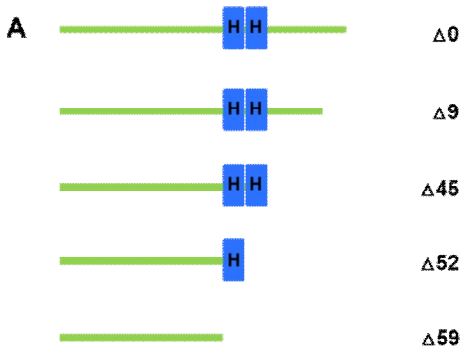
	2′FL (g/L)	Reaction (h)	Produc-tivity (g/L/h)	Rxn Cond'n
cFutC w/chaperone	0.68	12	0.057	5 mM substrates + 5 mM Lac
F-ePGK-fused cFutC	1.23	12	0.103	5 mM substrates + 5 mM Lac
F-ePGK-fused cFutC+Rxn optimization	2.05	8	0.256	5 mM substrates + 15 mM Lac

Figure 4.9 Increase in yield for *in vitro* synthesis of 2′-FL via improvement of soluble protein expression level of FutC and reaction optimization

(A) 2′-FL synthesis was conducted with GDP-Fuc and lactose. (B) 2′-FL synthesis was conducted with one-pot reaction from L-fucose using FKP and FutC enzymes. FutC, original α1,2-FucT; cFutC, codon-optimized FutC; cFutC w/chaperone, cFutC with co-expression of GroEL/ES; cFutC w/chaperone+Rxn optimization, cFutC with co-expression of GroEL/ES and reaction optimization for increase in lactose concentration (15 mM).

4.2.3 Soluble expression of α1,3-FucT

The soluble fraction of the expressed recombinant FutA was very low in spite of the various trials including co-expression with fusion partner proteins, chaperones, control of induction temperature and/or IPTG concentration, and cold shock expression (data not shown). FutA has a unique C-terminus with two heptad repeat region (DDLRVNY) that appears to be involved in protein dimerization, and the α-helices segments of hydrophobic, positively charged residues in the membrane anchoring region. Under the hypothesis that the α-helices in the membrane anchoring region hamper the soluble expression in cytosol, the truncation of the α-helices was expected to increase the expression level of the soluble protein. Four differentially-truncated forms of the C-terminal deletion for 9, 45, 52, and 59 residues were constructed to produce deletions of the partial and complete α-helices, one heptad repeat, and two heptad repeats, respectively (Figure. 4.10 A). The truncated mutants were evaluated via SDS-PAGE analysis for protein expression (Figure. 4.10 C) and 3-FL synthesis (Figure. 4.10 B). All variants except for Δ9 FutA exhibited higher relative activity when compared to the non-truncated form. Δ45, Δ52, and Δ59 truncated FutA mutants showed an increased soluble expression as well as total expression of FutA in the *E. coli* cells. Among these, Δ52 FutA, which has only one heptad repeat, showed the highest solubility, corresponding to a 2.5 times increase in activity (U/mL) relative to that of the non-truncated form. The result indicates that only one heptad repeat without a membrane anchoring region is sufficient for proper folding as an active form of FutA. Furthermore, the nucleotide sequence of Δ52 FutA was codon optimized, which resulted in 76% homology with the original sequence.



B

FutA	Relative Ac (% for U/mL)
△ 0	100
△ 9	2
△45	228
△52	248
△59	199

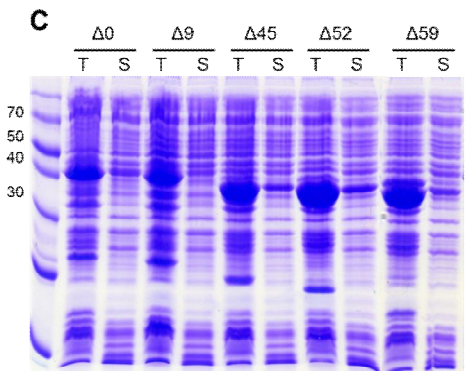


Figure 4.10 Improvement in soluble enzyme expression level by systemic deletion of C-terminal α-helices and heptad repeat region of FutA

(A) Systemic truncation of C-terminus was designed as deletion of partial (Δ9), complete α-helices (Δ45), one heptad repeat (Δ52), and two heptad repeat (Δ59). 3-FL synthesis reactions were performed with 5 mM GDP-Fuc, 2.5 mM Lac, 5 mM $MgCl_2$, and 30% (v/v) of cell-free extracts of truncated FutA variants in 50 mM sodium phosphate (pH 7.6) for 14 h at 37 ºC. Relative activities of U/mL were shown by average values of twice experiments (B). 3-FL yields (%) of Δ0 FutA and Δ52 FutA for GDP-Fuc were 2.4 and 6.0, respectively. After obtaining the soluble proteins, total (T, soluble and insoluble fraction) and soluble (S) fractions were analyzed by SDS-PAGE (C).

The newly synthesized gene was cloned under a T5 strong promoter, and induction was carried out with 1mM of IPTG concentration at 30 °C using an LB broth. The soluble protein expression level of the synthesized gene remarkably increased when cultured in rich media, such as terrific broth (TB), which contains additional carbon and nitrogen sources (Figure. 4.11). Further purification using nickel affinity chromatography yielded about 150-200 mg/L (culture broth) of codon-optimized Δ52 FutA (Figure. 4.12), showing the highest soluble protein expression level of α1,3-FucT known to date. At the same time, in order to increase the reaction rate for 3-FL production, relative cheap substrate, lactose, was able to be added up to 15 mM without substrate inhibition like the production of 2′-FL. The substantial increase in the soluble protein of codon optimized Δ52 FutA allowed an increase in the 3-FL production yield from 2.4% to 45% with a reduced amount of enzymes (i.e. 10% v/v, one third of the volume for original FutA) over a short amount of time (i.e., 3 h compared to 14 h for original FutA). Therefore, Δ52 FutA became the template for further protein engineering.

4.3 1st Generation mutagenesis: Site-saturation mutagenesis (SSM) for the functional residues in active site

4.3.1 Homology modeling and analysis of active site of α1,3-FucT

In order to attain a substantial increase in the yield and productivity for the synthesis of 3-FL, we aimed to increase the specific activity and change the substrate specificity of Δ52 FutA for lactose. To achieve this goal, semi-rational mutagenesis was performed by combining structure-guided computational analysis

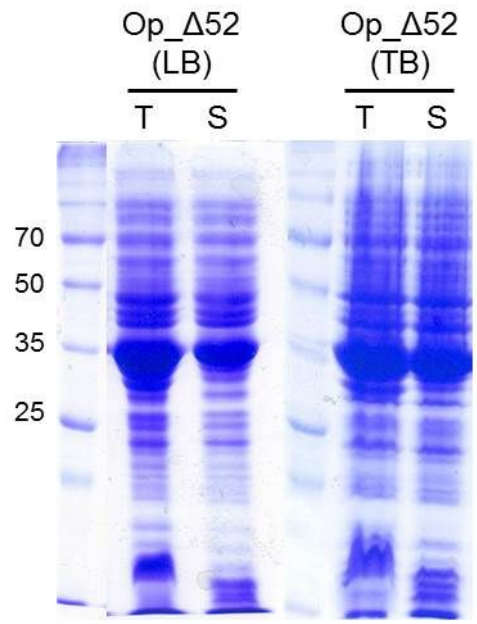


Figure 4.11 Improvement in soluble enzyme expression level of C-terminus truncated FutA (Δ52 FutA) followed by codon optimization

Codon optimized Δ52 FutA (Op_Δ52) was induced with 1 mM IPTG in LB or TB media, and the cells were further incubated at 30 ºC for 12 h. Harvested cells were suspended in 20 mM sodium phosphate buffer (pH 7.6) and concentrated 10 times. The soluble proteins were obtained by centrifugation at 22,250 × g for 30 min at 4 ºC after cell lysis. After loading of equal volumes of total (T, soluble and insoluble fractions) and soluble (S) fractions, gel was stained by Coomassie Brilliant Blue R-250.

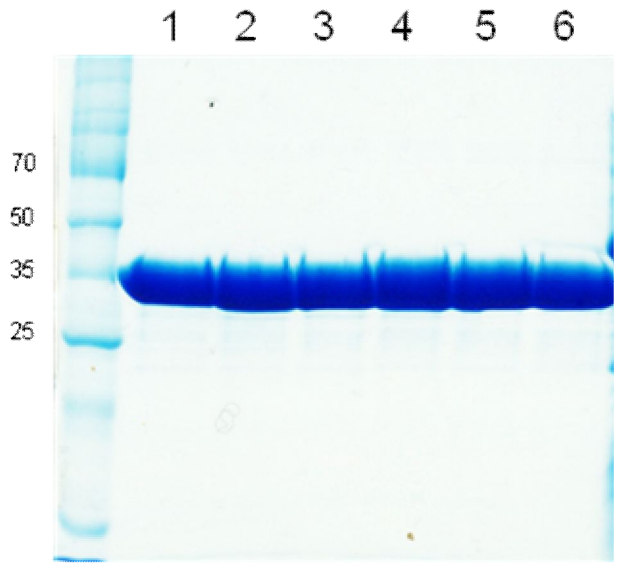


Figure 4.12 SDS-PAGE of purified wild-type and mutant enzymes of Δ52 FutA

After purification, equal volumes of purified enzymes were loaded on SDS-PAGE. Protein concentrations of wild-type and mutants were from 7.2 to 10.4 mg per milliliter from 50 mL culture. 1-6, Δ52 FutA wild-type and variants; 1, wild-type (Δ52 FutA); 2, A128N; 3, A128N/H129E; 4, A128N/H129E/Y132I; 5, A128N/H129E/S46F; 6, A128N/H129E/ Y132I/S46F

and saturation mutagenesis. As a first step, SSM was attempted to find the 'best hit'. To investigate the target region for mutagenesis in the active site, homology modeling of Δ52 FutA was conducted using 2NZY as a template, for which the structural data were acquired with the protein-ligand complex of GDP-Fuc. When the model structure for Δ52 FutA was superimposed onto the 2NZY structure, the structure was quite similar to 2NZY showing a GT-B fold structure consisting of two separate Rossmann-fold domains. According to the structure of the superimposed model, the catalytic key residue that functions as a general base to deprotonate from a reactive hydroxyl group of lactose acceptor was deduced as Glu96, which corresponded to Glu95 of 2NZY (Sun et al. 2007). When Glu96 of Δ52 FutA was substituted for alanine, a complete abolishment of the α1,3-FucT activity was observed, which was in good agreement with the suggested role for a general base residue (Figure 4.13).

4.3.2 Docking simulation of lactose for α1,3-FucT model structure

The docking simulation of lactose was conducted for Glu96, and the docking energies (*E*, kcal/mol) were calculated with the distance (δ, Å) between the glucose moiety (O3) of lactose and Glu96 ($O\varepsilon_2$) according to the inverting catalytic mechanism of GT (Lairson et al. 2008) such as FutA. The docking pose with the minimum energy value (E/δ) among the docking runs was selected. As a result, the docked model of lactose could be constructed in the active site cleft of Δ52 FutA where O3 of glucose was situated toward $O\varepsilon_2$ of Glu96 (Figure. 4.14 A).

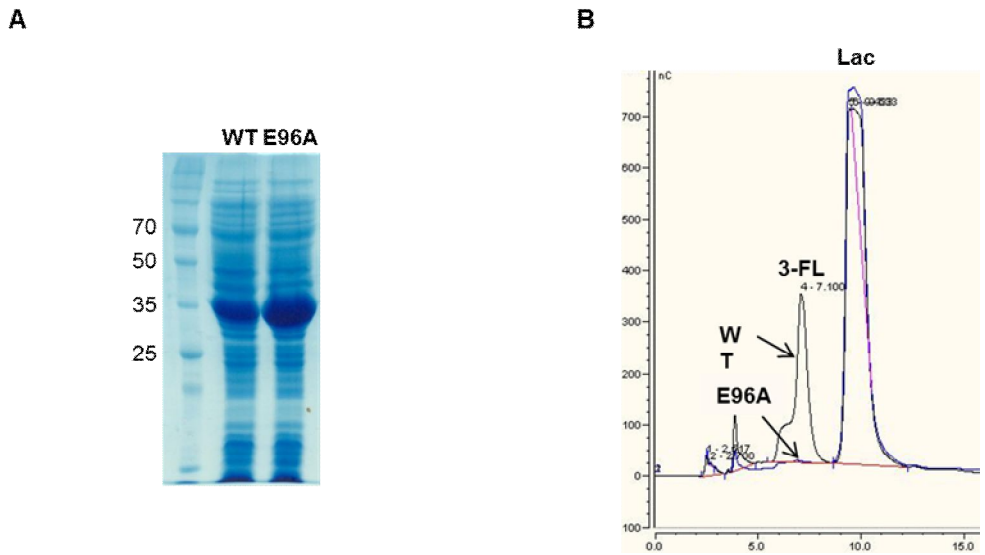


Figure 4.13 SDS-PAGE of purified wild-type and mutant enzymes of Δ52 FutA

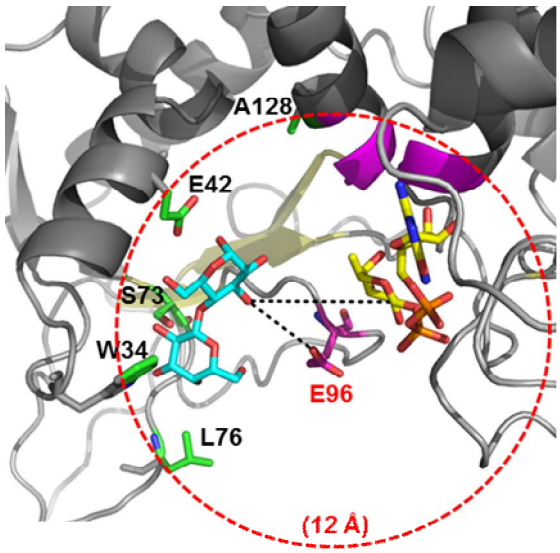
E96A mutant was expressed using the same expression method of wild-type (Δ52 FutA). After obtaining the soluble fractions, soluble expression levels were analyzed by SDS-PAGE (A). 3-FL synthesis reaction was performed with 5 mM GDP-Fuc, 15 mM Lac, 5 mM $MgCl_2$, and 10% (v/v) of cell-free extracts of the enzymes in 50 mM sodium phosphate buffer (pH 7.6) for 3 h at 37 ºC. Bio-LC analysis indicates that substitution of E96 with alanine (E96A) led to complete activity loss (B).

Since the distance between O3 of glucose and C1 of fucose was 9.7 Å, a sphere of 12 Å in radius from C_α of Glu96 was drawn to include all the residues within the distance, and 51 residues were identified. In parallel, with the docking simulation, HotSpot Wizard server was utilized to estimate the mutability of the residues in the active site or the lining of the access tunnels. Among the 51 residues identified from the docking simulation, only 5 amino acids, such as W34, E42, S73, L76, and A128 were ranked as residues with a high mutability (Figure. 4.14 A). The five high mutability residues in the active site were individually subjected to SSM, and the mutants showing high activity were screened using a colorimetric assay, as mentioned under the 'Materials and Methods' section.

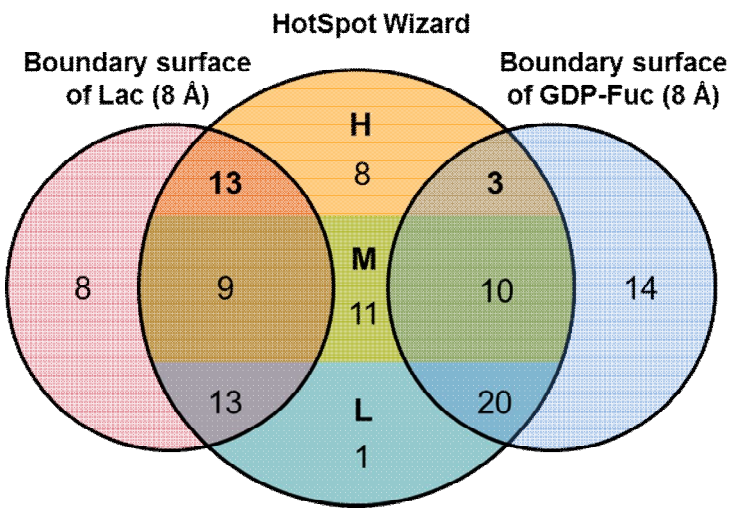
4.3.3 SSM for functional residues to find 'best hit'

Among the screened libraries, several mutants were found to exhibit improved activity only at the A128 site while no other mutants exhibited higher activity than that of wild-type (data not shown). The substitutions of A128 for Gly and Asn were revealed to be responsible for the 1.4 and 3.3-fold increase in activity, respectively, which were verified with the purified enzymes (Figure. 4.15). Based on the result of SSM, the A128N mutant was chosen as the head start for the next ISM.

A


A128
E42
S73
W34
E96
L76
(12 Å)

B


HotSpot Wizard
Boundary surface of Lac (8 Å)
Boundary surface of GDP-Fuc (8 Å)
H
8
13
3
M
11
8
9
10
14
13
L
1
20

C

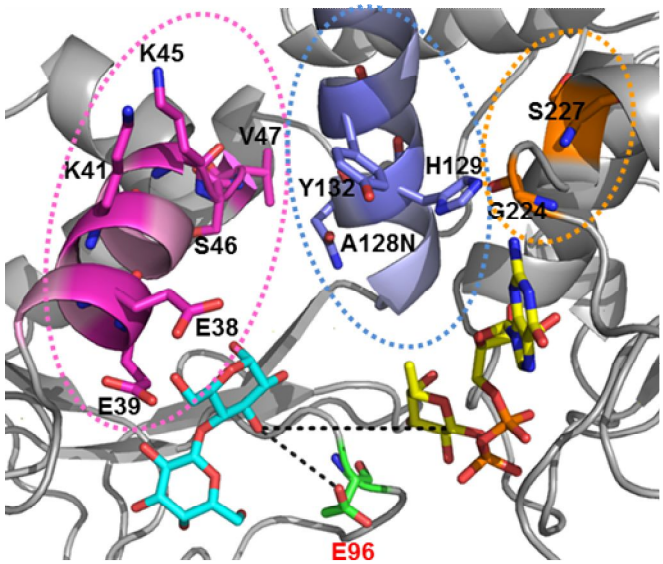
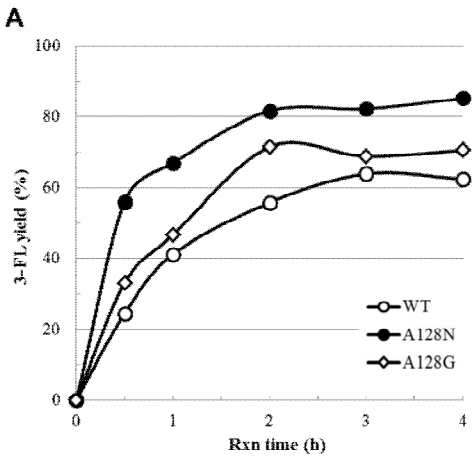


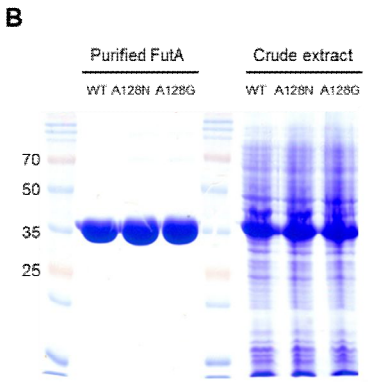
Figure 4.14 Rationale for the selection of candidates for 'best hit' from Cα of catalytic key residue and clustering of α-helices on substrate binding sites for ISM

(A) Five overlapping residues derived from docking (51 residues for a sphere of 12 Å from C_α of E96) and HotSpot Wizard with a high mutability are presented in green color (W34, E42, S73, L76, and A128 residues). The five residues were individually subjected to SSM.

(B) The 13 residues at the intersection of the boundary surface of lactose and with a high mutability from HotSpot Wizard were identified as W34, E38, E39, K41, E42, K45, S46, V47, S73, L76, A128, H129 and Y132. The 3 residues at the intersection of the boundary surface of the GDP-Fuc and high mutability from HotSpot Wizard were also identified as H129, G224, and S227. The ranges of the mutability levels are 1-3 (L, low), 4-6 (M, moderate), and 7-9 (H, high). (C) The helices of α5

(A128N, H129, and Y132), α2/loop (E38, E39, K41, K45, S46 and V47), and α10/loop (G224, and S227) were individually clustered for ISM. The residues on three helices are presented in blue (α5), magenta (α2/loop), and orange color (α10/loop), respectively. GDP-Fuc and lactose are presented in yellow and cyan color, respectively.





C

	Relative Specific activity (%)
WT	100
A128N	325
A128G	141

Figure 4.15 Evaluation of the mutants at A128 site for Δ52 FutA by determination of specific activities for the 3-FL synthesis

(A) 3-FL synthesis was performed with 5 mM GDP-Fuc, 15 mM lactose, 5 mM $MgCl_2$ and 20% (v/v) of crude extract of FutA mutants in 50 mM sodium phosphate buffer (pH 7.6) for 4 h at 37 ºC. (C) Specific activities were determined with 20 μg of purified enzymes (B) using the same reaction condition for 15 min. The mean values of relative specific activity are shown from duplicate experiments and the value of specific activity (U/mg) for the wild-type is 0.082±0.010.

4.4 2nd Generation mutagenesis: structure-guided iterative saturation mutagenesis (ISM) from 'best hit'

4.4.1 Selection of functional residues in substrate binding site using HotSpot Wizard

The A128 residue is located on the α5 helix in the deep cleft of the lactose binding site identified from the docking simulation (Figure. 4.14 A). Since the orientation of the side chain of Asn in the A128N mutant goes toward the glucose moiety, the A128N mutation appears to increase the α1,3-FucT activity due to the changes in lactose binding. This analysis indicates that more residues interacting with the lactose in the substrate binding site could be inferred to be good candidates for the following saturation mutagenesis. According to the docking simulation of Δ52 FutA, the distance from lactose to A128 was 7.9 Å, suggesting that all residues within an 8 Å distance from the boundary surface of lactose have some possibility to participate in the interactions between lactose and the lactose binding site or its surroundings to change the FucT activity. Among the poses docked with the lactose substrate, several reasonable coordinates with a similar orientation of lactose to the docked pose of A128 resulting in less than -4.7 kcal/mol of final docked energy were further screened. As shown in Figure. 4.16, five such poses of lactose were identified, and all residues within an 8 Å distance from the boundary surface of lactose were selected (Table 4.2). Among the residues that were screened, we sorted out the residues overlapping with the sites of a high mutability using HotSpot Wizard (Figure. 4.14 B).

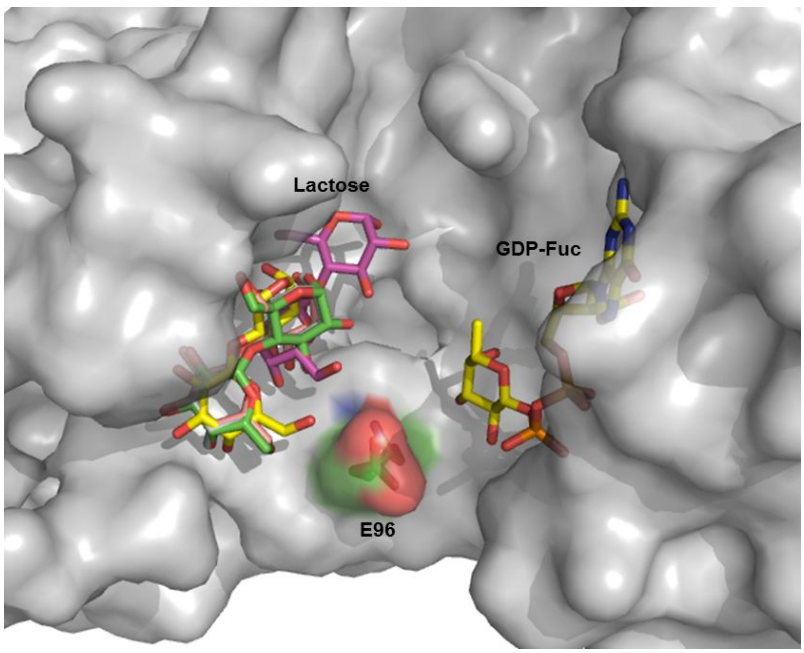


Figure 4.16 Coordinates of lactose for Δ52 FutA with GDP-Fuc

Lactose coordinates are presented according to the docking poses as shown at Table SII with GDP-Fuc and E96. Δ52 FutA is displayed by surface form.

Table 4.2 Residues within 8 Å distance from the boundary surface of lactose and GDP-Fuc for Δ52 FutA

8 Å from lactose							8 Å from GDP-Fuc	
Pose 1	**Pose 2**	**Pose 3**	**Pose 4**	**Pose 5**	**Sum**	**Mutability**	**Pose**	**Mutability**
V31	V31	V31	A32	V31	V31	6	T94	2
A32	A32	A32	W34	A32	A32	3	G95	2
W34	N33	N33	W35	N33	N33	1	E96	1
W35	W34	W34	E39	W34	W34	9	N97	2
E39	W35	W35	I40	W35	W35	4	E98	5
I40	G36	G36	E42	G36	G36	n.d	F111	3
E42	E38	E38	F43	E38	E38	9	L125	3
F43	E39	E39	S73	E39	E39	9	Y126	n.d
K44	I40	I40	N74	I40	I40	n.d	H129	9
K45	K41	E42	P75	K41	K41	7	S185	n.d
S46	E42	F43	L76	E42	E42	7	F186	4
V47	F43	S46	Y93	F43	F43	3	V187	1
L48	S46	L48	T94	K44	K44	n.d	A188	4
Y49	L72	L72	G95	S46	K45	9	S189	2
L72	S73	S73	E96	L48	S46	8	N190	1
S73	N74	N74	E98	L72	V47	9	A191	n.d
N74	P75	P75	L125	S73	L48	2	N192	n.d
P75	L76	L76		N74	Y49	n.d	A193	n.d
V91	F92	F92		P75	L72	5	R196	1
F92	Y93	Y93		L76	S73	8	G212	4
Y93	T94	T94		G77	N74	4	G213	1
T94	G95	G95		F92	P75	6	S214	n.d

G95	E96	E96	Y93	L76	8	V215	n.d
E96	E98	E98	T94	G77	4	Y221	n.d
N97	L125	I109	G95	V91	n.d	K222	6
E98	A128	L125	E96	F92	3	V223	4
M123		A128	N97	Y93	6	G224	9
P124			E98	T94	2	N225	4
L125			M123	G95	2	K226	4
Y126			L125	E96	1	S227	9
Y127				N97	2	E228	n.d
A128				E98	5	F229	n.d
H129				I109	1	L230	n.d
H131				M123	5	N236	3
Y132				P124	1	L237	1
E250				L125	3	C238	2
				Y126	n.d	F239	3
				Y127	n.d	E240	n.d
				A128	9	N241	1
				H129	9	S242	n.d
				H131	n.d	Y247	1
				Y132	9	V248	4
				E250	1	T249	1
						E250	1
						K251	1
						I252	n.d
						D254	5

n.d (not detected) means that resides not included in HotSpot region.

4.4.2 Structure-guided iterative saturation mutagenesis (ISM)

As illustrated in Figure. 4.14 C, the residues that were selected, except for the previously studied sites, were located in the two α-helices of α5 (A128N, H129, and Y132) and α2/loop (E38, E39, K41, K45, S46 and V47). Furthermore, the same method was used to identify the residues within an 8 Å distance from the GDP-Fuc as the third cluster for loop/α10 (G224 and S227). As a consequence, the candidate mutation sites were grouped into three for the next structure-guided ISM, including two α-helices around the lactose binding site and one α-helix around the GDP-Fuc binding site. To minimize the library size, ISM was performed for each cluster of α-helices in parallel, and the combinations of the mutations from each cluster were carried out as a final step. From the first cluster of the α5 helix containing the A128N mutation, ISM was sequentially conducted at the H129 and Y132 sites in order. Among the possible pathways, the H129 site was chosen for the first round of ISM since the H129 site is located in the same turn of the α5 helix with A128N and its side chain can be oriented to lactose with a side chain rotation, as confirmed with the pyMOL mutagenesis tool. Screening the first round ISM at the H129 site resulted in A128N/H129D and A128N/H129E mutants, showing a further increase in the α1,3-FucT activity at twice that of the A128N single mutant. Although the A128N/H129D variant was slightly less active than A128N/H129E, the substitution of H129 with Glu or Asp, such as acidic and hydrophilic residues, was responsible for the increase in activity. The A128N/H129E mutant was verified again through the 3-FL production using a purified enzyme, which resulted in an increase in the specific activity of 6.5-fold that of wild-type Δ52 FutA (Table 4.3).

Table 4.3 Relative specific activities of the various mutants of Δ52 FutA for 3-FL and Le^x production

	3-FL		Le^x	
	U/mg	Relative Specific Ac (%)	U/mg	Relative Specific Ac (%)
WT (Δ52 FutA)	0.08±0.01	100	1.75±0.52	100
A128N	0.27±0.03	338	4.38±0.09	250
A128N+H129E	0.52±0.01	650	8.33±0.68	475
A128N+H129E+Y132I	0.77±0.03	963	9.03±0.46	515
A128N+H129E+S46F	1.16±0.02	1450	14.49±2.09	828
A128N+H129E+Y312I+S46F	1.24±0.04	1550	13.17±1.25	753

The relative specific activity for the Δ52 FutA variants was compared against that of the parent enzyme. To produce 3-FL, reactions were carried out with 5 mM GDP-Fuc, 15 mM lactose, 5 mM $MgCl_2$ and 20 μg of purified enzyme in 50 mM sodium phosphate buffer (pH 7.6) for 15 min at 37 °C. In the case of the Le^x production, 5 mM LacNAc and 5 μg of purified enzyme were reacted for 10 min in the same reaction conditions above. The values consist of the means with standard deviations of the results from three experiments.

Next, the A128N/H129E variant was used as the starting point for the second round of ISM at the Y132 site. The α1,3-FucT activity was further enhanced by the Y132 mutants, which possessed a mutation in the hydrophobic residues such as Ile, Trp, Val, and Leu (data not shown). The mutations for Y132 seemed to change the interaction with lactose by eliminating the H-bonding. Among the Y132 variants, the A128N/H129E/Y132I mutant displayed the highest improvement in specific activity at 9.6-fold that of wild-type (Table 4.3).

In parallel, the SSM among the six residues (E38, E39, K41, K45, S46 and V47) on the second cluster α2 helix/loop was performed individually using the wild-type as a template to find the starting point for ISM. Although several mutants at the E38 and K45 sites, such as E38S, E38N, K45A, and K45T, displayed similar activity with that of wild-type, no better variants were identified from the E39, K41, and V47 libraries (data not shown). Only the S46 mutation appeared to show improved activity with S46F and S46L respectively showing 1.6- and 1.4-fold higher activity than that of the wild-type. In the case of S46F, the side chain of S46F pointing toward the glucose moiety of lactose can change the conformation of the lactose binding site. Since only the S46 site generated enhanced mutants, further ISM was not performed for the α2 helix/loop.

Only three residues (H129, G224, and S227) were chosen in the third cluster (loop/α10) of the GDP-Fuc binding site because a large portion of the residues exhibit a low mutability as a result of the highly conserved sequences (Table 4.2).

Since the H129 site is located in the intersection of the α5 and loop/α10 helices, G224 and S227 were subjected to SSM using the wild-type as a template. However, the screened mutants, such as G224E, G224D, G224S, S227G, and

S227T showed about the same or a somewhat lower activity than that of wild-type. Previous studies have reported that the C-terminal domains responsible for binding the nucleotide sugar donor are generally more conserved than the N-terminal domain (Albesa-Jove et al. 2014; Sun et al. 2007). In particular, the sequence alignment for FutA showed that the C-terminal domain has a high homology with various UDP- or GDP-dependent GTs. The stringent conservation of the sequences in the donor binding site would be essential to retaining its own activity.

Given that the A128N/H129E/Y132I and S46F mutations on the two clusters of the α5 helix and α2 helix/loop individually improved the activity toward lactose, combinations of the mutations were evaluated for any synergistic effects on their activity. When an additional S46F mutation on α2 helix/loop cluster was introduced to the A128N/H129E double and A128N/ H129E/Y132I triple mutant, synergic effects were consistently observed compared to the single mutational effect of S46F. The A128N/H129E/S46F triple mutant exhibited an increase in its specific activity by a factor of 14.5 relative to that of wild-type, which is a more noticeable improvement than that with Y132I, suggesting that the effect of the S46F mutation is quite cooperative with the effect of the mutations on the α5 helix for lactose binding.

As a consequence, the quadruple variant, i.e., A128N/H129E/Y132I/S46F, that was generated by combining all mutations on two α-helices displayed a synergistic effect on its specific activity of 15.5-fold that of the wild-type (Δ52 FutA) (Table 4.3).

4.5 Kinetic analysis and docking simulation of Δ52 α1,3-FucT mutants

To understand the underlying mechanisms for the individual mutations of Δ52 FutA to the changes in specific activities, the kinetic parameters for lactose and the GDP-Fuc substrates were evaluated using purified enzymes (Table 4.4 and Fig. 4.12). Notably, all mutants have K_m values that are substantially decreased for lactose by a factor of 2.3 to 7.6, indicating an increase in the binding affinity for lactose. In addition, the k_{cat} values of the mutants were improved for lactose, resulting in an improvement of the k_{cat}/K_m values by 3- to 20-fold compared to that of the wild-type (Δ52 FutA). On the other hand, in terms of the kinetic parameters for the GDP-Fuc donor, all mutants displayed little effect on the K_m values but a significant impact on k_{cat} by due to the increase by a factor of 1.8 to 24 relative to that of wild-type.

As shown in Table 4.4, the A128N single and A128N/H129E double mutants were the most effective in increasing the binding affinity of lactose, resulting in decreases in K_m by a factor of 2.3- and 4.4 relative to wild-type, respectively. In particular, the A128N/H129E double mutant substantially enhanced the k_{cat} values for GDP-Fuc (6.6-fold) as well as lactose (1.8-fold), resulting in a 7.9- and 5.9-fold improvement in the k_{cat}/K_m values for lactose and GDP-Fuc, respectively, compared to wild-type. According to the docked lactose for A128N/H129E double mutant model, the side chains of Asn128 and Glu129 located in the same turn of α5 helix go toward the lactose binding site as a pincer-like structure (Figure. 4.17 A). The structure could generate new H-bonds interactions between Asn128/Glu129 and the glucose moiety.

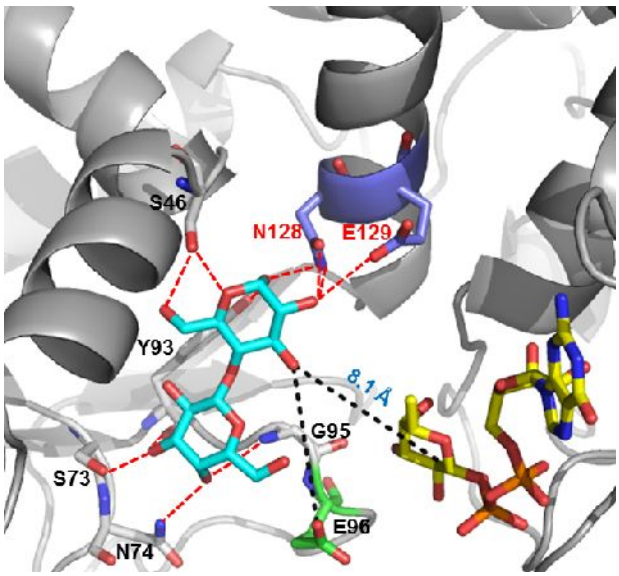
Table 4.4 Kinetic parameters of the wild-type (Δ52 FutA) and mutants from this study

Mutants of Δ52 FutA	Lactose [a]			GDP-fucose [b]		
	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m (min^{-1} mM^{-1})	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m (min^{-1} mM^{-1})
WT	135±4.3	543±0.7	0.25	3.8±0.08	0.33±0.01	11.45
A128N	171±0.8	232±10.8	0.74	6.8±0.13	0.37±0.04	18.49
A128N/ H129E	242±2.5	123±7.6	1.97	25.0±0.23	0.37±0.01	67.49
A128N/ H129E/Y132I	240±2.9	97±1.0	2.47	28.7±1.03	0.24±0.03	119.50
A128N/ H129E/S46F	337±5.4	81±0.7	4.16	79.7±1.08	0.28±0.01	284.54
A128N/H129E /Y132I/S46F	358±5.7	72±0.2	4.97	90.7±0.62	0.30±0.00	302.43

[a] Kinetic measurements at 2 mM GDP-Fuc and variable lactose concentrations (5-700 mM)

[b] Kinetic measurements at 10 mM lactose and variable GDP-Fuc concentrations (0.05-2 mM)

A


S46
N128
E129
Y93
8.1 Å
G95
S73
E96
N74

B

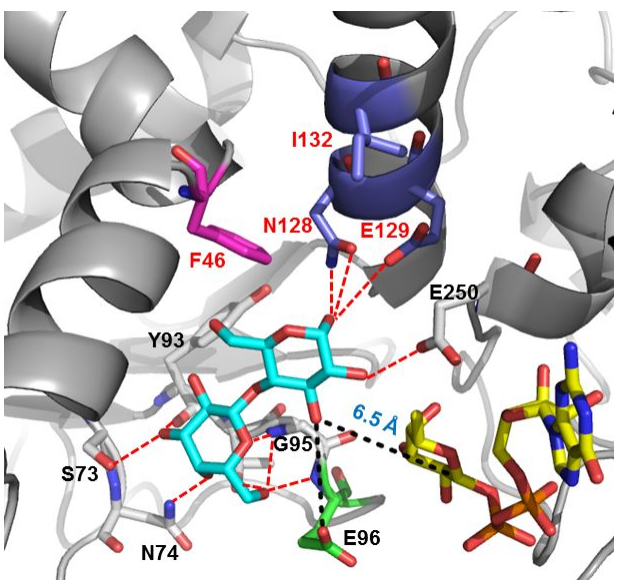

I132
N128
E129
F46
E250
Y93
6.5 Å
G95
S73
E96
N74

Figure 4.17 Structure modeling and docking simulation of lactose for the double (A128N/H129E) and quadruple (A128N/H129E/Y132I/S46F) mutants of Δ52 FutA

(A) For the double mutant, A128N and H129E mutations located on α5 helix are presented in blue color. (B) In the quadruple mutant, A128N, H129E, and Y132I mutations on α5 helix and S46F mutation on α2/loop are presented in blue and magenta, respectively. The newly generated H-bonds between the docked lactose (cyan) and corresponding residues are displayed with a red dotted line.

The H-bond interactions would stabilize the lactose binding to situate the lactose toward the interior of the lactose binding site, and the orientation of lactose can also make lactose come closer to GDP-Fuc compared to the wild-type by decreasing the distance from O3 of glucose to C1 of fucose for a nucleophilic attack. As the lactose moves into a deep binding site, the lactose in the active site keeps the H-bonding interactions with S46, S73, N74, Y93, and G95 residues, while the lactose in wild-type has H-bonds with W34, W35, E42, N74 and G95 that are located in the exterior of the lactose binding site.

Along with the A128N/H129E mutation on the α5 helix, S46F on the α2 helix/loop was responsible for the increase in both k_{cat} and the binding affinity for lactose according to the Table 4.4. The resulting triple mutant A128N/H129E/S46F has a substantial effect on the increase in k_{cat} for GDP-Fuc, like that of the addition of the H129E mutation. As a result, the A128N/H129E/S46F triple mutant showed an increase in the k_{cat}/K_m values by a factor of 17 and 25 for lactose and GDP-Fuc relative to wild-type. On the other hand, the Y132I mutation only affected a slight increase in the binding affinities of the two substrates for the A128N/H129E/Y132I mutant. As expected, combining all mutations (A128N/H129E/Y132I/S46F) on the two α-helices showed a higher activity than that of triple mutants (A128N/H129E/S46F and A128N/H129E/Y132I), resulting in an improvement by a factor of 20 and 27 in the k_{cat}/K_m values for lactose and GDP-Fuc, respectively, compared to wild-type. These results indicated that the two α-helices (α5 and α2/loop) work cooperatively for lactose binding and catalytic efficiency, and in particular, mutations such as A128N, H129E, and S46F located in the deep cleft of

the acceptor binding site are critical to improve the lactose binding and catalytic efficiency of both substrates.

As illustrated in Figure. 4.17 B, in the case of the quadruple mutant (A128N/H129E/Y132I/S46F), the glucose moiety can be more situated toward GDP-Fuc by the torsion angle of the glycosidic bond of lactose generating a new H-bond between Glu250 and the O2 of glucose. The movement in lactose is likely to be a result of the substitution with the bulky side chain of Phe at the 46 site and the abolishment of H-bonds with Ser46. The altered orientation of the lactose could make O3 of glucose come closer to C1 of fucose (6.5 Å) as well as to Oε2 of the Glu96 key residue. The decreased distances from lactose to GDP-Fuc would contribute to the stabilization of the oxocarbenium ion-like transition state, resulting in an increase in turnover rates (k_{cat}) for both substrates. In addition, the lowest docking energy (E = -5.93 kcal/mol) for lactose of the quadruple mutant among the variants suggests that a more stable lactose binding to the active site can be generated by the four mutations. In the case of Y132I, the substituted Ile132 cannot interact with glucose while the Tyr132 has a weak H-bond with O5 of glucose in the A128N/H129E mutant, indicating that the mutation into Ile132 might help lactose become more stabilized in the A128N/H129E/Y132I/S46F mutant.

4.6 Acceptor substrate specificity of Δ52 c mutants

Since the wild-type (Δ52 FutA) originally showed a much more favored acceptor substrate specificity for LacNAc over lactose, we also evaluated the substrate specificity of the mutants for LacNAc.

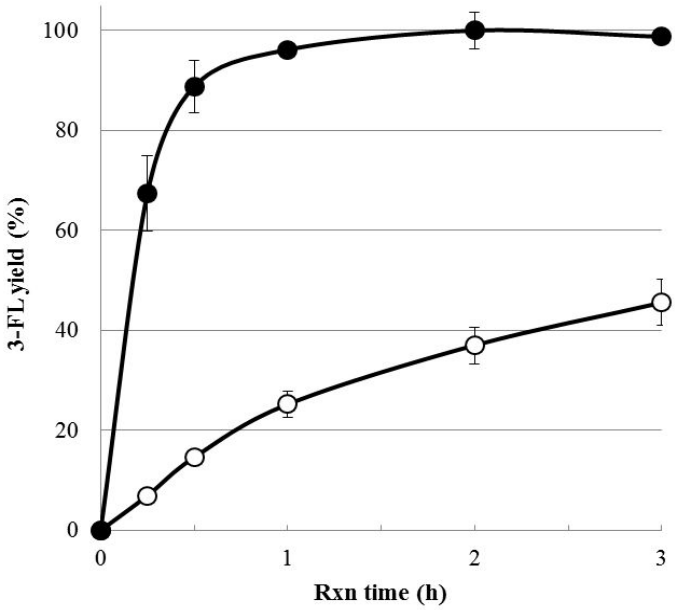
As shown in Table 4.3, all mutants also exhibited an increase in the specific activity for the LacNAc acceptor, and in particular, the triple mutant A128N/H129E/S46F showed the highest increase in the specific activity (by a factor of 8.3) relative to wild-type. The quadruple A128N/H129E/Y132I/S46F mutant increased the specific activity for 3-FL and Le^x by a factor of 15.2 and 7.5, respectively, relative to wild-type Δ52 FutA. In terms of the specific activities for 3-FL and Le^x, the ratio of 3-FL to Le^x for the wild-type was 0.047 while that for the A128N/H129E/Y132I/S46F mutant was 0.094, indicating a two times increase in the ratio of the substrate specificity for lactose over LacNAc. Although the relative specific activity for lactose of the A128N/H129E/Y132I/S46F mutant is still lower than that for LacNAc, the absolute value is almost comparable to the specific activity of LacNAc for the wild-type (Δ52 FutA), suggesting that the mutant can be efficiently used for the 3-FL synthesis.

4.7 Synthesis of 3-FL and Le^x

4.7.1 3-FL synthesis using quadruple mutant

Since the Δ52 FutA A128N/H129E/Y132I/S46F mutant showed the highest specific activity for lactose, the mutant was applied to the 3-FL production in order to improve the yield and productivity. When 10% v/v of cell extract of the mutant was used, a 96% 3-FL conversion yield was achieved after 1 h based on 5 mM GDP-Fuc (Figure. 4.18 A), and the yield reached 100% after 2 h. In contrast, the 3-FL yield for the wild-type (Δ52 FutA) was 45% after 3 h under the same conditions even though its soluble expression level was similar to that of the mutant.

A


100
80
60
40
20
0
3-FL yield (%)
0
1
2
3
Rxn time (h)

B

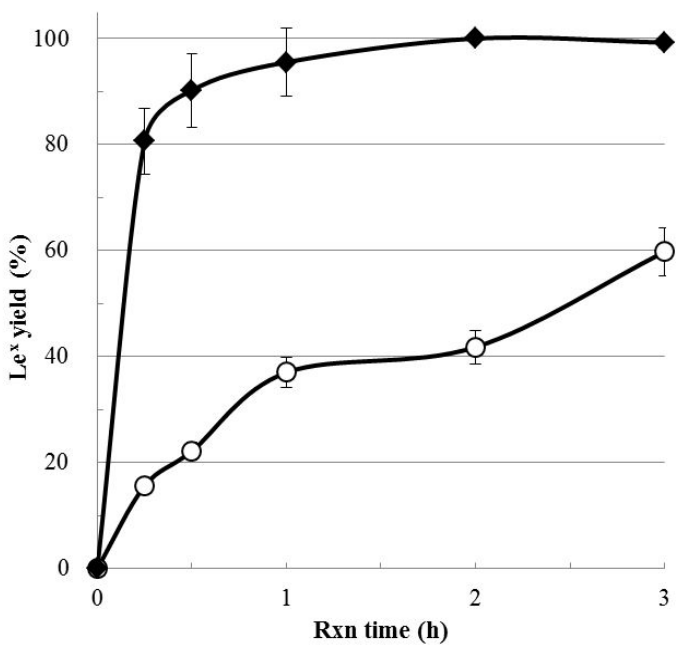

100
80
60
40
20
0
Le^x yield (%)
0
1
2
3
Rxn time (h)

Figure 4.18 Increased yield and productivity for the production of 3-FL (A) and Le^x (B) by Δ52 FutA mutants (A128N/H129E/S46F and A128N/H129E/Y132I/S46F)

3-FL (A) and Le^x (B) synthesis reactions were performed with 5 mM GDP-Fuc, 5 mM $MgCl_2$, 15 mM lactose or 5 mM LacNAc, and crude extract of the enzymes (10% v/v for the 3-FL, and 5% v/v for the Lex) in 50 mM sodium phosphate buffer (pH 7.6) at 37 °C. Δ52 FutA A128N/H129E/Y132I/S46F (closed circles) and Δ52 FutA A128N/H129E/S46F (closed diamonds) mutants were used to produce 3-FL and Le^x, respectively, in order to compare the yields against those of the wild-type Δ52 FutA (open circles). The yields of 3-FL and Le^x are determined according to the initial GDP-Fuc concentration, and mean values are shown from triple experiments with error bars.

4.7.2 Le^x synthesis using triple mutant

The Δ52 FutA A128N/H129E/S46F mutant with the highest specific activity for LacNAc was investigated for the Le^x production with 5 mM GDP-Fuc and 5 mM LacNAc using 5% v/v of soluble enzymes. As shown in Fig. 4. 18 B, the Le^x yield based on the GDP-Fuc improved to 95% after 1 h with the mutant while the yield of the wild-type was 60% after 3 h. As the results show, the A128N/H129E/Y132I/S46F and A128N/H129E/S46F mutants of Δ52 FutA substantially increased the yield and productivity for the production of 3-FL as well as Le^x, indicating that the mutants could be used efficiently to mass produce various 3-fucosyloligosaccharides possessing the lactose or LacNAc backbone.

4.8 Discussion

This is the first protein engineering study of FucT that belongs to GT-B fold family to improve the catalytic activity toward poor substrate lactose. In this work, FutA (α1,3-FucT from *H.pylori*) was used in protein engineering strategies to increase protein solubility and its catalytic activity as well as to change the substrate specificity for the synthesis of 3-FL. GT-B fold enzymes have distinct substrate binding domains for donor and acceptor substrates, and two domains are facing each other with the active site in the center. Therefore, modeling and docking substrates into the predicted substrate binding sites of the enzyme based on the key residues in the active site would become important clues for the mutagenesis to achieve better performance. This study consisted of a focused directed evolution that was conducted by combining the structure-guided analysis of active site/substrate binding site and clustered ISM considering the locations of

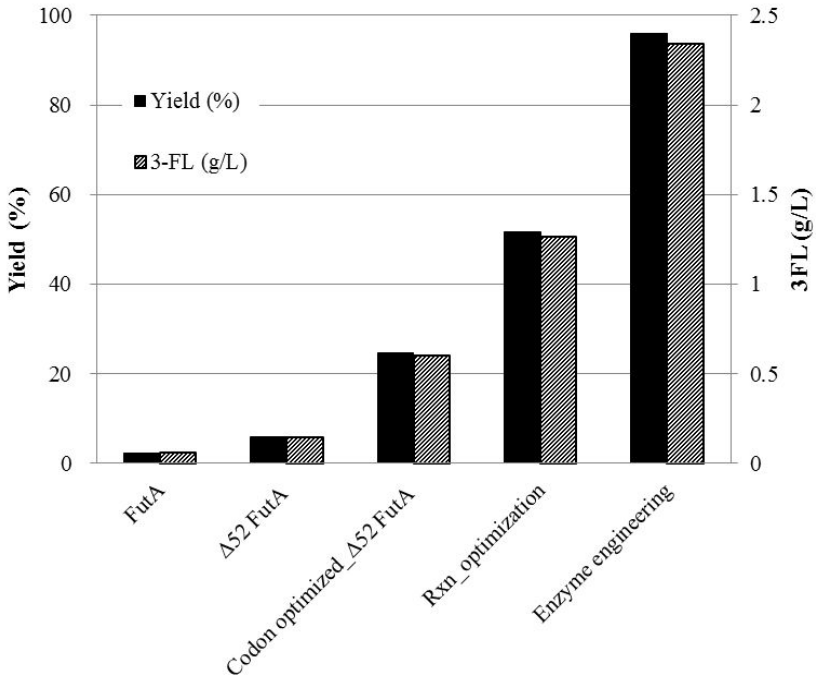
the α-helices. The functional residues that were selected through modeling, substrate docking, and HotSpot Wizard tool were subjected to SSM and structure-guided ISM, in order. As a result, we could efficiently screen mutants with enhanced activity from small and focused libraries. The Δ52 FutA quadruple mutant (A128N/H129E/Y132I/S46F) generated in this study displayed a 15.2-fold increase in specific activity for 3-FL synthesis relative to that of wild-type (Δ52 FutA). The mutations are mainly located on the two α-helices (i.e., α5 and α2/loop) that are situated in the deep cleft of the lactose binding site rather than at the entrance of the lactose access tunnel or GDP-Fuc binding site. The results indicate that such mutations could make lactose become situated inside the active site to generate a more stable and reactive conformation by forming new H-bonds with the lactose substrate via cooperative action of the two α-helices. Although the mutations are mainly located in the lactose binding site, the mutations could also affect the increase in the turnover rate of the GDP-Fuc caused by the stabilization of the transition-state complex, which would be triggered by inducing a closed conformation upon the binding of GDP-Fuc (Qasba et al. 2005).

The results indicate that the structure-guided ISM from the 'best hit' mutant is a powerful strategy to reduce the screening effort and to effectively improve the enzyme properties for the desired targets. In addition, in order to select functional residues for SSM and ISM, HotSpot Wizard can be a useful tool to predict the substrate binding site and mutability in company with a docking simulation of the enzyme. This strategy can be applied to other GTs with a GT-B fold structure.

In the production of 3-FL using 10% v/v of the cell-free extracts, the Δ52 FutA A128N/H129E/Y132I/S46F mutant showed a substantial increase in the 3-FL

yield (i.e., 96% based on 5 mM GDP-Fuc) after 1 h while Δ52 FutA and the original FutA achieved only a 45% yield after 3 h and a 2.4% yield after 14 h, respectively (Figure 4.19). At the same time, all mutants generated in this study exhibited an increase in activity for LacNAc as well, with the Δ52 FutA A128N/H129E/S46F mutant showing an increase by a factor of 8.3 in the specific activity for LacNAc relative to wild-type (Δ52 FutA). The results indicate that the mutants generated in this study are invaluable catalysts for the production of various 3-fucosyloligosaccharides containing 3-FL and Le^x.

In summary, this work will become a good example for the development of mutants of GT-B fold enzymes including FucTs through focused directed evolution for engineered substrate specificities and improved catalytic activity for mass production of useful oligosaccharides.



1

	3FL (g/L)	Reaction (h)	Productivity (g/L/h)	Rxn Cond'n
FutA	**0.06**	**14**	**0.004**	5 mM GDP-Fuc + 2.5 mM Lac + 30% v/v 3FT
Δ52 FutA	**0.15**	**14**	**0.011**	5 mM GDP-Fuc + 2.5 mM Lac + 30% v/v 3FT
Codon optimized_ Δ52 FutA	**0.60**	**15**	**0.040**	5 mM GDP-Fuc + 5 mM Lac + 20% v/v 3FT
Rxn optimization	**1.11**	**3**	**0.371**	5 mM GDP-Fuc + **15 mM Lac** + 10% v/v 3FT **(TB)**
Enzyme engineering	**2.35**	**1**	**2.35**	5 mM GDP-Fuc + **15 mM Lac** + 10% v/v 3FT **(TB)**

2

3

Figure 4.19 Improvement of 3-FL yield as increase in protein solubility and catalytic activity via protein engineering

FutA, original FutA from *H. pylori*; Δ52 FutA, C-terminus truncated FutA; Codon optimized_Δ52 FutA, C-terminus truncated FutA followed by codon optimization; Rxn optimization, reaction optimization for lactose concentration using codon optimized Δ52 FutA; Enzyme engineering, enzyme engineering for codon optimized Δ52 FutA

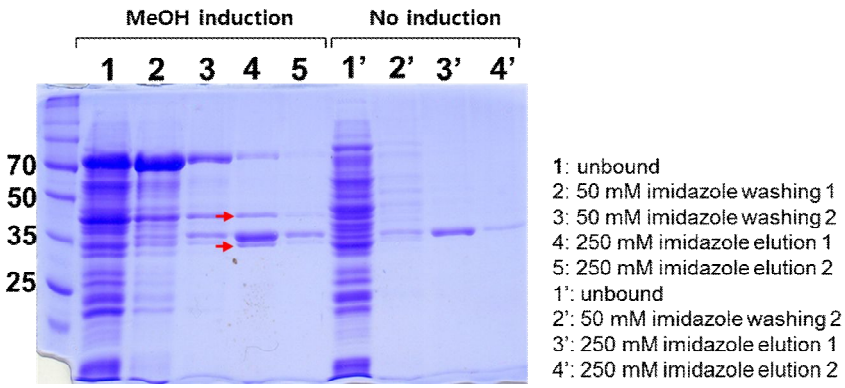
Chapter 5.

Production of 3-fucosyllactose in metabolically engineered *Escherichia coli* cells

5.1 *in vivo* production of GDP-Fuc in yeast

GMD and GFS were expressed in *P. pastoris* and *S. cerevisiae* as shown in Figure 5.1. When the purified enzymes from the cultured yeast cells were used for the *in vitro* synthesis of GDP-Fuc using the GDP-Man, GDP-Fuc was produced as a function of time (data not shown). The result indicates that the GMD and GFS were expressed in yeast strains as an active form. As shown in Figure 5.2, GDP-Fuc concentration of yeast cells harboring both GMD and GFS was higher than that of yeast cell harboring only GMD or GFS. Using the GMD and GFS, *S. cerevisiae* produced 0.21 mg/L (culture) of GDP-Fuc, whereas *P. pastoris* produced 11 mg/L of GDP-Fuc. To increase the yield of GDP-Fuc in *S. cerevisiae*, 1% of galactose was added every 24 h after induction with 4% galactose media. As shown in Figure 5.3, expression level of GMD and GFS was increased when 1% of galactose was added more. However, expression levels of the genes were noticeably decreased after 4 days, which might be due to the death phase of the *S. cerevisiae* cell after 4 days (Figure 5.3 A). As a result, GDP-Fuc production was also improved when 1% of galactose was more added, which resulted in 2.12 mg/L (culture) of GDP-Fuc after 3 days culture (Figure 5.3 B). The yield of GDP-Fuc was 31.6% based on the GDP-Man concentration detected at before induction (1 day). GDP-Man was almost consumed after 2 days, which indicates that GDP-Man would be consumed by other metabolic pathway during the cultivation. Therefore, we have to overexpress genes related to produce GDP-Man and knockout genes responsible for conversion of GDP-Man to other metabolites, to increase the production of GDP-Fuc in *S. cerevisiae*.

A



B

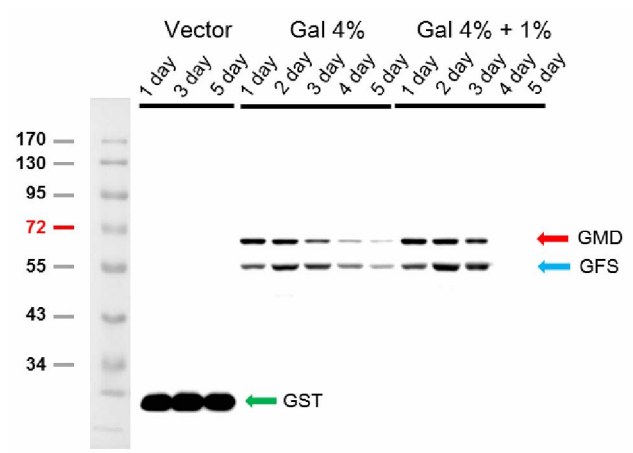
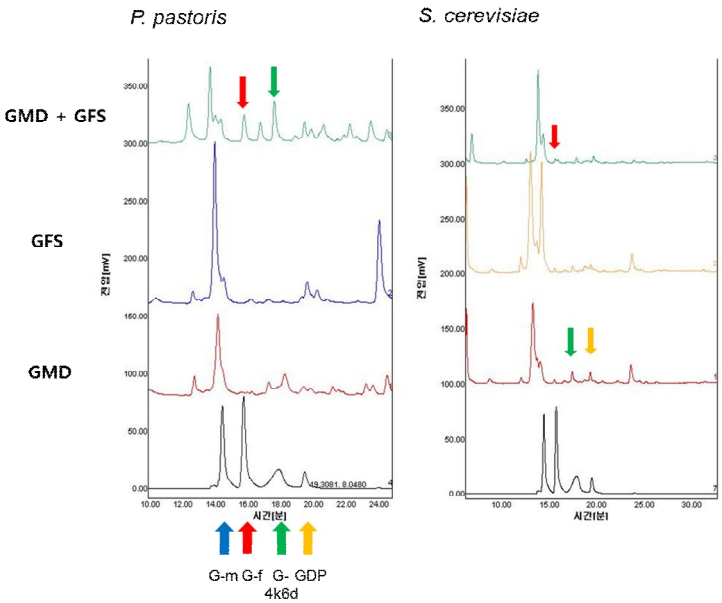


Figure 5.1 Expression of GMD and GFS in yeast strains

GMD (42 kDa) and GFS (34 kDa) expressed in *P. pastoris* were identified by Ni-NTA purification (A). GST-fused GMD (69 kDa) and GFS (61kDa) were expressed in *S. cerevisiae*, and their expressions were identified by western blotting using GST antibody (B). At this time, same amounts of protein (20 μg) were loaded into SDS-PAGE.

A



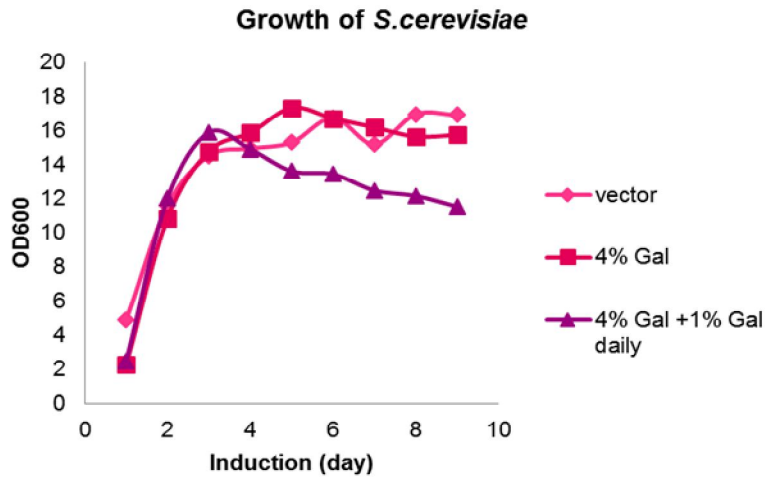
B

		OD $_{600}$	GDP-fuc	GDP-fuc
		induction	(mg/ 100 mL culture)	(mg/ 1L culture)
S. cerevisiae	GMD	15	1.50E-02	0.15
	GFS	15.05	1.60E-02	0.16
	GMD+GFS	11.86	2.10E-02	0.21
P. pastoris	GMD	11.11	1.90E-01	1.90
	GFS	9.48	1.50E-01	1.50
	GMD+GFS	10.72	1.10E+00	11.00

Figure 5.2 Production of GDP-Fuc in *P. pastoris* and *S. cerevisiae*

Production of GDP-Fuc for yeast cells harboring GMD and GFS was compared with yeast cells harboring only GMD or GFS (A). Produced GDP-Fuc were quantified by L culture of yeast cells.

A



B

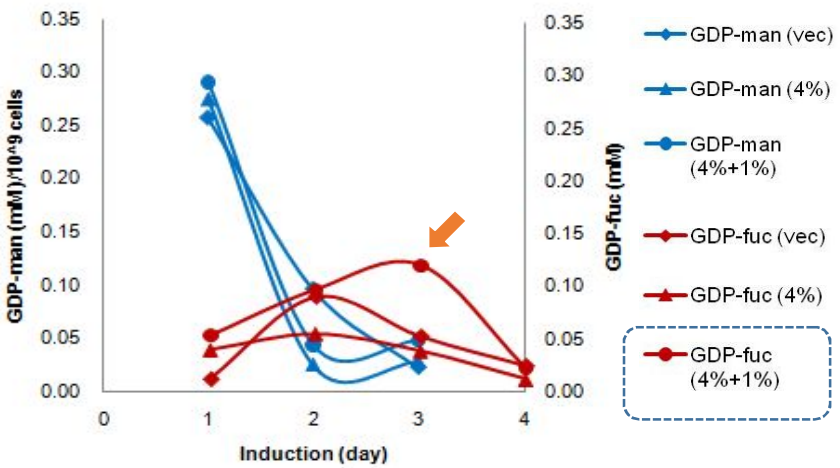


Figure 5.3 Production of GDP-Fuc in *S. cerevisiae* using galactose induction

Cell growth (A) and concentration of GDP-Man/GDP-Fuc (B) for *S. cerevisiae* cell with addition of 1% galactose (4% Gal + 1% Gal) were compared with cells that cultured in 4% galactose media (4% Gal).

Even though metabolic engineering of the genes in endogenous or newly introduced pathway could be conducted to increase the GDP-Man, the level of GDP-Man in yeast would be low to produce g scale of GDP-Fuc. Therefore, we have produced GDP-Fuc from glycerol or L-fucose using *de novo* and/or *salvage* pathway in *E. coli* system.

5.2 Construction of *E. coli* strain for the production of 3-FL

5.2.1 Construction of *E. coli* for *in vivo* production of GDP-Fuc

For *in vivo* production of 3-FL, *E. coli* BW25113 (DE3) was chosen as a parent strain. *E. coli* BW25113 (DE3) derived from *E. coli* K-12 is a β-galactosidase (*lacZ*)-negative but lactose permease (*lacY*) positive strain, which allows uptake of lactose into the cell and block the lactose degradation. To produce GDP-Fuc in *E. coli*, *de novo* and salvage pathway were constructed, respectively. Firstly, to introduce *fkp* of salvage pathway into *E. coli* as a plasmid expression, *fucIK* gene locus was deleted to block the degradation of L-fucose, which should be supplemented from extracellular media. As a result, pC-F and pJ401-FutA vectors harboring *fkp* and Δ52 FutA quadruple mutant, respectively, were co-expressed in BW ΔfucIK.

On the other hand, genes, which are capable of synthesis of GDP-Fuc by *de novo* pathway were cloned into pCDFm vector. As shown in Figure 1.8, GDP-Fuc is synthesized from GDP-Man via GMD and WcaG (GFS). GDP-Man is produced from fructose-6-P by three reactions carried out by ManA, ManB, and ManC. Among the genes, ManB, ManC, Gmd, and WcaG were cloned into

pCDFm vector to allow co-expression with pJ401-FutA for the production of 3-FL.

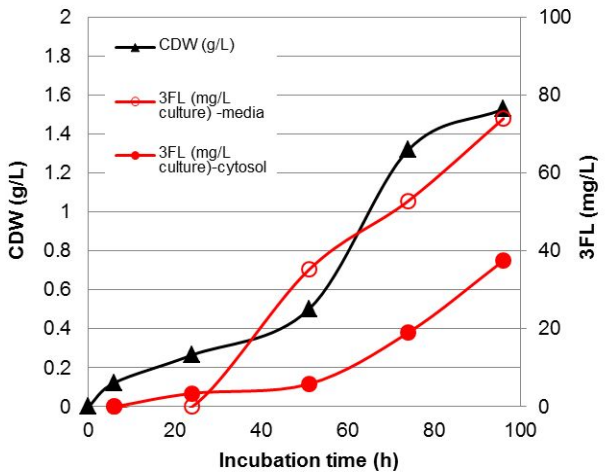
5.2.2 Batch and fed-batch culture in minimal media using salvage pathway

BW ΔfucIK harboring pC-F and pJ401-FutA was used for the production of 3-FL in minimal media supplemented with 5 g/L of glycerol. L-fucose and lactose substrates, which were added at the induction time, could transfer into the cells through inherent lactose permease (*lacY*) and glycerol facilitator (*GlpF*). For the batch culture, total 110 mg per L culture (74 mg/L in media and 37 mg/L in cytosol) of 3-FL was produced after 96 h as shown at Figure 5.4 A. To increase the production of 3-FL, fed-batch culture was roughly conducted in minimal media using glycerol as carbon source. As a result, 152 mg/L of 3-FL (130 mg/L in media and 22 mg/L in cytosol) was produced after 169 h (Figure 5.4 B). Instead of using glycerol as sole carbon source, glucose was used at initial growth until the OD_{600} reached 0.7, and then glycerol was used after induction time. Resultingly, 192 mg/L of 3-FL (40 mg/L in media and 152 mg/L in cytosol) was produced after 24 h, but 3-FL was not increased anymore and produced 3-FL was secreted after 24 h (data not shown).

5.2.3 Media optimization

In the co-expression system using pC-F and pJ4-1-FutA in the BW ΔfucIK cell, the soluble expression of FutA was very low in minimal media as shown in Figure 5.5 A.

A



B

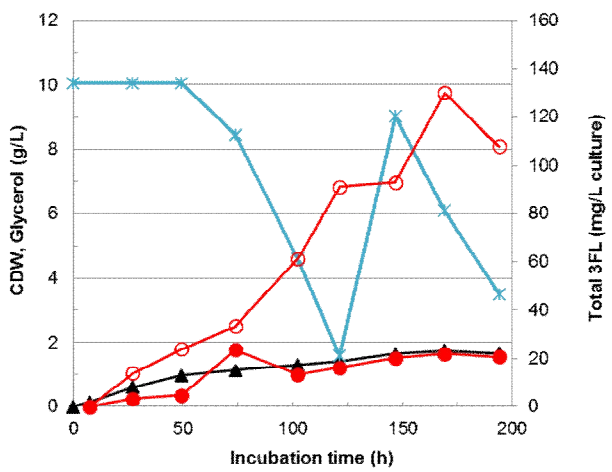
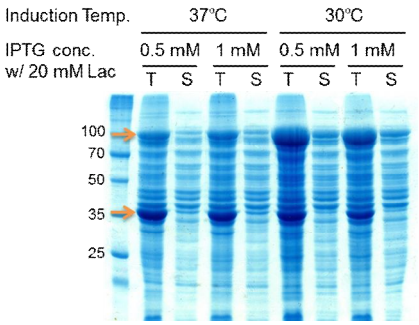


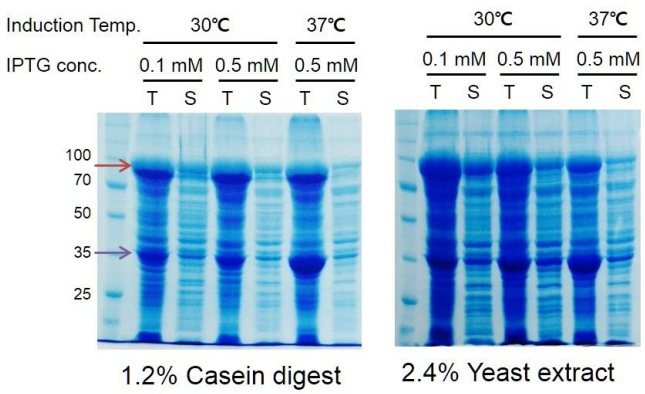
Figure 5.4 Production of 3-FL in BW ΔfucIK harboring FKP and Δ52 FutA mutant in minimal media

3-FL was produced by batch culture (A) and fed-batch culture (B) using glycerol as C-source. Open circle, extracellular 3-FL; closed circle, intracellular 3-FL; closed triangle, dry cell weight; x, glycerol concentration

A



B



C

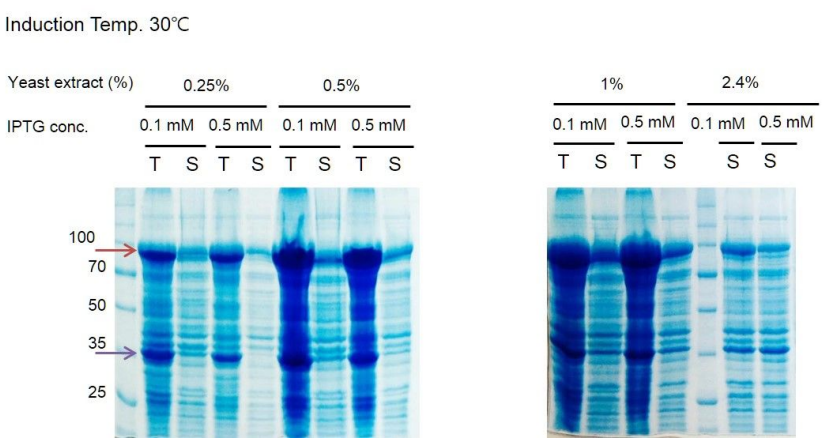


Figure 5.5 Media optimization to increase the solubility of FutA

To increase the level of soluble expression of FutA, media optimization was conducted. In our previous study, total and soluble expression of FutA substantially increased in TB broth, which contains 1.2% casein digest and 2.4% yeast extract. In order to identify which component increased the solubility, 1.2% of casein digest and 2.4% of yeast extract were added in minimal media, respectively, for 30 ºC and 37 ºC induction. As a result, addition of 2.4% yeast extract increased solubility of FutA at 30 ºC as shown in Figure 5.5 B. Since culture in TB broth did not increase 3-FL production in previous study even though solubility of FutA increased, concentration of yeast extract in minimal media was optimized. Considering the soluble expression of FutA, 1% yeast extract was most effective among 0.25% to 2.4% concentration (Figure 5.5 C). In accordance with yeast extract, 1 mM of Mn, which is cofactor of fucose-1-phosphate guanylyltransferase reaction of FKP, was added into the optimized media. In addition, lactose concentration (i.e. 20 and 40 mM) was tested; there is little difference for 3-FL production. As shown at Figure 5.6, 3-FL production was increased up to 403 mg/L (378 mg/L in media and 25 mg/L in cytosol) in optimized media containing 1% yeast extract, 1 mM $MnSO_4$, and 40 mM lactose. We confirmed that 3-FL production increases as other sugars in yeast extract were depleted, which made us assume that sugars in yeast extract could inhibit 3-FL production. Therefore, same ratios of amino acids contained in 2.4% yeast extract were added in the minimal media instead of yeast extract. However, addition of amino acids did not increase the yield and productivity of 3-FL (data not shown).

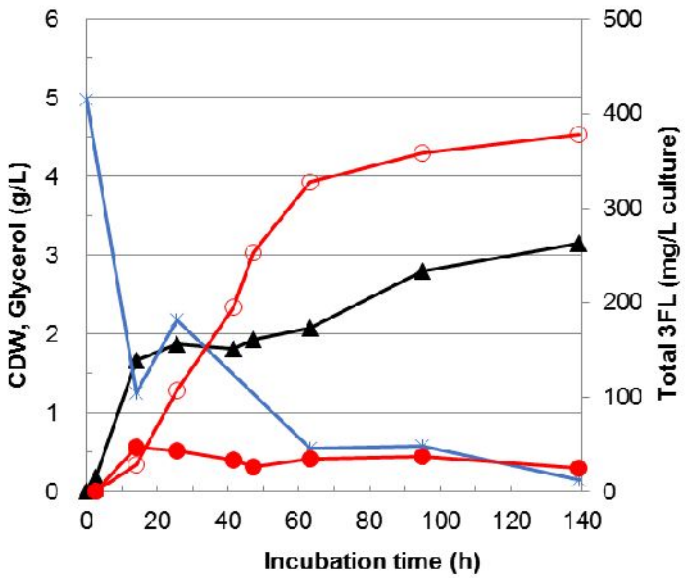


Figure 5.6 3-FL production in optimized media

3-FL was produced by fed-batch culture using glycerol as C-source. Open circle, extracellular 3-FL; closed circle, intracellular 3-FL; closed triangle, dry cell weight; x, glycerol concentration

5.3 Increase in production of GDP-Fuc

5.3.1 Determination of rate determining step

To determine the rate limiting step of 3-FL produciton in *E. coli* system, intermediate substrates in cytosol of cultured cells for fed-batch culture were quntitatively analysed by HPLC. For all samples, little (<< 1 μM) or no accumulation of GDP-Fuc intermediate was identified in cytosol (Figure 5.7 A). Compared kinetic parameter of FKP with Δ52 FutA quadruple mutant, the value of FKP is much higher than that of Δ52 FutA mutant. In addition, 3-FL yield for batch culture using the quadruple mutant of FutA was only two-times higher than that for batch culture using the wild-type of FutA, although the specific activity of the mutant is higher than wild-type by 15-fold. These results indicate that GDP-Fuc production is a rate limiting step for 3-FL production in *E. coli* (Figure 5.7 D).

When we detected GMD, GDP, and GTP in cytosol of the cultured cell, small amount of GMP (< 20 μM) was detected only at initial sample, and little or no GDP and GTP was detected (Figure 5.7 C). It can be deduced that produced GTP from *E. coli* metabolism was used to rapidly produce GDP-Fuc through fucose-1-phosphate guanylyltransferase reaction of FKP. And the result elucidates that bottleneck of GDP-Fuc production would be due to the lack of nucleoside phosphate substrates such as GTP supplemented from *E. coli* metabolism. In addition, very low amount of L-fucose (<< 2 μM) detected in cytosol would reflect the low level of uptake of L-fucose and affect the limitation of GDP-Fuc synthesis (Figure 5.7 B).

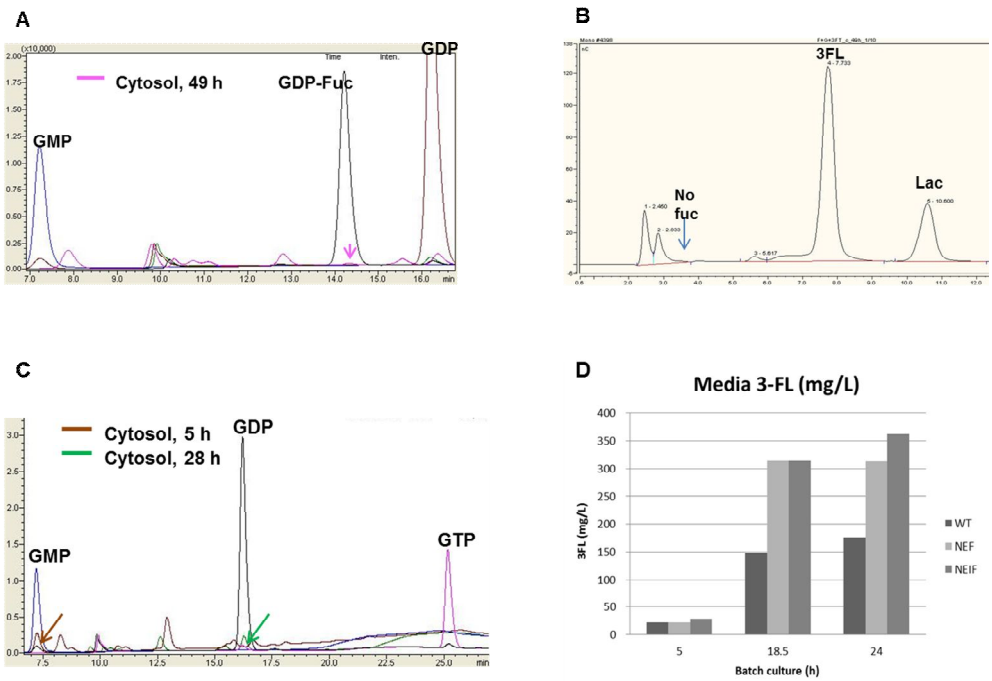


Figure 5.7 Determination of rate limiting step for 3-FL production

GDP-Fuc (A), Fuc (B), GMP, GDP, and GTP (C) in cytosol of cultured *E. coli* cells were quantitatively determined by HPLC. 3-FL production in *E. coli* cells by overexpression of mutants of FutA were compared with that of wild-type (D).

5.3.2 Construction various vector systems to increase GDP-Fuc

Based on the quantitative analysis of GMP, GDP, and GTP, we assumed that synthesis of these metabolite would be rate determining step for the production of 3-FL. For the production of GDP-Fuc, Lee et al. (2012) focused on the increase in GTP for ManC reaction capable of conversion of mannose 1-phosphate into GDP-Man by *de novo* pathway. Recently, Zhai et al. (2015) enhanced GDP-Fuc production via increase in GTP by combinatorial expression of genes for salvage pathway.

In this study, we have constructed combinatorial vector systems for increasing of GDP-Fuc as illustrated at Figure 5.8 A. In *de novo* pathway, ManB, ManC, Gmd, and WcaG were used, whereas, in salvage pathway, FKP, Gsk, Gpt, Gmk, and Ndk were used for the production GDP-Fuc. Therefore, pC-BCGW, pC-BCGW+G, pC-F, pC-FG, pC-GpGmNF, and pC-BCGW+G+F constructions were generated (Table 2.4). The genes of *de novo* and salvage pathway were combined, as well. When these constructions were used for GDP-Fuc synthesis by batch culture, pC-FG carrying FKP and Gsk showed the highest GDP-Fuc production per g of dry cell weight (Figure 5.8 B). The GDP-Fuc synthesis for the constructions of salvage pathway were higher than those of de novo pathway. In the case of pC-GpGmNF, several genes (Gpt, Gmk, Ndk, and FKP) cloned into one pCDFm vector affected overall amount of protein expression, according to the SDS-PAGE (data not shown) and resulted in decrease in GDP-Fuc production. Along with the Gsk expression, Ndk capable of conversion for GDP into GTP was added downstream at Gsk to allow construction of dicistron with Gsk to increase the GTP pool (Figure 5.9).

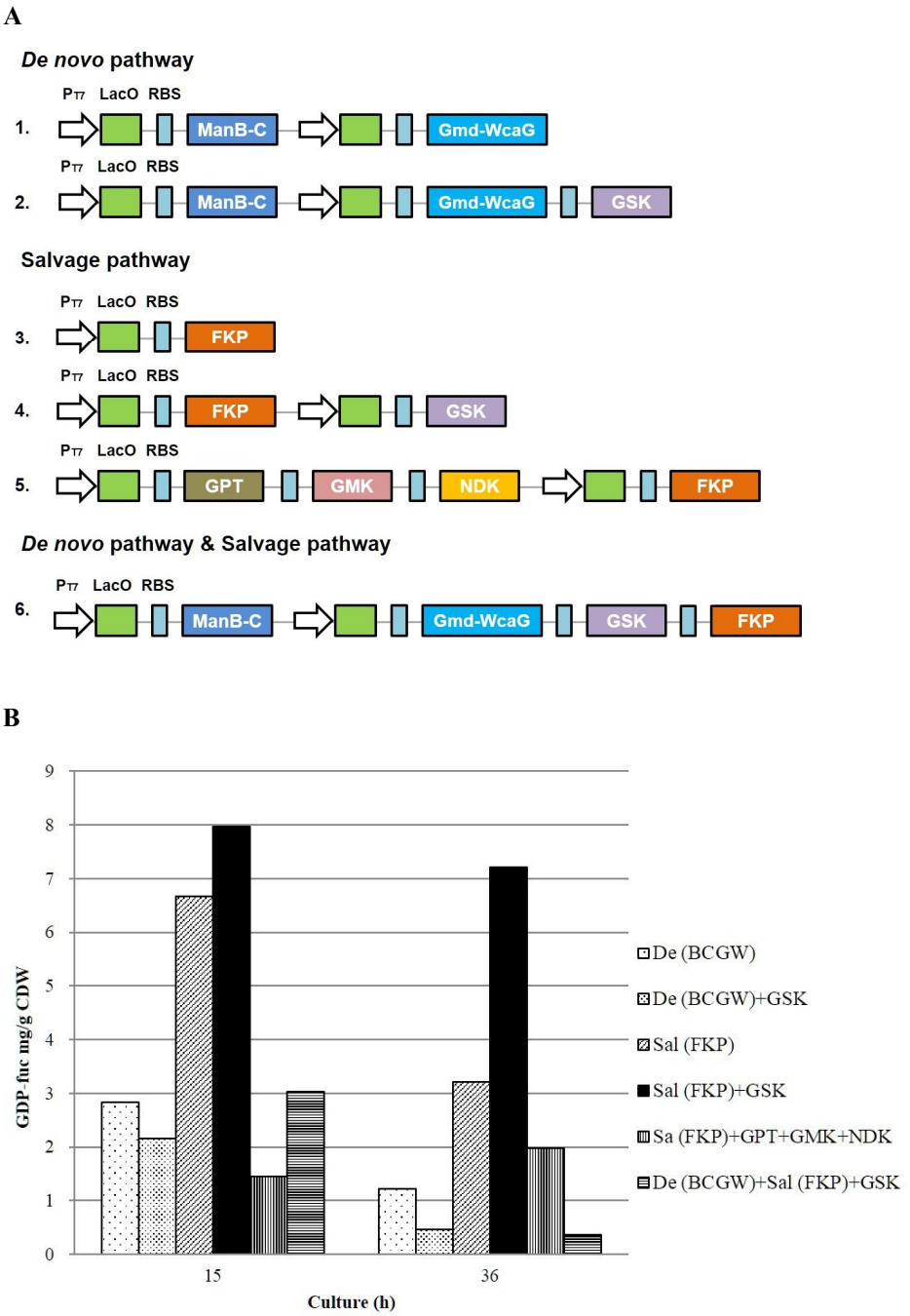
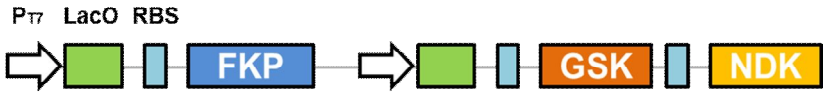


Figure 5.8 Construction of vector sets to increase GDP-Fuc



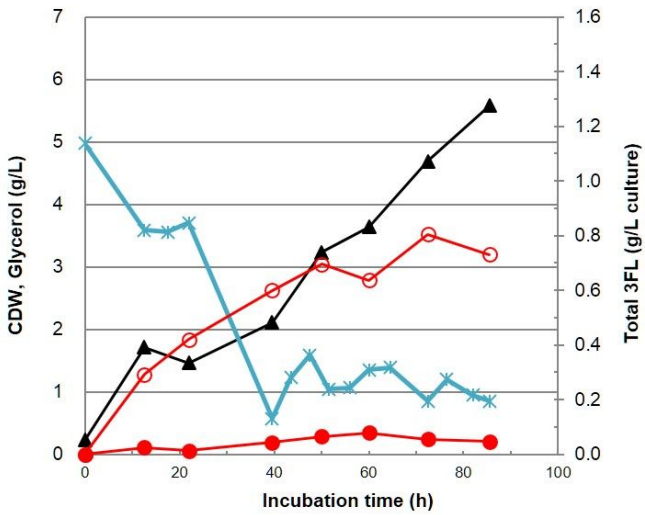


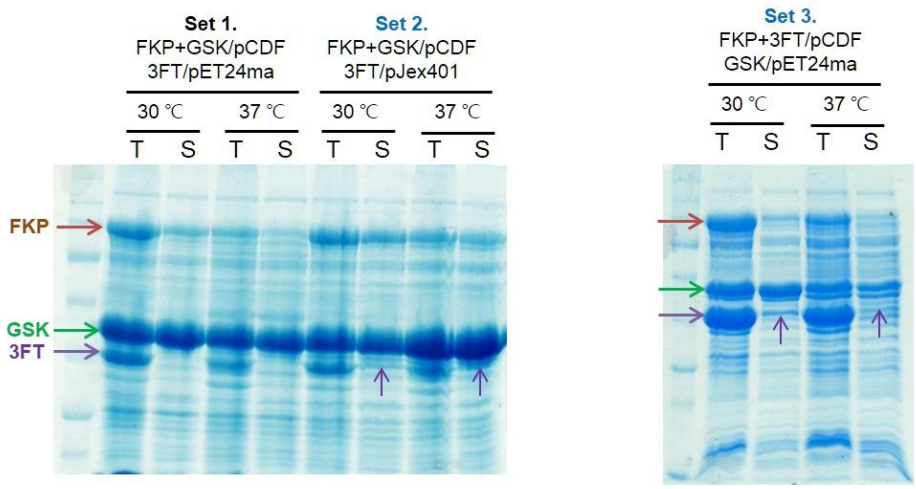
Figure 5.9 3-FL production in BL21 ΔlacZ ΔfucIK strain harboring FKP, Gsk, Ndk, and Δ52 FutA mutant

We found that Ndk was expressed well, but 3-FL production yield was not increased, unexpectedly, resulting in 860 mg/L 0f 3-FL in 72.5 h.

Therefore, the construction of FKP and Gsk was the most effective for the production of GDP-Fuc and 3-FL. To investigate the optimal combination of *fkp*, *gsk*, and *futA* genes regarding protein expression and 3-FL production, several sets of vectors were constructed as shown in Figure 5.10. As a result, set 2 (pCDFm carrying *fkp*+*gsk* and pJ401 carrying Δ52 *futA* quadruple mutant) and set 3 (pCDFm carrying *fkp*+Δ52 *futA* quadruple mutant and pET24ma carrying *gsk*) showed higher level of soluble expression for FutA than set 1 (pCDFm carrying *fkp*+*gsk* and pET24ma carrying Δ52 *futA* quadruple mutant). Therefore, 3-FL production of set 2 and set3 was compared as shown at Figure 5.11. The set 2 and set 3 accomplished the 1 g/L of 3-FL after 100 h. And the productivity of set 2 was higher than set 3, resulting in 1.05 g/L of 3-FL production after 102 h.

5.4 Comparison of 3-FL production in *E. coli* BL21 (DE3) and BW25113 (DE3)

The acetate production pattern of *E. coli* B (BL21) is different from that of *E. coli* K-12 especially when the bacteria grow to high densities at high glucose concentration. In the metabolic pathway of *E. coli* BL21, all the following pathways (glyoxylate shunt, TCA cycle, gluconeogenesis pathway, glycogen biosynthesis, acetate converstion to acetyl CoA, oxaloacetate conversion to phosphoenolpyruvate, and fatty acid degradation)were active indepentently on glucose feeding strategy. *E. coli* JM109 unlike in *E. coli* BL21 was dependent on the glucose feeding strategy: At high glucose, pathways were low or inactive



	pCDF duet - T7 promoter - 20-40 copy #	pJex401 - T5 promoter - ~40 copy#	pET24ma - T7 promoter - 10-12 copy #
Set 1	FKP + GSK		FutA
Set 2	FKP + GSK	FutA	
Set 3	FKP + 3FT		GSK

Figure 5.10 Protein expression of vector sets containing *fkp*, *gsk*, and FutA

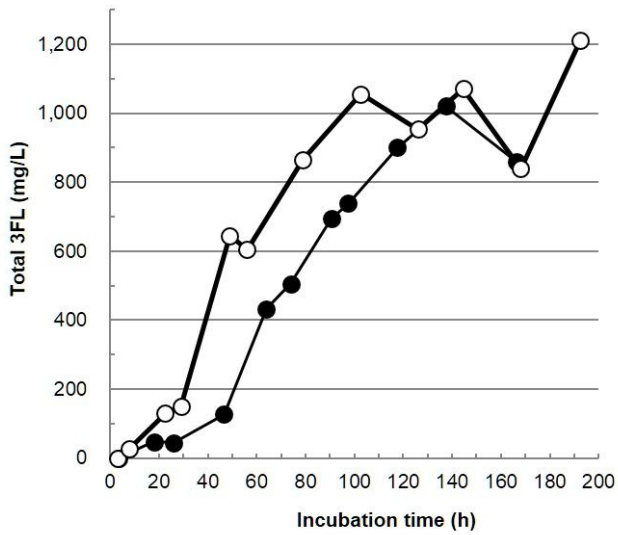


Figure 5. 11 3-FL production by fed-batch culture for set 2 and set 3

Open circle, set 2 (culture on temperature shift 30 ºC for 3 h to 37 ºC after induction); closed circle, set 3 (37 ºC culture)

exception of pyruvate conversion to acetate by pyruvate oxidase and malate conversion to pyruvate by malate dehydrogenase. At low concentration of glucose, metabolic pathways such as gluconeogenesis, fatty acid degradation, oxaloacetate conversion to phosphoenolpyruvate, acetate conversion to acetyl-CoA, and glyoxylate shunt were more active than in high glucose (Phue et al. 2005). For these reasons, *E. coli* K-12 accumulates acetate up to 11 g/L and its growth rate slows down. On the other hand, *E. coli* B accumulates acetate to about 3 g/L and its growth rate is not affected.

To increase the cell growth and productivity, we changed *E. coli* strain into *E. coli* B strain. Prior to construct 3-FL production system in *E. coli* B strain, protein expression test of set 2 was proceeded in BL21 (DE3), BL21 Star (DE3). As illustrated at Figure 5.12 A, protein expressions of FKP, Gsk, and FutA in BL21 (DE3) were higher than in BL21 Star (DE3). Using the BL21 (DE), *lacZ* gene locus was deleted from chromosome, to allow blocking of lactose degradation and change the lactose flux from the lactose utilizing pathway to the 3-FL production pathway. As a result, BL21 ΔlacZ ΔfucIK strain carrying FKP, Gsk, and ΔFutA quadruple mutant was constructed as illustrated in Figure 5.13. Using the BL21 ΔlacZ ΔfucIK strain, we compared 3-FL production with BW 25113 ΔfucIK. As a result, cell growth rate and productivity of 3-FL production of BL21 ΔlacZ ΔfucIK strain were increased relative to BW ΔfucIK (Figure 5.12 B). 3-FL yield of BL21 ΔlacZ ΔfucIK strain was 1.15 g/L in 69 h.

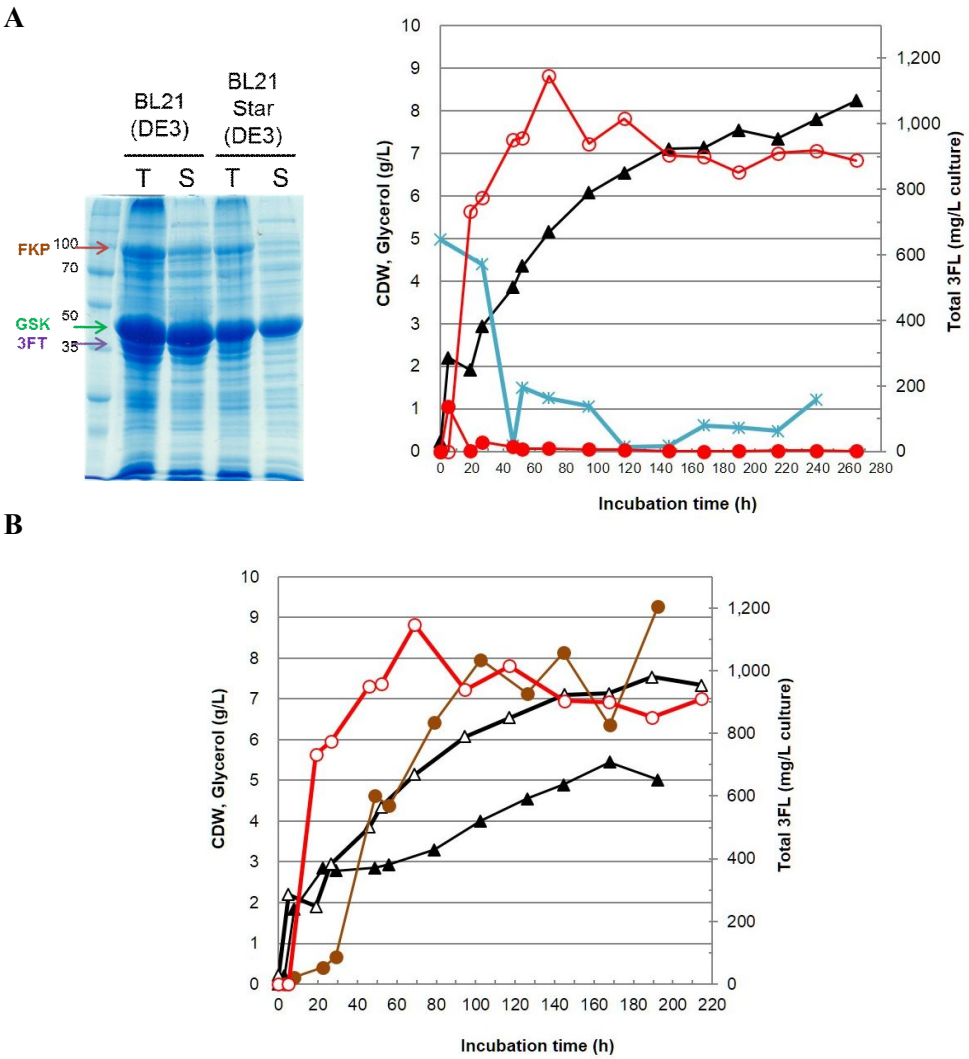


Figure 5.12 Comparison of 3-FL production in *E. coli* BL21 (DE3) and BW25113 (DE3)

3-FL production in BL21 (DE3) (A), comparison of 3-FL production in BL21 (DE3), open circle; 3-FL production in BW25113 (DE3), closed circle; dry cell weight of BL21 (DE3), open triangle; dry cell weight of BW25113 (DE3), closed triangle.

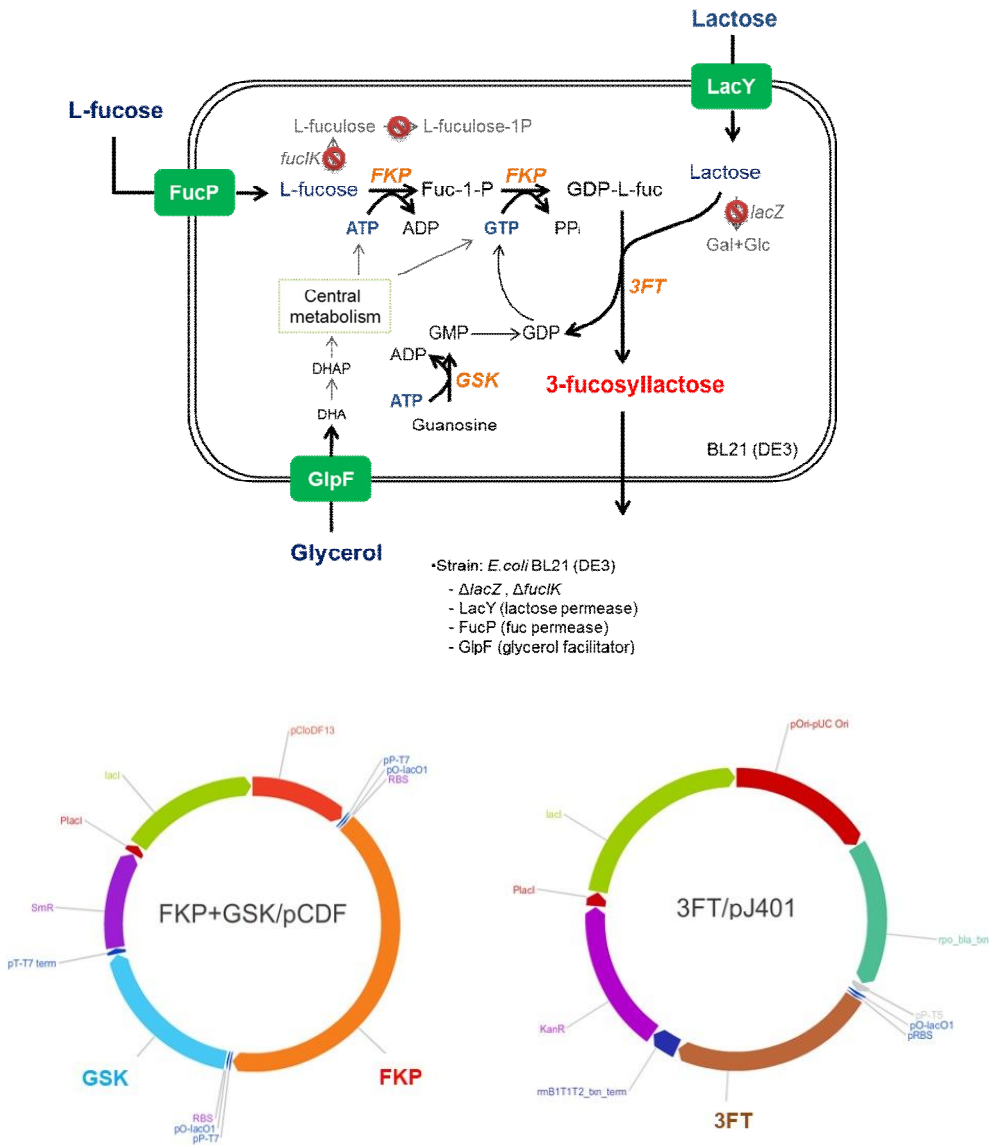
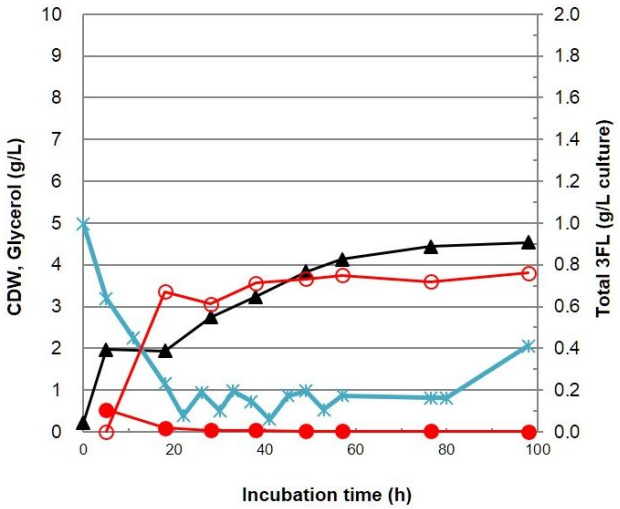


Figure 5.13 Construction of *E. coli* BL21 (DE3) strain for the production of 3-FL

Using the BL21 ΔlacZ ΔfucIK strain, we re-optimized fed-batch culture condition in flask. In previous result, set 2 showed high yield and productivity when temperature was maintained at 30 ºC for 3 h after induction, and then shifted at 37 ºC. Instead of shifting temperature, 30 ºC was maintained after induction. In addition, glycerol as a carbon source was fed into the medium for 1.5 g/L at a time to maintain below 1.5 g/L of glycerol concentration. As a result, yield and productivity were substantially increased, yielding 2 g/L of 3-FL in 76. 5 h (Figure 5.14 B). Furthermore, we reconfirmed Gsk effect for the biosynthesis of 3-FL in the newly engineered *E. coli* BL21 (DE3). As illustrated in Figure 5.14, addition of Gsk improved 3-FL production as glycerol feeding proceeded, but the strain carrying only Fkp and FutA increased 3-FL a little even though glycerol was fed as a function of time.

A



B

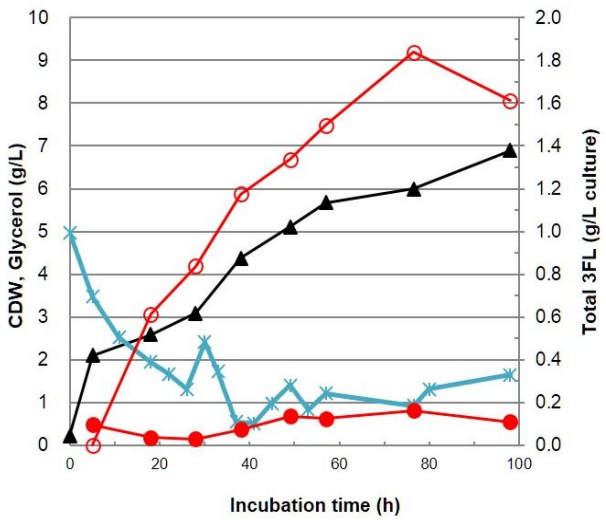


Figure 5.14 3-FL production in BL21 ΔlacZ ΔfucIK strain

3-FL was synthesized in BL21 ΔlacZ ΔfucIK strain harboring FKP and FutA (A) and FKP, Gsk, and FutA (B). Open circle, extracellular 3-FL; closed circle, intracellular 3-FL; closed triangle, dry cell weight; x, glycerol concentration

Chapter 6.

Conclusion and further suggestion

6.1 Conclusion

For the efficient production of sialyl- and fucosyllactose from human milk oligosaccharides, enzymatic biosynthesis were designed and protein engineering of glycosyltransferases using hybrid approach was developed to improve the reaction process and increase the yield and productivity for the desired production.

In this study, sialyltransferase and fucosyltransferase involved in GT-B fold structure possessing inverting mechanism were engineered for the production of silayl- and fucosyllactose. GT-B fold enzymes have distinct substrate binding domains for donor and acceptor substrates, and two domains are facing each other with the active site in the center. Therefore, modeling and docking substrates into the predicted substrate binding sites of the enzyme based on the key residues in the active site would become important clues for the mutagenesis to achieve better performance.

α2,3 and α2,6 sialyltransferases were successfully engineered using a hybrid approach that combines rational design with site-saturation mutagenesis. Narrowly focused on the substrate-binding pocket of the STs, putative functional residues were selected by multiple sequence alignment and alanine scanning, and subsequently subjected to site-saturation mutagenesis. Saturation mutagenesis of the functional residues could generate the mutants with enhanced activity and reduced side activity. These mutants of sialyltransferase were used efficiently to improve the reaction process for mass production through the construction of one-pot batch reaction and enhancement of the yield for the production of 3′-SL and 6′-SL.

For the production of fucosyllactose (2′-FL and 3-FL), α1,2 and α1,3-fucosyltransferase were used in protein engineering strategies to increase protein solubility and catalytic activity for lactose acceptor. Focused directed evolution was conducted for α1,3-fucosyltransferase by combining the structure-guided analysis of active site/substrate binding site and clustered iterative saturation mutagenesis. As a result, a quadruple mutant generated in this study was applied for the production of 3-FL, resulting in increase in yield and productivity by 40- and 560-fold based on 5 mM GDP-Fuc, compared with original enzyme.

The mutant was applied to *in vivo* production of 3-FL in engineered *E. coli*. Media optimization increased the solubility of α1,3- fucosyltransferase, and GDP-Fuc production has been determined to be a rate limiting step for the production of 3-FL in *E. coli*. In order to increase the GDP-Fuc production, salvage pathway enzyme FKP was overexpressed with Gsk, which is capable of conversion of guanosine to GMP. *E. coli* BL21 (DE3) strain knocked-out for genes corresponding to the degradation of substrates was used for fed-batch culture, resulted in production of 2 g/L of 3-FL from glycerol, L-fucose, and lactose.

In summary, this work will become a good example for the development of mutants of GT-B fold enzymes through hybrid approach for engineered substrate specificities and improved catalytic activity for mass production of useful oligosaccharides with increased yield and productivity.

6.2 Further suggestion

Based on the quantitative analysis of nucleoside phosphate in cultured cell for 3-FL production, we assumed that bottleneck of GDP-Fuc production is caused by lack of GMP, GDP, and GTP. To increase the GTP pool in *E. coli*, we identified the chip data (GSE33147) of genes related with GTP synthesis of *E. coli* in M9 minimal medium supplemented with 2 g/L of glycerol for 44 days. According to the chip data, three genes (*gsk*, *purF*, and *purL*) were found to be expressed at relatively low level (Figure 6.1). Using the chip data, metabolic modeling was used to search target genes for increasing the GTP, and then the same three genes were identified as target genes for the production of GTP. Therefore, metabolic engineering could be proceeded to increase the expression of the three genes.

In addition, modular optimization could be performed by casting metabolic network as a collection of distinct modules (Xu et al. 2013). For example, production of GDP-Fuc, GTP, FL, and transport of substrates/product could be arranged into individual modules. Tuning transcriptional efficiency by altering promoter strength or translational efficiency by changing plasmid copy number and ribosome binding sites (RBS) could be conducted combinatorially for the modules to optimize a multi-gene pathway for FL production in *E. coli*. To construct modules, *de novo* pathway would be considered for GDP-Fuc production as well as salvage pathway. Although the salvage pathway enzyme has advantages for use of single enzyme and higher efficiency for the production of GDP-Fuc than *de novo* pathway (Figure 5.8), *de novo* pathway would be also considered for the economical mass production. Salvage pathway enzyme (FKP)

uses L-fucose as a substrate, whereas glucose (or glycerol) is used to produce GDP-Fuc in *de novo* pathway. For further study, *de novo* and salvage pathway could be combined to optimize FL production yield from the carbon source and substrates supplemented from extracellular media.

Finally, 3-FL production in *E. coli* system should be applied to the liter scale culture using fermentor. Compared with the FL production for the BL21 (DE3) *E. coli* strain constructed in this study and those for other studies, titer (produced FL g per dry cell weight g) for the BL21 (DE3) *E. coli* strain is higher than those of other strains. The comparison indicates that increase of cell mass and optimization of culture condition in fermentor are required to increase the production of 3-FL in *E. coli*.

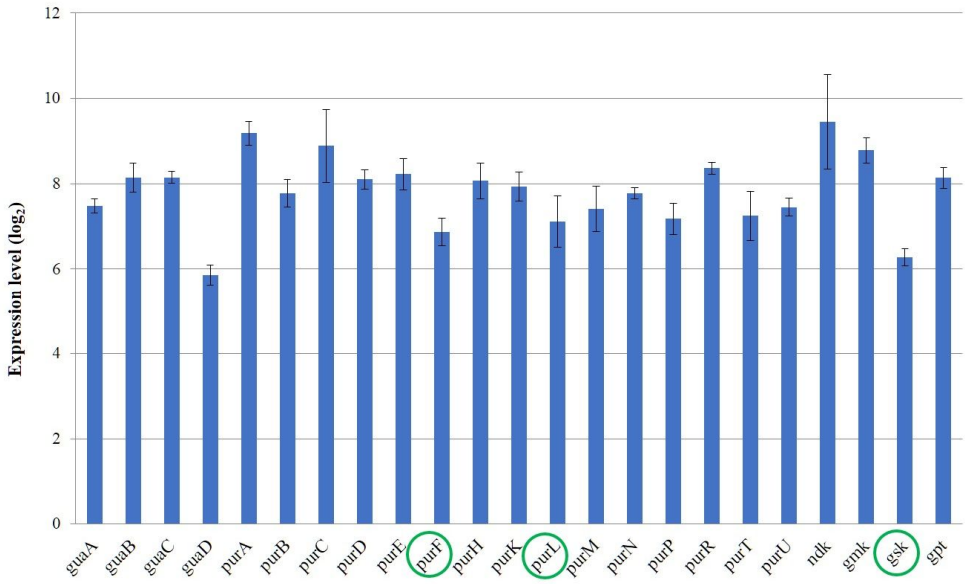


Figure 6.1 Expression profiling by array from GSE33147

gsk, guanosine inosine kinase; purF, amidophosphoribosyltransferase; purL, phosphoribosylformylglycinamidine synthase

Table 6.1 *in vivo* production of FL

Strain		Modifications	Titer (g/g CDW)	Titer (g/L)	Media/ Feeding	Culture type	Refs
JM109 (DE3)	De novo (plasmid)	Col1F', rcsA ΔlacZ	?	11 2′-FL	MM/ Glc, Lac	Fed-batch (1 L fermentor)	Angew. Chem. Int. Ed. 45, 1778–1780 (2006)
JM109 (DE3)	De novo + Salvage (chromosome)	lacZΔ15 ΔfuclK	0.31	20.2 2′-FL	MM/ Glycerol, Lac, Fuc	Fed-batch (13.5 L fermentor)	Microbial cell Fac. 12, 1–13 (2013)
BL21 Star (DE3)	De novo (plasmid)	lacZΔ15	0.09	6.4 2′-FL	MM/ Glycerol, Lac	Fed-batch (1 L fermentor)	J. Biotechnol. 210 107–115 (2015)
JM109 (DE3)	Salvage (plasmid)	FucP SetA	?	0.05 3-FL	MM/ Glc, Glyerol, Lac, Fuc	Batch (100 mL)	US patent US8652808 B2 (2014)
BL21(DE3)	Salvage (plasmid)	ΔlacZ ΔfuclK	0.33	2.0 3-FL	MM+YE/ Glycerol, Lac, Fuc	Fed-batch (50 mL flask)	In this study
BW25113 (DE3)	Salvage (plasmid)	ΔfuclK	0.26	1.06 3-FL	MM+YE/ Glycerol, Lac, Fuc	Fed-batch (50 mL flask)	In this study

References

Acevedo-Rocha CG, Reetz MT, Nov Y. 2015. Economical analysis of saturation mutagenesis experiments. Scientific Reports 5.

Aharoni A, Thieme K, Chiu CP, Buchini S, Lairson LL, Chen H, Strynadka NC, Wakarchuk WW, Withers SG. 2006. High-throughput screening methodology for the directed evolution of glycosyltransferases. Nat Methods 3(8):609-14.

Albesa-Jove D, Giganti D, Jackson M, Alzari PM, Guerin ME. 2014. Structure-function relationships of membrane-associated GT-B glycosyltransferases. Glycobiology 24(2):108-124.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389-402.

Antoine T, Priem B, Heyraud A, Greffe L, Gilbert M, Wakarchuk WW, Lam JS, Samain E. 2003. Large-scale in vivo synthesis of the carbohydrate moieties of gangliosides GM1 and GM2 by metabolically engineered Escherichia coli. Chembiochem 4(5):406-12.

Asakuma S, Hatakeyama E, Urashima T, Yoshida E, Katayama T, Yamamoto K, Kumagai H, Ashida H, Hirose J, Kitaoka M. 2011. Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. J Biol Chem 286(40):34583-92.

Ashida H, Miyake A, Kiyohara M, Wada J, Yoshida E, Kumagai H, Katayama T, Yamamoto K. 2009. Two distinct alpha-L-fucosidases from Bifidobacterium bifidum are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. Glycobiology 19(9):1010-7.

Banas P, Otyepka M, Jerabek P, Petrek M, Damborsky J. 2006. Mechanism of enhanced conversion of 1,2,3-trichloropropane by mutant haloalkane dehalogenase revealed by molecular modeling. J Comput Aided Mol Des 20(6):375-83.

Baumgartner F, Seitz L, Sprenger GA, Albermann C. 2013. Construction of Escherichia coli strains with chromosomally integrated expression

cassettes for the synthesis of 2'-fucosyllactose. Microb Cell Fact 12:40.

Biedermannova L, Prokop Z, Gora A, Chovancova E, Kovacs M, Damborsky J, Wade RC. 2012. A Single Mutation in a Tunnel to the Active Site Changes the Mechanism and Kinetics of Product Release in Haloalkane Dehalogenase LinB. Journal of Biological Chemistry 287(34):29062-29074.

Blum JK, Ricketts MD, Bommarius AS. 2012. Improved thermostability of AEH by combining B-FIT analysis and structure-guided consensus method. Journal of Biotechnology 160(3-4):214-221.

Bode L. 2012. Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology 22(9):1147-62.

Bode L, Jantscher-Krenn E. 2012. Structure-Function Relationships of Human Milk Oligosaccharides. Advances in Nutrition 3(3):383s-391s.

Bolam DN, Roberts S, Proctor MR, Turkenburg JP, Dodson EJ, Martinez-Fleites C, Yang M, Davis BG, Davies GJ, Gilbert HJ. 2007. The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. Proceedings of the National Academy of Sciences of the United States of America 104(13):5336-5341.

Bornscheuer UT. 2013. Protein Engineering as a Tool for the Development of Novel Bioproduction Systems. Fundamentals and Application of New Bioproduction Systems 137:25-40.

Breton C, Fournel-Gigleux S, Palcic MM. 2012. Recent structures, evolution and mechanisms of glycosyltransferases. Current Opinion in Structural Biology 22(5):540-549.

Breton C, Imberty A. 1999. Structure/function studies of glycosyltransferases. Current Opinion in Structural Biology 9(5):563-571.

Breton C, Snajdrova L, Jeanneau C, Koca J, Imberty A. 2006. Structures and mechanisms of glycosyltransferases. Glycobiology 16(2):29r-37r.

Buck R, Duska-McEwen GO, Schaller JP. 2014. Nutritional formulations including human milk oligosaccharides and antioxidants and uses thereof.

U.S. Patent No 8,703,737.

Campbell JA, Davies GJ, Bulone V, Henrissat B. 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochemical Journal 326:929-939.

Chang A, Singh S, Phillips GN, Thorson JS. 2011. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. Current Opinion in Biotechnology 22(6):800-808.

Chaturvedi P, Warren CD, Altaye M, Morrow AL, Ruiz-Palacios G, Pickering LK, Newburg DS. 2001. Fucosylated human milk oligosaccharides vary between individuals and over the course of lactation. Glycobiology 11(5):365-72.

Chen K, Arnold FH. 1993. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci U S A 90(12):5618-22.

Cheng J, Huang S, Yu H, Li Y, Lau K, Chen X. 2010. Trans-sialidase activity of Photobacterium damsela alpha2,6-sialyltransferase and its application in the synthesis of sialosides. Glycobiology 20(2):260-8.

Chessa D, Winter MG, Jakomin M, Baumler AJ. 2009. Salmonella enterica serotype Typhimurium Std fimbriae bind terminal alpha(1,2)fucose residues in the cecal mucosa. Molecular Microbiology 71(4):864-875.

Chica RA, Doucet N, Pelletier JN. 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotechnol 16(4):378-84.

Chin YW, Kim JY, Lee WH, Seo JH. 2015. Enhanced production of 2'-fucosyllactose in engineered Escherichia coli BL21star(DE3) by modulation of lactose metabolism and fucosyltransferase. J Biotechnol 210:107-15.

Choi YH, Kim JH, Park JH, Lee N, Kim DH, Jang KS, Park IH, Kim BG. 2014. Protein engineering of alpha2,3/2,6-sialyltransferase to improve the yield

and productivity of in vitro sialyllactose synthesis. Glycobiology 24(2):159-69.

Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P and others. 2012. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. Plos Computational Biology 8(10).

Coppa GV, Zampini L, Galeazzi T, Facinelli B, Ferrante L, Capretti R, Orazio G. 2006. Human milk oligosaccharides inhibit the adhesion to Caco-2 cells of diarrheal pathogens: Escherichia coli, Vibrio cholerae, and Salmonella fyris. Pediatr Res 59(3):377-82.

Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol 328(2):307-17.

Dalby PA. 2011. Strategy and success for the directed evolution of enzymes. Current Opinion in Structural Biology 21(4):473-480.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97(12):6640-5.

Davids T, Schmidt M, Bottcher D, Bornscheuer UT. 2013. Strategies for the discovery and engineering of enzymes for biocatalysis. Curr Opin Chem Biol 17(2):215-20.

Drouillard S, Driguez H, Samain E. 2006. Large-scale synthesis of H-antigen oligosaccharides by expressing Helicobacter pylori alpha1,2-fucosyltransferase in metabolically engineered Escherichia coli cells. Angew Chem Int Ed Engl 45(11):1778-80.

Drouillard S, Mine T, Kajiwara H, Yamamoto T, Samain E. Efficient synthesis of 6'-sialyllactose, 6,6'-disialyllactose, and 6'-KDO-lactose by metabolically engineered E. coli expressing a multifunctional sialyltransferase from the Photobacterium sp. JT-ISH-224. Carbohydr Res 345(10):1394-9.

Dumon C, Samain E, Priem B. 2004. Assessment of the two Helicobacter pylori alpha-1,3-fucosyltransferase ortholog genes for the large-scale synthesis

of LewisX human milk oligosaccharides by metabolically engineered Escherichia coli. Biotechnol Prog 20(2):412-9.

Endo T, Koizumi S, Tabata K, Ozaki A. 2000. Large-scale production of CMP-NeuAc and sialylated oligosaccharides through bacterial coupling. Appl Microbiol Biotechnol 53(3):257-61.

Errey JC, Lee SS, Gibson RP, Fleites CM, Barry CS, Jung PMJ, O'Sullivan AC, Davis BG, Davies GJ. 2010. Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors. Angewandte Chemie-International Edition 49(7):1234-1237.

Fierfort N, Samain E. 2008. Genetic engineering of Escherichia coli for the economical production of sialylated oligosaccharides. J Biotechnol 134(3-4):261-5.

Floor RJ, Wijma HJ, Colpa DI, Ramos-Silva A, Jekel PA, Szymanski W, Feringa BL, Marrink SJ, Janssen DB. 2014. Computational Library Design for Increasing Haloalkane Dehalogenase Stability. Chembiochem 15(11):1659-1671.

Goldsmith M, Tawfik DS. 2012. Directed enzyme evolution: beyond the low-hanging fruit. Curr Opin Struct Biol 22(4):406-12.

Guo Q, Zhao F, Guo SY, Wang XC. 2004. The tryptophane residues of dimeric arginine kinase: roles of Trp-208 and Trp-218 in active site and conformation stability. Biochimie 86(6):379-386.

Gupta RD, Tawfik DS. 2008. Directed enzyme evolution via small and effective neutral drift libraries. Nature Methods 5(11):939-942.

Han NS, Kim TJ, Park YC, Kim J, Seo JH. 2012. Biotechnological production of human milk oligosaccharides. Biotechnology Advances 30(6):1268-1278.

Hanahan D. 1983. Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166(4):557-80.

Heinzelman P, Komor R, Kanaan A, Romero P, Yu XL, Mohler S, Snow C, Arnold F. 2010. Efficient screening of fungal cellobiohydrolase class I

enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. Protein Engineering Design & Selection 23(11):871-880.

Hoffmann G, Bonsch K, Greiner-Stoffele T, Ballschmiter M. 2011. Changing the substrate specificity of P450cam towards diphenylmethane by semi-rational enzyme engineering. Protein Engineering Design & Selection 24(5):439-446.

Hufner E, Parkot J, Jennewein S; 2010. Synthesis of fucosylated compounds.

Idota T, Kawakami H, Murakami Y, Sugawara M. 1995. Inhibition of cholera toxin by human milk fractions and sialyllactose. Biosci Biotechnol Biochem 59(3):417-9.

Ishikawa M, Koizumi S. 2010. Microbial production of N-acetylneuraminic acid by genetically engineered Escherichia coli. Carbohydr Res 345(18):2605-9.

J.S. Woo JKS, B. G Kim, S.Y Kang, D.H. Kim, K.S. Jang, J.Y. Yang, W.M. Seo, T.G. Kil, S.H. Sim, I. K. Heo. 2008. Novel N-Acetylglucosamine-2-Epimerase and Method for Producing CMP-neuraminic acid Using the Same. Korea patent.

Kakuta Y, Okino N, Kajiwara H, Ichikawa M, Takakura Y, Ito M, Yamamoto T. 2008. Crystal structure of Vibrionaceae Photobacterium sp. JT-ISH-224 alpha2,6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: catalytic mechanism and substrate recognition. Glycobiology 18(1):66-73.

Katayama T, Sakuma A, Kimura T, Makimura Y, Hiratake J, Sakata K, Yamanoi T, Kumagai H, Yamamoto K. 2004. Molecular cloning and characterization of Bifidobacterium bifidum 1,2-alpha-L-fucosidase (AfcA), a novel inverting glycosidase (Glycoside hydrolase family 95). Journal of Bacteriology 186(15):4885-4893.

Kille S, Acevedo-Rocha CG, Parra LP, Zhang ZG, Opperman DJ, Reetz MT, Acevedo JP. 2013. Reducing Codon Redundancy and Screening Effort of

Combinatorial Protein Libraries Created by Saturation Mutagenesis. Acs Synthetic Biology 2(2):83-92.

Kim DU, Yoo JH, Lee YJ, Kim KS, Cho HS. 2008. Structural analysis of sialyltransferase PM0188 from Pasteurella multocida complexed with donor analogue and acceptor sugar. BMB Rep 41(1):48-54.

Kim JH. 2013. Enzymatic biosynthesis of 3-fucosyllactose and improvement of protein solubility by random mutagenesis. Seoul national university.

Kobata A. 2010. Structures and application of oligosaccharides in human milk. Proceedings of the Japan Academy Series B-Physical and Biological Sciences 86(7):731-747.

Kong XD, Yuan SG, Li L, Chen S, Xu JH, Zhou JH. 2014. Engineering of an epoxide hydrolase for efficient bioresolution of bulky pharmaco substrates. Proceedings of the National Academy of Sciences of the United States of America 111(44):15717-15722.

Krauth C, Fedoryshyn M, Schleberger C, Luzhetskyy A, Bechthold A. 2009. Engineering a function into a glycosyltransferase. Chem Biol 16(1):28-35.

Kunz C, Rudloff S, Baier W, Klein N, Strobel S. 2000. Oligosaccharides in human milk: structural, functional, and metabolic aspects. Annu Rev Nutr 20:699-722.

Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: Structures, functions, and mechanisms. Annual Review of Biochemistry 77:521-555.

Lazarus MB, Nam YS, Jiang JY, Sliz P, Walker S. 2011. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. Nature 469(7331):564-U168.

Lee SG, Lee JO, Yi JK, Kim BG. 2002. Production of cytidine 5'-monophosphate N-acetylneuraminic acid using recombinant Escherichia coli as a biocatalyst. Biotechnol Bioeng 80(5):516-24.

Lee WH, Han NS, Park YC, Seo JH. 2009. Modulation of guanosine 5 '-

diphosphate-D-mannose metabolism in recombinant Escherichia coli for production of guanosine 5 '-diphosphate-L-fucose. Bioresource Technology 100(24):6143-6148.

Lee WH, Pathanibul P, Quarterman J, Jo JH, Han NS, Miller MJ, Jin YS, Seo JH. 2012a. Whole cell biosynthesis of a functional oligosaccharide, 2'-fucosyllactose, using engineered Escherichia coli. Microb Cell Fact 11:48.

Lee WH, Shin SY, Kim MD, Han NS, Seo JH. 2012b. Modulation of guanosine nucleotides biosynthetic pathways enhanced GDP-L-fucose production in recombinant Escherichia coli. Appl Microbiol Biotechnol 93(6):2327-34.

Lehmann F, Tiralongo E, Tiralongo J. 2006. Sialic acid-specific lectins: occurrence, specificity and function. Cell Mol Life Sci 63(12):1331-54.

Lin SW, Yuan TM, Li JR, Lin CH. 2006. Carboxyl terminus of Helicobacter pylori alpha1,3-fucosyltransferase determines the structure and stability. Biochemistry 45(26):8108-16.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42(Database issue):D490-5.

Lowe JB. 2003. Glycan-dependent leukocyte adhesion and recruitment in inflammation. Curr Opin Cell Biol 15(5):531-8.

Lutz S. Beyond directed evolution--semi-rational protein engineering and design. Curr Opin Biotechnol 21(6):734-43.

Lutz S. 2010. Beyond directed evolution--semi-rational protein engineering and design. Curr Opin Biotechnol 21(6):734-43.

Ma B, Audette GF, Lin S, Palcic MM, Hazes B, Taylor DE. 2006. Purification, kinetic characterization, and mapping of the minimal catalytic domain and the key polar groups of Helicobacter pylori alpha-(1,3/1,4)-fucosyltransferases. J Biol Chem 281(10):6385-94.

Magalhaes A, Reis CA. 2010. Helicobacter pylori adhesion to gastric epithelial

cells is mediated by glycan receptors. Brazilian Journal of Medical and Biological Research 43(7):611-618.

Mattila P, Rabina J, Hortling S, Helin J, Renkonen R. 2000. Functional expression of Escherichia coli enzymes synthesizing GDP-L-fucose from inherent GDP-D-mannose in Saccharomyces cerevisiae. Glycobiology 10(10):1041-1047.

Moretti R, Chang A, Peltier-Pain P, Bingman CA, Phillips GN, Thorson JS. 2011. Expanding the Nucleotide and Sugar 1-Phosphate Promiscuity of Nucleotidyltransferase RmlA via Directed Evolution. Journal of Biological Chemistry 286(15):13235-13243.

Morrow AL, Ruiz-Palacios GM, Altaye M, Jiang X, Guerrero ML, Meinzen-Derr JK, Farkas T, Chaturvedi P, Pickering LK, Newburg DS. 2004. Human milk oligosaccharides are associated with protection against diarrhea in breast-fed infants. Journal of Pediatrics 145(3):297-303.

Newburg DS, Grave G. 2014. Recent advances in human milk glycobiology. Pediatr Res 75(5):675-9.

Newburg DS, Pickering LK, Mccluer RH, Cleary TG. 1990. Fucosylated Oligosaccharides of Human-Milk Protect Suckling Mice from Heat-Stabile Enterotoxin of Escherichia-Coli. Journal of Infectious Diseases 162(5):1075-1080.

Newburg DS, Ruiz-Palacios GM, Morrow AL. 2005. Human milk glycans protect infants against enteric pathogens. Annual Review of Nutrition 25:37-58.

Ni L, Chokhawala HA, Cao H, Henning R, Ng L, Huang S, Yu H, Chen X, Fisher AJ. 2007. Crystal structures of Pasteurella multocida sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. Biochemistry 46(21):6288-98.

Ni LS, Sung MC, Yu H, Chokhawala H, Chen X, Fisher AJ. 2006. Cytidine 5 '-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from Pasteurella multocida. Biochemistry 45(7):2139-

2148.

Ninonuevo MR, Park Y, Yin HF, Zhang JH, Ward RE, Clowers BH, German JB, Freeman SL, Killeen K, Grimm R and others. 2006. A strategy for annotating the human milk glycome. Journal of Agricultural and Food Chemistry 54(20):7471-7480.

Park HH, Choi J, Lee HJ, Ryu J, Park JH, Rhee WJ, Park TH. 2015. Enhancement of human erythropoietin production in Chinese hamster ovary cells through supplementation of 30Kc19-30Kc6 fusion protein. Process Biochemistry 50(6):973-980.

Park I-H. 2012. Development of immobilization method for multi enzymes system for CMP-NeuAc synthesis and enzymatic synthesis of 2-fucosyllactose. Seoul national university.

Park SH, Park HY, Sohng JK, Lee HC, Liou K, Yoon YJ, Kim BG. 2009. Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. Biotechnol Bioeng 102(4):988-94.

Pavelka A, Chovancova E, Damborsky J. 2009. HotSpot Wizard: a web server for identification of hot spots in protein engineering. Nucleic Acids Res 37(Web Server issue):W376-83.

Pavlova M, Klvana M, Prokop Z, Chaloupkova R, Banas P, Otyepka M, Wade RC, Tsuda M, Nagata Y, Damborsky J. 2009. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. Nature Chemical Biology 5(10):727-733.

Persson K, Ly HD, Dieckelmann M, Wakarchuk WW, Withers SG, Strynadka NCJ. 2001. Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. Nature Structural Biology 8(2):166-175.

Phue JN, Noronha SB, Hattacharyya R, Wolfe AJ, Shiloach J. 2005. Glucose metabolism at high density growth of E. coli B and E. coli K: differences in metabolic pathways are responsible for efficient glucose utilization in

E. coli B as determined by microarrays and Northern blot analyses. Biotechnol Bioeng 90(7):805-20.

Prabhu P, Jeya M, Lee JK. 2010. Probing the molecular determinant for the catalytic efficiency of L-arabinose isomerase from Bacillus licheniformis. Appl Environ Microbiol 76(5):1653-60.

Priem B, Gilbert M, Wakarchuk WW, Heyraud A, Samain E. 2002. A new fermentation process allows large-scale production of human milk oligosaccharides by metabolically engineered bacteria. Glycobiology 12(4):235-40.

Qasba PK, Ramakrishnan B, Boeggeman E. 2005. Substrate-induced conformational changes in glycosyltransferases. Trends Biochem Sci 30(1):53-62.

Ramakrishnan B, Qasba PK. 2010. Crystal Structure of the Catalytic Domain of Drosophila beta 1,4-Galactosyltransferase-7. Journal of Biological Chemistry 285(20):15619-15626.

Reetz MT, Carballeira JD. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat Protoc 2(4):891-903.

Reetz MT, D Carballeira J, Vogel A. 2006a. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. Angewandte Chemie-International Edition 45(46):7745-7751.

Reetz MT, Prasad S, Carballeira JD, Gumulya Y, Bocola M. 2010. Iterative Saturation Mutagenesis Accelerates Laboratory Evolution of Enzyme Stereoselectivity: Rigorous Comparison with Traditional Methods. Journal of the American Chemical Society 132(26):9144-9152.

Reetz MT, Wang LW, Bocola M. 2006b. Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. Angewandte Chemie-International Edition 45(8):1236-1241.

Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS. 2003. Campylobacter jejuni binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk

inhibit its binding and infection. J Biol Chem 278(16):14112-20.

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779-815.

Shimada K, Nagano M, Kawai M, Koga H. 2005. Influences of amino acid features of glutathione S-transferase fusion proteins on their solubility. Proteomics 5(15):3859-63.

Sinclair HR, Smejkal CW, Glister C, Kemp F, van den Heuvel E, de Slegte J, Gibson GR, Rastall RA. 2008. Sialyloligosaccharides inhibit cholera toxin binding to the GM1 receptor. Carbohydr Res 343(15):2589-94.

Song JA, Lee DS, Park JS, Han KY, Lee J. 2011. A novel Escherichia coli solubility enhancer protein for fusion expression of aggregation-prone heterologous proteins. Enzyme and Microbial Technology 49(2):124-130.

Song JA, Lee DS, Park JS, Han KY, Lee J. 2012. The N-domain of Escherichia coli phosphoglycerate kinase is a novel fusion partner to express aggregation-prone heterologous proteins. Biotechnology and Bioengineering 109(2):325-335.

Soriano del Amo D, Wang W, Besanceney C, Zheng T, He Y, Gerwe B, Seidel RD, 3rd, Wu P. 2010. Chemoenzymatic synthesis of the sialyl Lewis X glycan and its derivatives. Carbohydr Res 345(9):1107-13.

Steiner K, Schwab H. 2012. Recent advances in rational approaches for enzyme engineering. Comput Struct Biotechnol J 2:e201209010.

Stemmer WP. 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature 370(6488):389-91.

Sturgeon RJ, Deamer RL, Harbison HA. 1979. Improved spectrophotometric determination of glycerol and its comparison with an enzymatic method. J Pharm Sci 68(8):1064-6.

Sugiarto G, Lau K, Li Y, Khedri Z, Yu H, Le DT, Chen X. Decreasing the sialidase activity of multifunctional Pasteurella multocida alpha2-3-sialyltransferase 1 (PmST1) by site-directed mutagenesis. Mol Biosyst 7(11):3021-7.

Sun HY, Lin SW, Ko TP, Pan JF, Liu CL, Lin CN, Wang AH, Lin CH. 2007. Structure and mechanism of Helicobacter pylori fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. J Biol Chem 282(13):9973-82.

Sun M, Li Y, Chokhawala HA, Henning R, Chen X. 2008. N-Terminal 112 amino acid residues are not required for the sialyltransferase activity of Photobacterium damsela alpha2,6-sialyltransferase. Biotechnol Lett 30(4):671-6.

Unligil UM, Rini JM. 2000. Glycosyltransferase structure and mechanism. Curr Opin Struct Biol 10(5):510-7.

Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH. 2002. Protein building blocks preserved by recombination. Nature Structural Biology 9(7):553-558.

Wang B. 2009. Sialic acid is an essential nutrient for brain development and cognition. Annu Rev Nutr 29:177-222.

Wang G, Rasko DA, Sherburne R, Taylor DE. 1999. Molecular genetic basis for the variable expression of Lewis Y antigen in Helicobacter pylori: analysis of the alpha (1,2) fucosyltransferase gene. Mol Microbiol 31(4):1265-74.

Wang W, Hu T, Frantom PA, Zheng T, Gerwe B, Del Amo DS, Garret S, Seidel RD, 3rd, Wu P. 2009. Chemoenzymatic synthesis of GDP-L-fucose and the Lewis X glycan derivatives. Proc Natl Acad Sci U S A 106(38):16096-101.

Weichert S, Jennewein S, Hufner E, Weiss C, Borkowski J, Putze J, Schroten H. 2013. Bioengineered 2'-fucosyllactose and 3-fucosyllactose inhibit the adhesion of Pseudomonas aeruginosa and enteric pathogens to human intestinal and respiratory cell lines. Nutr Res 33(10):831-8.

Welch M, Govindarajan S, Ness JE, Villalobos A, Gurney A, Minshull J, Gustafsson C. 2009. Design Parameters to Control Synthetic Gene Expression in Escherichia coli. Plos One 4(9).

Williams GJ, Zhang C, Thorson JS. 2007. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. Nat Chem Biol 3(10):657-62.

Woo J.S. KBG, Kim D.H., Choi Y.H., Song J.K. , Kang S.Y , Seo W.M. , Yang J.Y. , Lee S.M. . 2013. Method of Preparing Sialyl Derivative. Korea patent 10-2014-0141228.

Wu S, Acevedo JP, Reetz MT. 2010a. Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. Proc Natl Acad Sci U S A 107(7):2775-80.

Wu SA, Tao NN, German JB, Grimm R, Lebrilla CB. 2010b. Development of an Annotated Library of Neutral Human Milk Oligosaccharides. Journal of Proteome Research 9(8):4138-4151.

Xu P, Gu Q, Wang W, Wong L, Bower AG, Collins CH, Koffas MA. 2013. Modular optimization of multi-gene pathways for fatty acids production in E. coli. Nat Commun 4:1409.

Yamamoto T, Ichikawa M, Takakura Y. 2008. Conserved amino acid sequences in the bacterial sialyltransferases belonging to Glycosyltransferase family 80. Biochem Biophys Res Commun 365(2):340-3.

Yang GY, Rich JR, Gilbert M, Wakarchuk WW, Feng Y, Withers SG. 2010. Fluorescence Activated Cell Sorting as a General Ultra-High-Throughput Screening Method for Directed Evolution of Glycosyltransferases. Journal of the American Chemical Society 132(30):10570-10577.

Yang JP, Ma XX, He YX, Li WF, Kang Y, Bao R, Chen Y, Zhou CZ. 2011. Crystal structure of the 30 K protein from the silkworm Bombyx mori reveals a new member of the beta-trefoil superfamily. J Struct Biol 175(1):97-103.

Yeom SJ, Kim HJ, Lee JK, Kim DE, Oh DK. 2008. An amino acid at position 142 in nitrilase from Rhodococcus rhodochrous ATCC 33278 determines the substrate specificity for aliphatic and aromatic nitriles. Biochem J 415(3):401-7.

Yu H, Chokhawala H, Karpel R, Yu H, Wu B, Zhang J, Zhang Y, Jia Q, Chen X. 2005. A multifunctional Pasteurella multocida sialyltransferase: a powerful tool for the synthesis of sialoside libraries. J Am Chem Soc 127(50):17618-9.

Yu ZT, Chen C, Kling DE, Liu B, McCoy JM, Merighi M, Heidtman M, Newburg DS. 2013a. The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. Glycobiology 23(2):169-77.

Yu ZT, Chen C, Newburg DS. 2013b. Utilization of major fucosylated and sialylated human milk oligosaccharides by isolated human gut microbes. Glycobiology 23(11):1281-92.

Yuan Z, Bailey TL, Teasdale RD. 2005. Prediction of protein B-factor profiles. Proteins-Structure Function and Bioinformatics 58(4):905-912.

Zhai Y, Han D, Pan Y, Wang S, Fang J, Wang P, Liu XW. 2015. Enhancing GDP-fucose production in recombinant Escherichia coli by metabolic pathway engineering. Enzyme Microb Technol 69:38-45.

Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol 16(3):258-61.

Zivkovic AM, German JB, Lebrilla CB, Mills DA. 2011. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. Proc Natl Acad Sci U S A 108 Suppl 1:4653-8.

국문 초록

본 논문은 모유 유래의 기능성 올리고당인 시알릴락토오스 및 푸코실 락토오스 생산에 있어서 생산 수율 및 생산성 증가를 위한 당전이 효소의 하이브리드 (hybrid) 단백질 공학적 변이 전략 및 생체 내·외의 효율적인 생합성 방법 개발에 관한 연구를 골자로 하고 있다.

첫 번째로, α2,3-시알릴락토오스 (3′-SL)와 α2,6-시알릴락토오스 (6′-SL)의 생체 외 생합성 경로에서 생산 수율 및 생산성 증가를 위해 시알산 전이효소의 단백질 공학적 변이를 수행하였다. 각각의 시알릴락토오스는 당 공여체 기질인 시티딘 1인산-*N*-아세틸뉴라민산 (CMP-Neu5Ac)을 생산하는 다효소반응과 시알산 전이효소 반응의 결합에 의하여 생성된다. α2,3-시알산 전이효소의 경우 CMP-Neu5Ac를 생산하는 pH 7.0 이하에서 6′-SL의 부산물을 생산하였으며, α2,6-시알산 전이효소의 경우 상대적으로 낮은 활성을 갖는 문제점이 있었다. 본 연구에서는 논리적 (rational) 방법과 위치 특이적 포화변이 (site saturation mutagenesis)가 결합된 하이브리드 방법을 개발함으로써 적은 수의 양질의 변이체를 탐색하고자 하였다. 단백질 결정 및 모델 구조를 이용하여 시알산 전이효소의 기질 결합 포켓으로부터 다수 서열정렬과 알라닌 스캐닝을 통해 기능적 잔기를 선별하였으며 선별된 잔기에 대해 포화변이를 수행함으로써 개량된 변이체를 획득할 수 있었다. α2,3-시알산 전이효소의 경우 R313N 단일 아미노산 변이체는 2.3배의 활성증가를 나타냈으며, 두 개의 잔기가 치환된 R313N/T265S와 R313H/T265S 또한 2 배의

이상의 활성 증가를 나타냈다. 특히 R313 잔기의 N, D, Y, H, T로의 치환 변이체들은 pH 7.0 이하에서 6′-SL을 생산하는 부반응성이 완전히 제거된 것이 확인되었다. α2,6-시알산 전이효소의 경우, 단일 아미노산 변이체인 L433S/T와 두 개의 아미노산 변이체인 I411T/L433T가 야생형 대비 각각 3배 및 5배의 활성 증가를 나타냈다. 이들 변이체 각각은 일체형 회분식 공정의 구축을 가능케 하였고 수율을 향상시킴으로써 시알릴락토오스의 반응공정을 향상시키는 데에 효율적으로 사용되었다.

두 번째로, α1,2-푸코실락토오스 (2′-FL)와 α1,3-푸코실락토오스 (3-FL)를 생산하기 위하여 α1,2- 및 α1,3-푸코실 전달효소를 *Helicobacter pylori* 균주로부터 클로닝하였으며 이들 효소의 위치 특이성에 따른 기질특이성을 확인하였다. 또한 2′-FL과 3-FL을 경제적으로 생산하기 위하여 당 공여체 기질인 GDP-fucose를 *Bacteroids fragilis*의 salvage 경로의 효소를 이용하여 L-fucose로부터 생산하고 이를 푸코실 전달효소와 함께 결합하여 원포트 (one-pot)로 생산하는 반응 시스템을 구축하였다. 본 합성 경로의 속도 결정단계가 되는 푸코실 전달반응의 효율을 증가시키기 위하여 난발현성의 푸코실 전달효소에 대한 가용성 단백질 발현양 증가와 활성 증가를 위한 엔지니어링을 수행하였다. α1,2-푸코실 전달효소의 경우 염기서열 최적화와 융합 단백질 도입을 통하여 5 mM GDP-fucose 대비 2′-FL의 생산 수율을 90%로 향상시키고 생산성을 46배 이상 증가시켰다. 또한, α1,3-푸코실 전달효소의 경우 단백질의 C-말단을 체계적으로 제거한 이후 염기서열 최적화를 통해 1 L 대장균 배양 당 150-200 mg의 단백질을 획득함으로써 세포추출물을 이용한 반응에 있어서 야생형을 이용한 반응보다 18배의 3-FL 생산 수율 증가를 나타냈다. 이후 α1,3-푸코실

전달효소의 lactose에 대한 기질특이성을 향상시키기 위해 단백질의 구조에 기인한 컴퓨터 분석과 반복적 포화변이 (iterative saturation mutagenesis, ISM)가 결합된 위치 집중적 방향 진화 (focused directed evolution)가 도입되었다. 효소의 활성 부위 또는 기질 결합부위에서의 변이를 수행할 기능적 잔기를 선택하기 위해 단백질 모델링, 기질 docking및 HotSpot Wizard가 사용되었다. 선택된 각 잔기는 α-helix의 위치에 따라 집단화 (clustering)되었으며, 각 클러스터에 ISM이 수행되었다. 결과적으로, 야생형 대비 15.2배의 활성이 증가된 4개의 아미노산이 치환된 변이체를 탐색하였다. 본 변이체는 락토오스 기질에 대한 결합력 향상 및 GDP-fucose 기질에 대해서도 k_{cat}이 향상된 결과를 나타냈다. 또한 본 변이체는 3-FL의 생산 수율을 1시간 내에 96% 이상으로 향상시켰으며 야생형 대비 생산성을 대폭 향상시켰다.

세 번째로, 대장균 내에서 대사공학 연구를 통하여 3-FL을 생산하는 연구를 수행하였다. 대장균 내에 α1,3-푸코실 전달효소 변이체 및 salvage 경로의 GDP-fucose 생산관련 유전자들을 과발현하고 기질의 분해경로를 차단하였다. 결과적으로 락토오스와 푸코오스가 기질로 사용되었을 때, 글리세롤을 탄소원으로 이용함으로써 유가배양 (fed-batch)을 통해 76 시간 안에 2 g/L의 3-FL을 생산할 수 있었다.

본 연구의 하이브리드 단백질 공학적 변이는 다양한 당전이 효소의 기질 특이성 및 활성을 향상시키기 위해 효율적으로 사용될 수 있다. 또한 본 연구를 통해 생성된 당전이 효소의 변이체 및 *in vivo*·*in vitro*의 효율적 생합성 방법은 다양한 기능성 올리고당의 생물공학적 대량생산에 있어서 수율 및 생산성을 향상시킴으로써 매우 효율적으로 사용 될 수 있을 것으로 기대한다.

주요어: 기능성 모유 올리고당, 시알릴락토오스, 푸코실락토오스, 시알산 전이효소, 푸코실 전달효소, 단백질 공학적 변이, 하이브리드 접근법

학번: 2010-30265

1