공 학 석 사 학 위 논 문

High-Throughput Flow Cytometry Screening coupled with a split-GFP for Enhancing Soluble Expression of α1,2-Fucosyltransferase in *E. coli*

대장균에서의 α1,2-퓨코당 전이효소의 용해도 향상을 위한 split-GFP 및 유세포 분석기를 결합한 고속 대량 스크리닝 시스템의 적용

2021 년 2 월

서울대학교 대학원

화학생물공학부

김 정 화

High-Throughput Flow Cytometry Screening coupled with a split-GFP for Enhancing Soluble Expression of α1,2-Fucosyltransferase in *E. coli*

By

Jeonghwa Kim

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

February, 2021

School of Chemical and Biological Engineering

Seoul National University

High-Throughput Flow Cytometry Screening Coupled with a split-GFP for Enhancing Soluble Expression of α1,2-Fucosyltransferase in *E. coli*

대장균에서의 α1,2-퓨코당 전이효소의 용해도 향상을 위한 split-GFP와 유세포 분석기를 결합한 고속 대량 스크리닝 시스템의 적용

지도교수 김병기

이 논문을 공학석사 학위논문으로 제출함

2021년 2월

서울대학교 대학원

화학생물공학부

김 정 화

김정화의 공학석사 학위논문을 인준함

2021년 2월

위 원 장 한 지 숙 (인)

부위원장 김 병 기 (인)

위 원 서 상 욱 (인)

Abstract

High-Throughput Flow Cytometry Screening coupled with a split-GFP for Enhancing Soluble Expression of α1,2-Fucosyltransferase in *E. coli*

Jeonghwa Kim

Chemical and biological engineering

The Graduate School

Seoul National University

Recently, various researches on the biological advantages of human milk oligosaccharides (HMO) have been described. HMOs can be used as prebiotics for *Bifidobacterium* strains, helping the growth of intestinal microorganisms. Furthermore, HMOs help immunity by preventing the attachment of pathogens to the intestinal mucosa. 2'-Fucosyllactose (2'-FL), one of the major human milk oligosaccharides, was produced in several microorganisms engineered to produce 2'-FL for high titer such as *Escherichia coli*. Generally, α1,2-Fucosyltransferase (α1,2-FucT) can synthesize 2'-FL from GDP-L-fucose and D-lactose. However, the low solubility of α1,2-FucT acts

as a bottleneck for the production of 2'-FL through microorganisms, which limits the industrial application of α1,2-FucT.

In order to solve this solubility issue, the following studies were conducted to improve the soluble expression of α1,2-FucT. First, we generated mutant libraries by site-saturation mutagenesis based on the α-helix rule. We plotted the helical wheel of α-helix on the surface of the enzyme. Subsequently, hydrophobic amino acids present in the hydrophilic surfaces of each α-helix were targeted and grouped. Site saturation mutagenesis was performed for each group. Second, a split-GFP system was applied for screening of mutants with enhanced soluble expression. The GFP11 fragment is linked to the C-terminal of the α1,2-FucT so that the expression of GFP11 is affected by the solubility and expression of the α1,2-FucT. Each mutant library was screened via FACS to separate soluble mutants for high-throughput screening. Cells showing the highest 10% fluorescence were separated in two rounds of enrichment. The selected cells were picked for further assay using 96-well deep plates and each fluorescence intensity/OD600 of the colonies was measured by fluorescence spectrometer.

As a result, a quadruple mutants L80C/A121D/P124A/L125V was generated, which showed 1.73-fold improvement in the 2'-FL titer relative to wild-type. The results of this study show us the possibility of industrial application of α1,2-FucT for the increased productivity of 2'-fucosyllactose in *E. coli.*

Keywords: Human milk oligosaccharides (HMO), 2'-fucosyllactose, α1,2-Fucosyltransferase, α-helix rule, split-GFP, FACS, Protein engineering

Student number : 2019-24630

Contents

List of Figures

List of Tables

1. Introduction

Human milk oligosaccharides (HMOs) are the essential oligosaccharides that amount to the third most abundant oligosaccharides in human milk. Also, it is known that more than two hundred different HMOs have been reported. [1, 2] HMOs play important roles related to health promotion and immune system development, such as prebiotics to stimulate the growth of gut microorganisms. [3] Furthermore, HMOs help to enhance immunity by preventing the attachment of pathogens and inhibiting pathogenic infections in the intestinal mucosa. [4] (Figure1.1) With these advantages, HMOs are high value-added substances that have recently been developed as food ingredients, cosmetics [5], and pharmaceutical applications. [6] 2'-Fucosyllactose(2'FL) is one of the fucosylated oligosaccharides in HMO and contained about 30% in HMOs. [7] (Figure1.2) 2'-FL is produced by separation from human breast milk, chemically synthesized, and obtained by the enzymatical synthesis *in vitro*. [8] Also, it has been reported that 2'-FL can be produced in several microorganisms for high titer and productivity. [9]

In engineered *E. coli*, α1,2-fucosyltransferase (α1,2-FucT) synthesized 2'-FL by transferring the fucose from the substrate GDP-L-fucose to another sugar backbone, lactose. [10] (Figure1.3) The method of synthesizing GDP-L-fucose in *E. coli* consists of the *de novo* pathway and salvage pathway. GDP-L-fucose is synthesized

from D-Glucose via the *de novo* pathway, whereas from L-fucose through fucose-1-phosphate in salvage pathway. [8] (Figure1.4) However, despite the various engineering studied on the *E. coli* strains, the industrial application of α1,2-FucT is limited by the low soluble expression and efficiency. [11] To overcome this solubility issue, we tried to increase the soluble expression of α1,2-FucT from *Helicobacter pylori* by site saturation mutagenesis.

To enhance the soluble expression of the protein, several studies applied engineering techniques such as optimization of growth conditions, co-expression with molecular chaperones, and fusion with peptide tags. [12, 13] Also, the low solubility of protein can be increased to identify and mutate aggregation hotspots in insoluble proteins, applying the α-helix rule and the hydropathy contradiction rule by Asano group. [14] We tested the α-helix rule to identify aggregation hotspots in α1,2-FucT and targeted two or three hydrophobic amino acids present in hydrophilic surfaces of each α-helix to applying site saturation mutagenesis using degenerated codons.

For screening the enhancement of soluble expression of the enzyme, we applied the split-GFP system that was utilized to monitor protein interactions, protein folding, and assay protein solubility. [15, 16] A split-GFP consists of two GFP fragments, containing GFP1-10 detector fragment and 15-amino-acid GFP11 fragment. [16] Two GFP fragments can combine spontaneously to form a complete GFP and fluorescence. [17] The GFP11 fragment is very small, soluble, and

hardly affects protein solubility and folding, so it can be used for screening of protein solubility, protein interactions, and protein folding *in vivo* and *in vitro*. [16] We applied this system for high-throughput screening of soluble mutants. C-terminal of the α1,2-FucT was linked to the GFP11 fragment so that whole-cell fluorescence is affected by the soluble expression of the α1,2-FucT.

The mutant library was screened using fluorescence-activated cell sorter (FACS) and fluorescence spectrometer for the second screening. FACS is a specialized instrument for the powerful and quickly screening of large libraries of mutants up to 10^{10}. [18] After sorting the cells with the top 10% fluorescence twice, whole-cell fluorescence/OD600 of the colonies measured by fluorescence spectrometer for further screening. (Figure1.5)

As a result, we generated a quadruple mutant which showed 1.73-fold improvement in 2'-FL titer than that of wild-type. The results of this study show the possibility of industrial application of α1,2-FucT for the production of 2'-FL in *E. coli.*

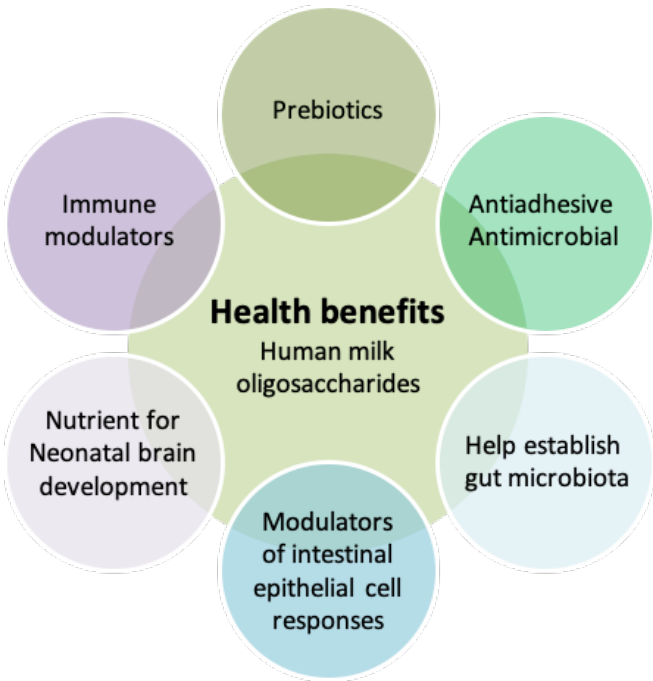


Figure1.1 Various effects promoted by HMOs

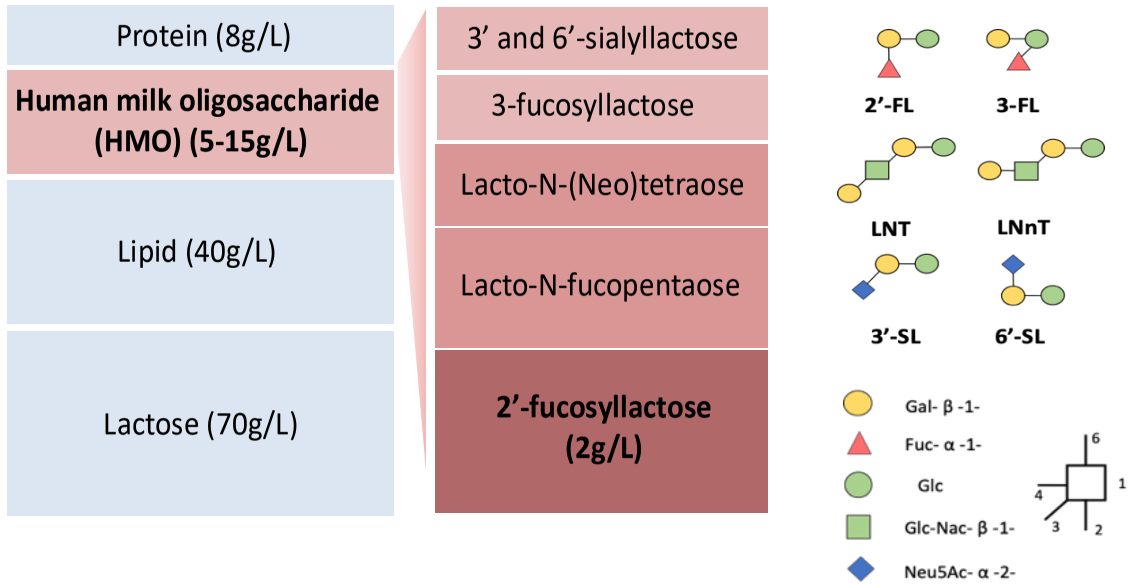


Figure1.2 Composition of human-milks and structures of common HMOs

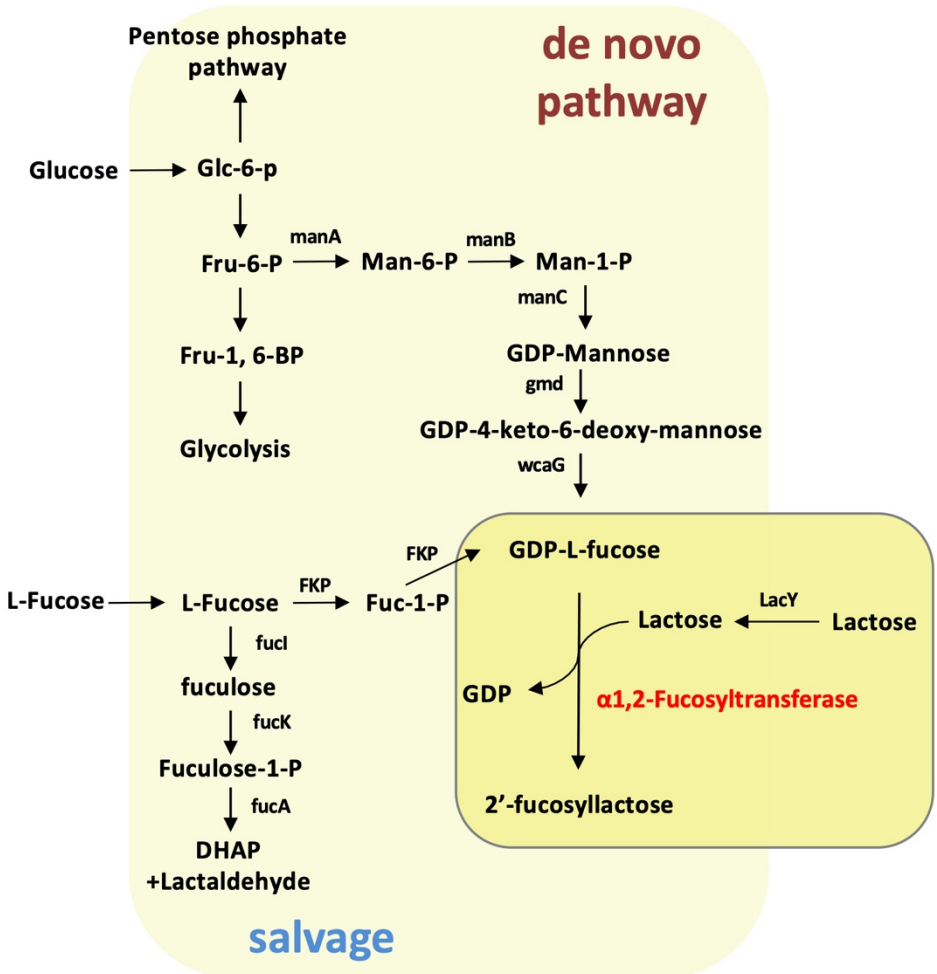


Figure1.3 Overall reaction of 2'-FL synthesis in *E. coli*

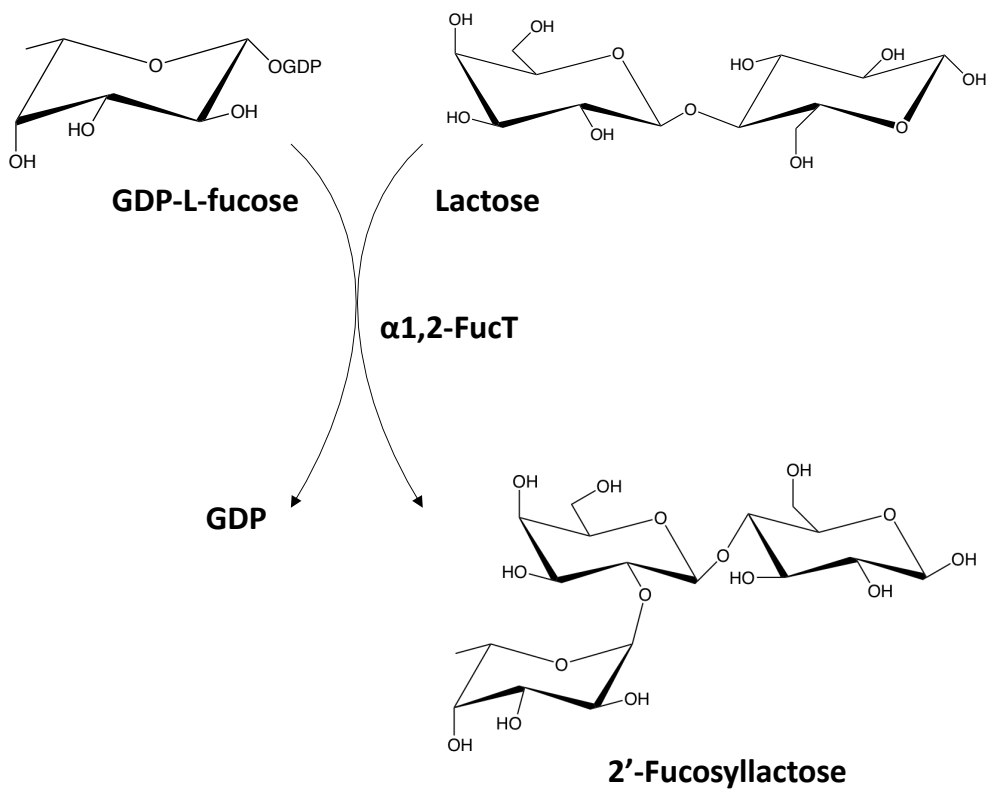
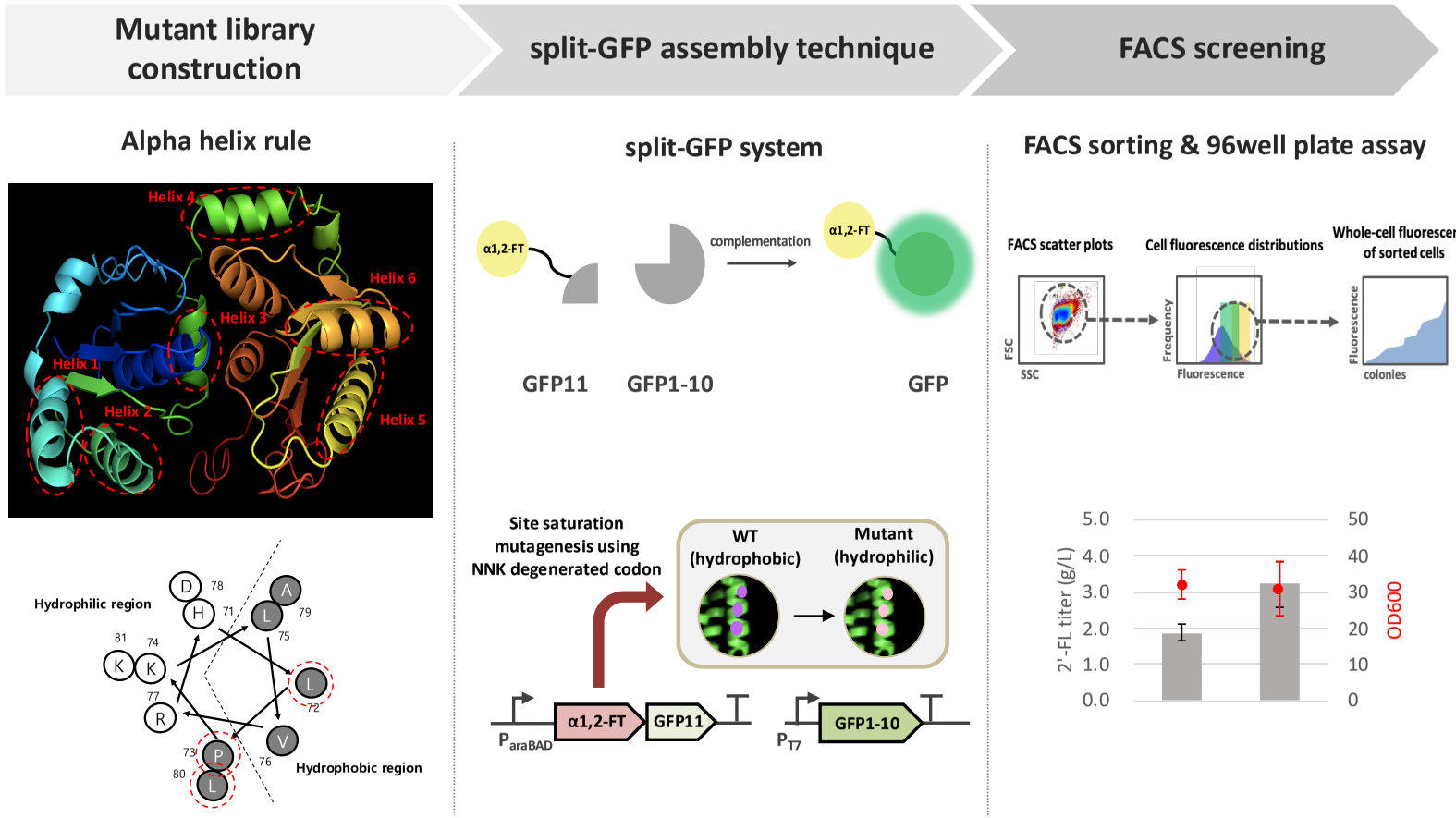


Figure1.4 Synthesis of 2'-fucosyllactose by α1,2-FucT with GDP-L-fucose and lactose sugar backbone



1.5 Overall scheme of high-throughput screening platform

2.MATERIALS AND METHODS

2.1. Chemicals and materials

E. coli BL21(DE3) and *E. coli* :: PtacΔZ (modified by Bumseok Park, Park et al, 2020.) were used as host strains for expression of cFutC, cFutC-GFP11, F-ePGK-cFutC-GFP11, and GFP1-10 gene. *E. coli* DH5α was used for several gene cloning. Luria Bertani (LB) broth was used with kanamycin or ampicillin for the growth and induced with L-arabinose or isopropyl β-D-1-thiogalactopyranoside (IPTG). The minimal media with yeast extract (1.2%), D-(+)-glucose (1.92%), and Lactose monohydrate (18mM) was used for the batch culture in *E. coli* :: PtacΔZ.

As previously reported by Choi et al, minimal media containing monopotassium phosphate (3 g/L), dipotassium phosphate (12 g/L), ammonium sulfate (5 g/L), sodium chloride (0.1 g/L), magnesium sulfate heptahydrate (0.3 g/L), calcium chloride dihydrate (0.015 g/L), ferrous sulfate heptahydrate (0.11 g/L) in sodium citrate (1.5 g/L), thiamine (7.5 μg/L) and trace element was used for the production of 2'-FL. The trace element was prepared with ethylenediaminetetraacetic acid (5g/L), ferric chloride hexahydrate (0.83 g/L), zinc chloride (84 mg/L), copper(II) chloride dihydrate (13 mg/L), cobalt(II) chlorate dihydrate (10 mg/L), boric acid (10 mg/L), manganese(II) chloride tetrahydrate (1.6 mg/L) with pH 7.5. [19] Other chemicals were obtained from Merck (Merck, St. Louis, Missouri, USA) and Duksan (Duksan Pharmaceutical, Seoul, South Korea).

All restriction enzymes, DNA polymerase, DNA ligase, and PCR reagents were purchased from Thermo Fisher Scientific (Thermo Fisher Scientific Korea, Seoul, South Korea) and NEB (New England Biolabs, Ipswich, MA, USA). Other enzymes and buffers were purchased from Takara (Takara Bio, Shiga, Japan) and Promega (Promega Corporation, Madison, WI, USA).

2.2. Construction of expression vector for split-GFP system

The pET24ma vector was used for expression of cFutC (codon-optimized FutC, modified by Yunhee Choi) for batch-culture. [20] For cloning and expression of cFutC-GFP11 and F-PGK-cFutC-GFP11 genes, both pET15b and pBAD/HisA vectors were used in this study. Modified pET28a (changed the origin to P15A by our lab) was used for expression of the GFP1-10 gene. All host strains and plasmids in this study are listed in Tables 2.1. Also, all primers in this study were synthesized by Bionics and listed in Table 2.2-2.3.

The cFutC and F-ePGK-cFutC genes were fused to GFP11 tag with linker in C-terminal of the genes by amplifying with primers in Table 2.3. First, cFutC and F-ePGK-cFutC inserts were prepared by PCR using primers that introduce the linker sequence and the GFP11 gene. Also, primers were synthesized to contain NcoI and XhoI restriction enzyme sites for cloning into the IPTG inducible pET15b vector. cFutC-GFP11 and F-PGK- cFutC-GFP11 genes were inserted into pET15b using the restriction enzymes NcoI and XhoI. By heat-shock, each reaction mixture was transformed into *E. coli* DH5α cells for

ligation. The recombinant strains were sequenced for expression of each of GFP11 fused genes to verify the feasibility of split-GFP for screening enhanced soluble expression of enzyme.

2.3. Verify the feasibility of split-GFP for screening enhanced soluble protein of α1,2-FucT

By heat-shock, each of cFutC-GFP11 and F-PGK-cFutC-GFP11 genes was transformed into *E. coli* BL21(DE3) cells with pET28a GFP1-10 plasmid for protein co-expression. The transformed cells were spread on LB-agar plate containing kanamycin (50ug/ml) and ampicillin (100ug/ml). A single colony was picked into 3mL LB media with kanamycin and ampicillin and incubated at 37 °C for 16 h. Cells were transferred into 50 mL of LB media containing the same antibiotics and incubated at 37 °C. After that, cells were induced with IPTG (0.1mM) until OD600 reached at 0.6.

After induction at 18°C for 16 h, cells were centrifuged and washed using sodium phosphate buffer (20mM, pH 7.6). Cells were lysed via sonication and the soluble protein was collected from each supernatant using centrifuge at 22,250g at 4 °C for 30 min. Next, green fluorescence intensity of soluble fraction was quantified using a fluorescence spectrophotometer (Spark multimode microplate reader, Tecan, Männedorf, Switzerland)

Strains	Description
E. coli ::Ptac ΔZ	*E. coli* BL21 (DE3) ΔlacZ ΔfucIK ::Ptac ***Park et al, 2020 (modified by Bumseok Park)**
Plasmids	**Description**
pCDFm	T7 promoter, CloDF13 ori, Sm^r, modified for *de novo* pathway ***Park et al, 2020 (modified by Bumseok Park)**
pET28a-P15A ori	T7 promoter, P15A ori, Km^r **(changed the origin in our lab)**
pET15b	T7 promoter, pBR322 ori, Amp^r
pBAD/HisA	araBAD promoter, pBR322 ori, Amp^r

Table 2.1 Host strains and plasmids used in this study

Primers	5' → 3'	Restriction sites
2'FT-NcoI-F	ATATCCATGGCCTTTAAGGTGGTGCA	NcoI
F-ePGK-2'FT-NcoI-F	ATATCCATGGGCTCTGTAATTAAGATGACCGATC	NcoI
2'FT-linker-GFP11-XhoI-R	ATATCTCGAGTGTGATTCCAGCAGCGTTCACGTACTCGTGCAGCACCATGTGGTCTCT*AGAGCCGCCGCCAGAGCCGCCGCC**ATCTAAAGCGTTATACTTCTG *: GFP11 **: linker sequence	XhoI
Primers	5' → 3'	
2'FT-GFP11-Kpn1-F	GGTACCGGCGGCGGCTCTGGCGGCGG	
2'FT-GFP11-Kpn1-R	ATCTAAAGCGTTATACTTCTGGGAT	

Table 2.2 Primers used in the construction of expression vector for screening system

Primers	5' → 3'
L72P73L80NDT-F	AAACTAGTACGCGATGCGNDTAAGTGTATGGGATTCGACCG
L72P73L80AHN-R	ACACTTAHNCGCATCGCGTACTAGTTTAHNAHNGTGTTGCATTTTCGCTATGGC
F94P98L101-NNK-F	GATTGTTNNKGAGTACGAANNKAAGCTGNNKAAACCATCGCGCCTGACATAT
F94P98L101- MNN -R	TTTMNNCAGCTTMNNTTCGTACTCMNNAACAATCTCCTGACTAACACGGTCG
A121P124L125NNK-F	NNKATATCANNKNNKATTAAGCAAACCTTTACGCTGC
A121P124L125MNN-R	GGTTTGCTTAATMNNMNNTGATATMNNGTCAAAGTATCGTGGATCCTGG
P134P134NNK-F	CGCTGCCGNNKNNKCCTGAAAATAATAAAAATAA
P134P134MNN-R	TATTATTTTCAGGMNNMNNCGGCAGCGTAAAGGTTTGC
L179A192-NNK-F	TGTCAGNNKGGTATTGACTATCAAAAAAAGGCGCTTGAGTATATGNNKAAACGCGTGCCGAACAT
L179A192-MNN -R	TTTMNNCATATACTCAAGCGCCTTTTTTTGATAGTCAATACCMNNCTGACAACCGATCCCCACATAATCGC
L207L215-NNK-F	GACNNKGAATTCACGCAGAATCTCGATNNKGGCTACCCTTTTATGGACATGAC
L207L215-MNN-R	GCCMNNATCGAGATTCTGCGTGAATTCMNNGTCTTCGCAAAAAACAAACAGTTC

Table 2.3 Primers used for the construction of each mutant library

2.4. Construction of mutant library using Site-Saturation Mutagenesis

The cFutC-GFP11 gene was inserted into pBAD/HisA vector for titratable regulation of protein expression. First, cFutC-GFP11 insert was prepared by PCR and primers that contain NcoI and XhoI restriction sites were used for construction into the pBAD/HisA vector. The cFutC-GFP11 gene was digested and inserted into pBAD/HisA vector using the restriction enzymes NcoI and XhoI. By heat-shock, each reaction mixture was transformed into *E. coli* DH5α cells to construct mutant library.

After that, helical wheel diagrams of 6 α-helix in cFutC were drawn and hydrophobic amino acids in hydrophilic regions were targeted and grouped by applying the α-helix rule. The GFP11 fused cFutC mutant libraries were constructed by site saturation mutagenesis using inverse PCR and primers containing degenerated codons in table 2.3.

By electroporation, each GFP11 fused cFutC mutant library was transformed into *E. coli* :: PtacΔZ with GFP1-10 gene for co-expression of split-GFP and spread on LB-agar plate with kanamycin (50ug/ml) and ampicillin (100ug/ml). Each mutant library contains more than 10^5 colonies and was prepared for FACS screening after induction by IPTG (0.1mM).

2.5. High-Throughput screening using flow cytometry

Each library was collected into 10mL LB media with antibiotics and cultured at 37 °C for 12 h. Precultured cells were transferred into 50

mL of LB media with antibiotics and incubated at 37 °C. After that, cells were induced with IPTG (0.1mM) until OD600 reached at 0.6. After induction at 30°C for 16 h, cells were centrifuged and washed using phosphate-buffered saline (1X PBS, pH7.4) to prepare samples.

Each sample was screened by flow cytometry, FACS (S3e cell sorter, Bio-Rad, Hercules, CA, USA). Cells showing the 10% highest fluorescence intensity were separated. By the same method, 10% of enriched cells were separated and recovered on LB-agar plate containing kanamycin (50ug/ml) and ampicillin (100ug/ml). Single colonies picked for further screening in a 96-well deep plate. In each plate, four wild type controls were included. After induction by arabinose and IPTG, whole-cell fluorescence intensity and OD600 were measured using fluorescence spectrometry. Mutants with more than 1.8-fold in fluorescence intensity/OD600 than average of four wild types were selected for sequencing.

After sequencing, each selected mutant gene was inserted into pET24ma vector to verify soluble protein fraction using primers that introduce mutations into cFutC gene. After that, SDS-PAGE analysis was performed to examine the increase in protein soluble expression of the mutants. Each of cFutC mutant was transformed into *E. coli* BL21(DE3) by heat-shock transformation for expression cFutC mutants. The transformed cells were cultured on LB-agar plate with kanamycin (50ug/ml). Each single colony was picked into 3mL LB media containing kanamycin (50ug/ml) and incubated at 37 °C for 16

h. Cells were inoculated into 50 mL of fresh LB with kanamycin (50ug/ml) and induced with IPTG (0.1mM) until OD600 reached at 0.6. After induction at 18°C for 16 h, cells were removed using centrifugation and washed using sodium phosphate buffer (20mM, pH 7.6). Cells were lysed by sonification and obtained to prepare SDS-PAGE samples from each supernatant by centrifugation at 22,250g at 4 °C for 30 min.

2.6. *In vivo* batch-culture

For the production of 2'-FL through the *de novo* pathway, pCDFm gene for the synthesis of GFP-L-fucose and each cFutC mutant gene inserted into pET24ma vector were co-transformed into *E. coli* :: PtacΔZ using electroporation.

Each transformant was picked into 3mL LB broth with kanamycin (50ug/ml) and streptomycin (100ug/ml) and cultured at 37 °C with 200rpm for 12 h. Precultured cells were transferred into 50 mL of minimal media with yeast extract, D(+)-glucose containing kanamycin and streptomycin and induced with IPTG and lactose monohydrate until OD600 reached at 0.8. After induction at 30°C with 200rpm for 16 h, cells were centrifuged and each supernatant was obtained to prepare Bio-LC sample.

2.7. Bio-LC analysis

To analyze of 2'-FL titer, each supernatant was obtained from culture medium by centrifugation at 22,250g at 4°C for 5 min. Each supernatant was heated at 95°C for 3 min to stop the remained reaction. After that, media was centrifuged at 22,250g at 4°C for 10 min to obtain 2'-FL synthesized in *E. coli*. Samples were analyzed as previously reported by Choi et al. [20, 21] Each supernatant was diluted with HPLC grade water (Duksan) and analyzed through a Bio-LC DX-300 (Dionex Co., Sunnyvale, CA, USA). Also, a CarboPac PA-1 column (Dionex Co., Sunnyvale, CA, USA), was used to separate 2'-FL and lactose [20, 21]

3. RESULTS AND DISCUSSION

3.1. Construction of the split-GFP screening system

First, we tested whether the split-GFP system could be utilized for screening increased soluble protein of α1,2-FucT. Therefore, we verified the feasibility of split-GFP using a mutant that fused F-ePGK (full length of phospho-glycerate kinase from *E. coli*) into the N-terminal of codon-optimized FutC α1,2-FucT (cFutC) as a control. cFutC has 87% homology with its original sequence, and total protein expression was dramatically enhanced 5-fold than original FutC, while the protein solubility was not improved. [20] F-ePGK is one of the fusion proteins of 40kDa in size and is well known as water-soluble and stable in *E. coli*. Also, F-ePGK fused cFutC was known to dramatically increased the soluble expression of the protein. [20]

The GFP11 fragment was fused to the C-terminal of the cFutC and F-ePGK-cFutC. Each GFP11 fused gene was inserted into pET15b vector and the GFP1-10 gene was inserted into pET28a vector to construct expression plasmids. (Figure3.1) Each GFP11 fused gene was transformed into *E. coli* BL21(DE3) cells with GFP1-10 gene for protein co-expression. After induction by IPTG, the fluorescence intensity of F-ePGK-cFutC-GFP11 was more than 3-fold than fluorescence intensity of cFutC-GFP11 induced by 1mM IPTG after 16hr. (Figure3.2) Also, F-ePGK-cFutC-GFP11 showed increased soluble protein than that of cFutC-GFP11 on SDS-PAGE. (figure 3.3)

Therefore, we decided it is reasonable to use split-GFP for screening of increased soluble protein of α1,2-FucT.

3.2. Optimization of expression levels of split-GFP

To optimize the expression level of GFP11, the cFutC-GFP11 gene was inserted into pBAD/HisA vector for titratable expression. (figure 3.4) One of the popular expression plasmids, the pET15b vector contains the T7 promoter and lac operon system. The pBAD/HisA vector is one of the pBAD expression systems and contains the araBAD promoter and araBAD operator sequence. The pBAD expression system is known to be effective for the expression of proteins with toxicity and optimizing solubility in *E. coli* by allowing tightly controlled and titratable expression of proteins. [22] When the expression vector of the cFutC-GFP11 gene is changed to the pBAD expression system, cFutC-GFP11 and GFP1-10 genes are more tightly regulated on SDS-PAGE. (figure 3.5)

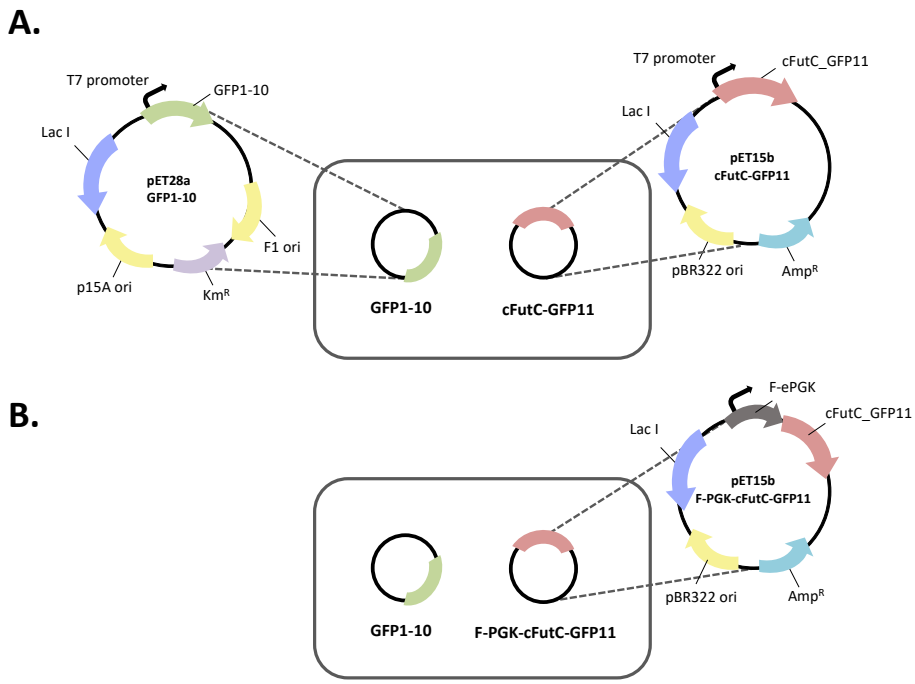


Figure3.1 Vector maps of the split-GFP screening system

(A) Co-expression with GFP1-10 for cFutC-GFP11 in pET15b vector

(B) Co-expression with GFP1-10 for F-ePGK-cFutC-GFP11 in pET15b vector

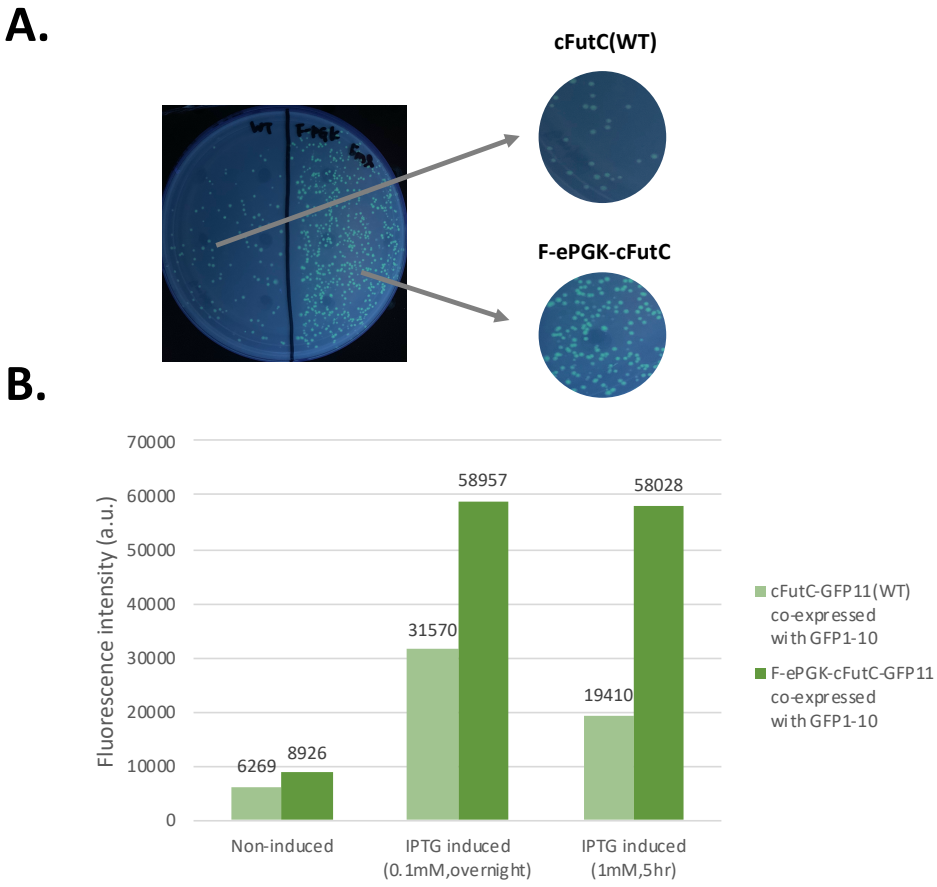


Figure3.2 Verify the feasibility of split-GFP using a mutant that fused F-ePGK into the N-terminal of cFutC as a control

(A) Fluorescence comparison on LB-agar plate

(B) Fluorescence intensity of F-ePGK-cFutC-GFP11 and cFutC-GFP11 induced by IPTG after 16hr at 30°C

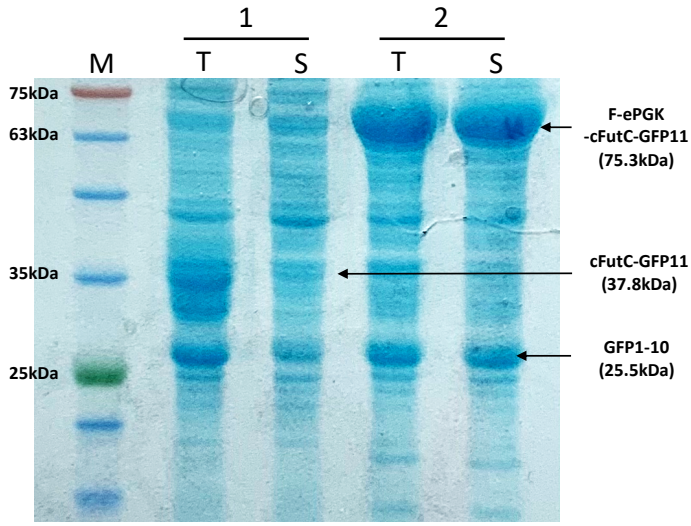


Figure 3.3 Protein expression of cFutC-GFP11 and F-ePGK-cFutC-GFP11

Total (T) and soluble (S) proteins of co-expression with GFP1-10 for cFutC-GFP11 and F-ePGK-cFutC-GFP11 were analyzed by SDS-PAGE. M : protein marker; lane1 : co-expression with GFP1-10 for cFutC-GFP11; lane2 : co-expression with GFP1-10 for F-ePGK-cFutC-GFP11

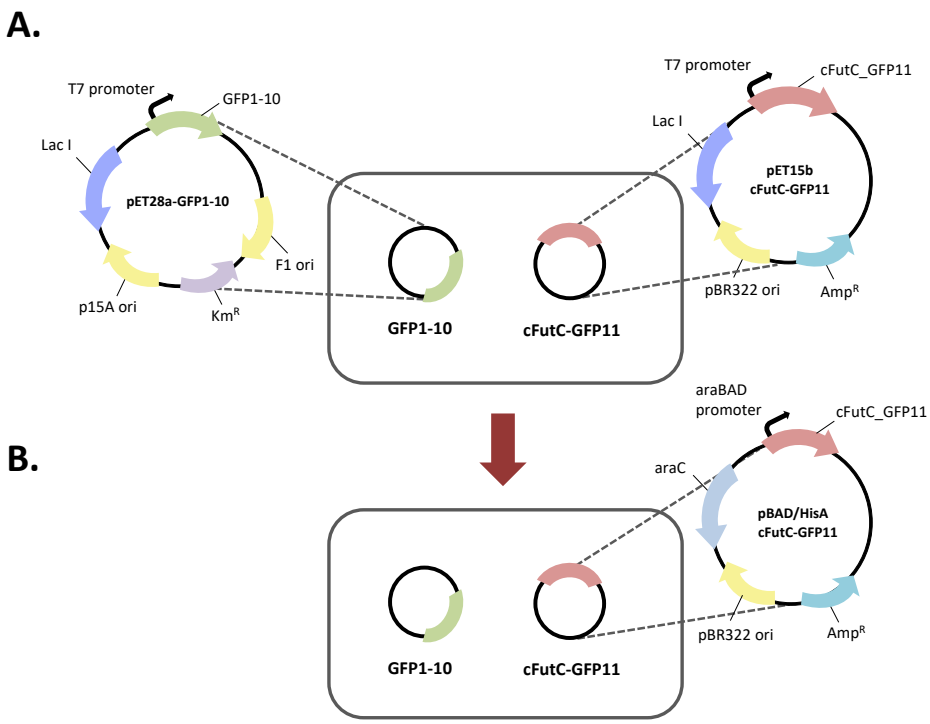


Figure3.4 Optimization of expression levels of split-GFP

(A) Co-expression with GFP1-10 for cFutC-GFP11 in pET15b vector

(B) Co-expression with GFP1-10 for cFutC-GFP11 in pBAD/HisA vector for tight regulation

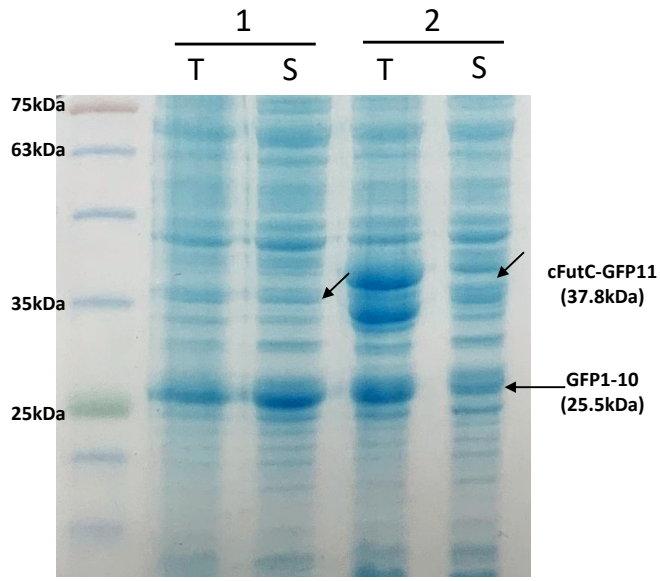


Figure 3.5 Protein co-expression with GFP1-10 for cFutC-GFP11 in pBAD/HisA and pET15b vecor

Total (T) and soluble (S) proteins of co-expression with GFP1-10 for cFutC-GFP11 were analyzed by SDS-PAGE.

M : protein marker;

lane1 : co-expression with GFP1-10 for cFutC-GFP11 in pBAD/HisA vector; lane2 : co-expression with GFP1-10 for cFutC-GFP11 in pET15b vector

3.3. Construction of the mutant library using SSM

To construct the mutant library, helical wheel diagrams of 6 α-helix in cFutC were structured by applying the α-helix rule. (Figure 3.6) After that, two or three hydrophobic amino acids present in hydrophilic regions of each α-helix were targeted. (table3.1) Amino acids targeted in one α-helix were grouped into one group and performed site saturation mutagenesis. Each GFP11 fused mutant library was developed using site saturation mutagenesis with NNK degenerated codon. Each GFP11 fused mutant library was transformed into *E. coli* :: PtacΔZ with GFP1-10 gene for protein co-expression and it contains more than 10^5 colonies. After induction by IPTG, each mutant library was prepared for FACS screening.

3.4. High-throughput screening using flow cytometry

First, each mutant library was screened using flow cytometry to determine whether there is a correlation between the increase in soluble expression of cFutC and the fluorescence intensity of split GFP. First, cells showing the highest 10% fluorescence were separated and recovered for second round sorting. (figure 3.7) After that, enriched cells showing the highest 10% fluorescence were separated by the same method and picked for further screening in a 96-well deep plate. In each plate, four wild type controls were included. After induction by arabinose and IPTG in a 96-well deep plate, fluorescence intensity and OD600 were measured. (figure 3.8) After that, each mutant with

more than 1.8-fold in fluorescence intensity/OD600 than average of four wild types was selected.

According to the sequencing, 15 mutants contained the amino acid mutations to more hydrophilic amino acids than wild type amino acid were obtained. (table3.2) Therefore, according to the α-helix rule, when hydrophobic amino acids are substituted with hydrophilic amino acids, soluble protein expression was increased, which leads to an increase in split GFP intensity.

3.5. Soluble expression of α1,2-FucT

We expected that the mutants with higher fluorescence intensity/OD600 than the wild type cFutC have higher soluble protein fraction on SDS-PAGE. Therefore, SDS-PAGE analysis was performed to examine the increase in protein soluble expression of the mutants selected through high-throughput screening. To analysis the expression level of soluble protein, the mutant gene was inserted into pET24ma vector, One of the popular expression plasmids. After construction, the selected mutants were cultured to analyze total and soluble fraction of protein by SDS-PAGE gel. The results demonstrated that the fraction of soluble protein was similar to the wild type or slightly increased in the majority of mutants. (figure 3.9)

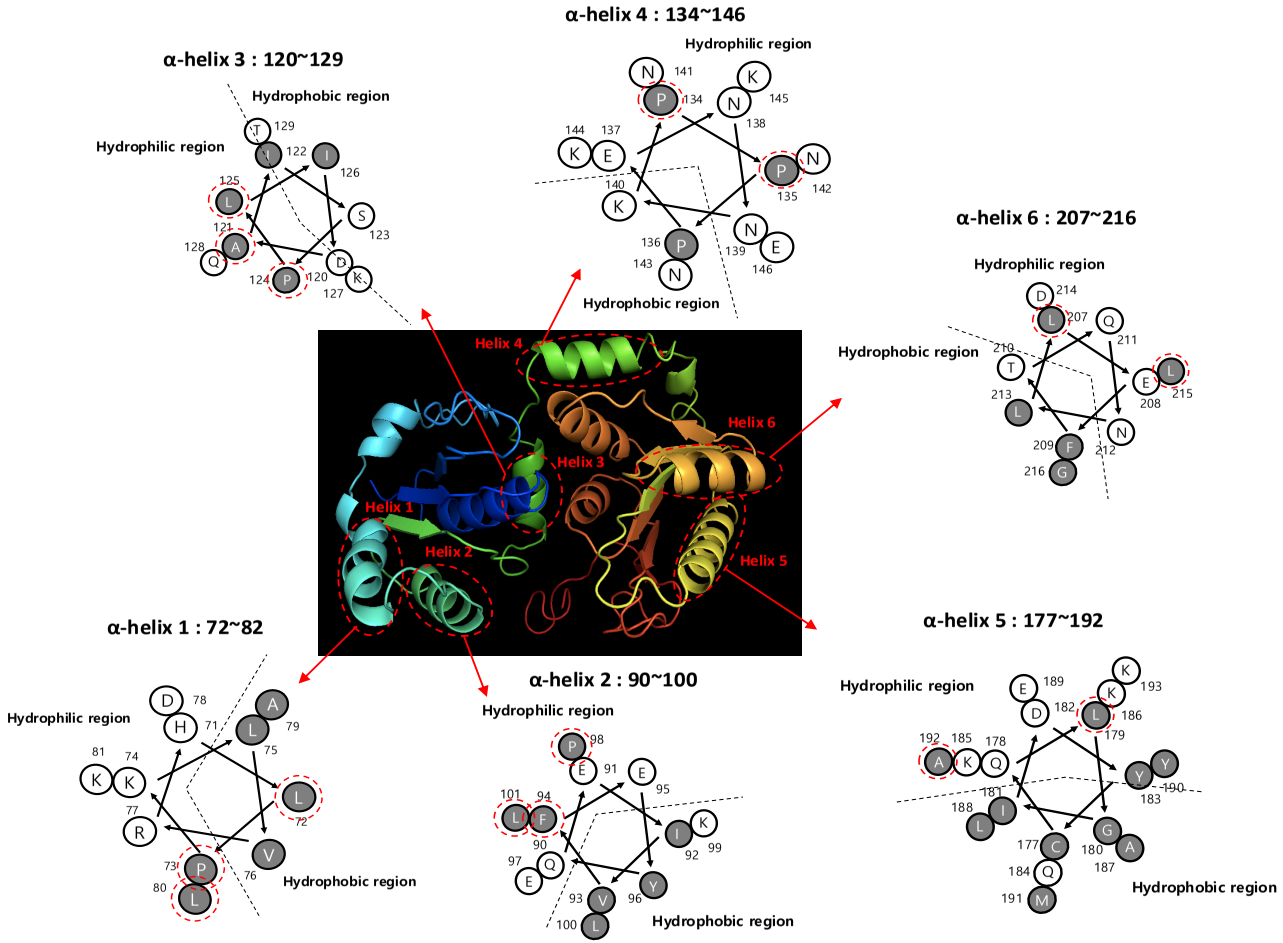


Figure 3.6 Selected hydrophobic targets in hydrophilic regions using α-helix rule

Group	Selected targets
α-helix 1	L72, P73, L80
α-helix 2	F94, P98, L101
α-helix 3	A121, P124, L125
α-helix 4	P134, P135
α-helix 5	L179, A192
α-helix 6	L207, L215

Table 3.1 Selected targets applying the α-helix rule

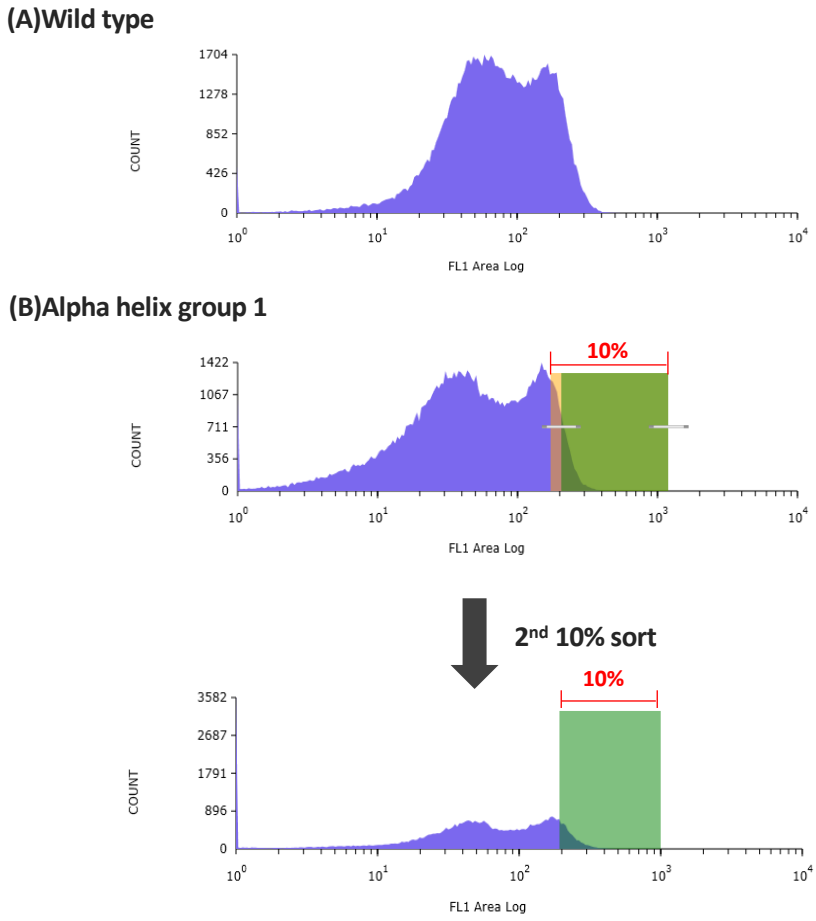


Figure 3.7 FACS analysis of alpha-helix group1 library

Cells with the 10% highest fluorescence intensity were separated and recovered for second round sorting.

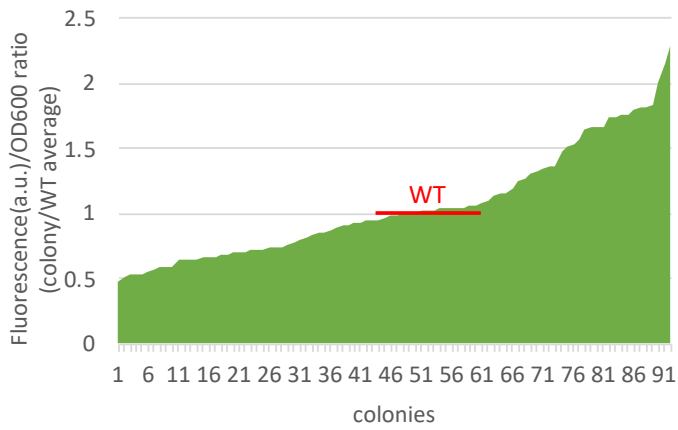


Figure 3.8 Further screening of alpha-helix group1 library in 96-well plate

Each whole-cell fluorescence intensity and OD600 were measured using fluorescence spectrometry. In each plate, four wild type controls were included. (four wild type controls).

Group	fluorescence/OD600	Mutation
α-helix 1	2.29	L80Y(TAT)
	2.01	L80C(TGT)
α-helix 2	1.84	F94K(AAG)/P98P(CCG)
α-helix 3	2.40	A121D(GAT)/P124A(GCG)/L125V(GTT)
	2.02	A121A(GCG)/P124G(GGT)/L125R(AGG)
	1.89	A121S(TCG)/P124A(GCG)/L125M(ATG)
	1.81	A121A(GCG)/P124T(ACG)/L125P(CCT)
α-helix 4	2.34	P134Q(CAG)/P135A(GCG)
	2.24	P134P(CCG)/P135R(AGG)
	2.09	P134W(TGG)/P135L(CTT)
α-helix 6	2.53	L215V(GTT)
	2.23	L207R(AGG)
	2.10	L215I(ATT)

Table 3.2 Sequence data of improved mutants

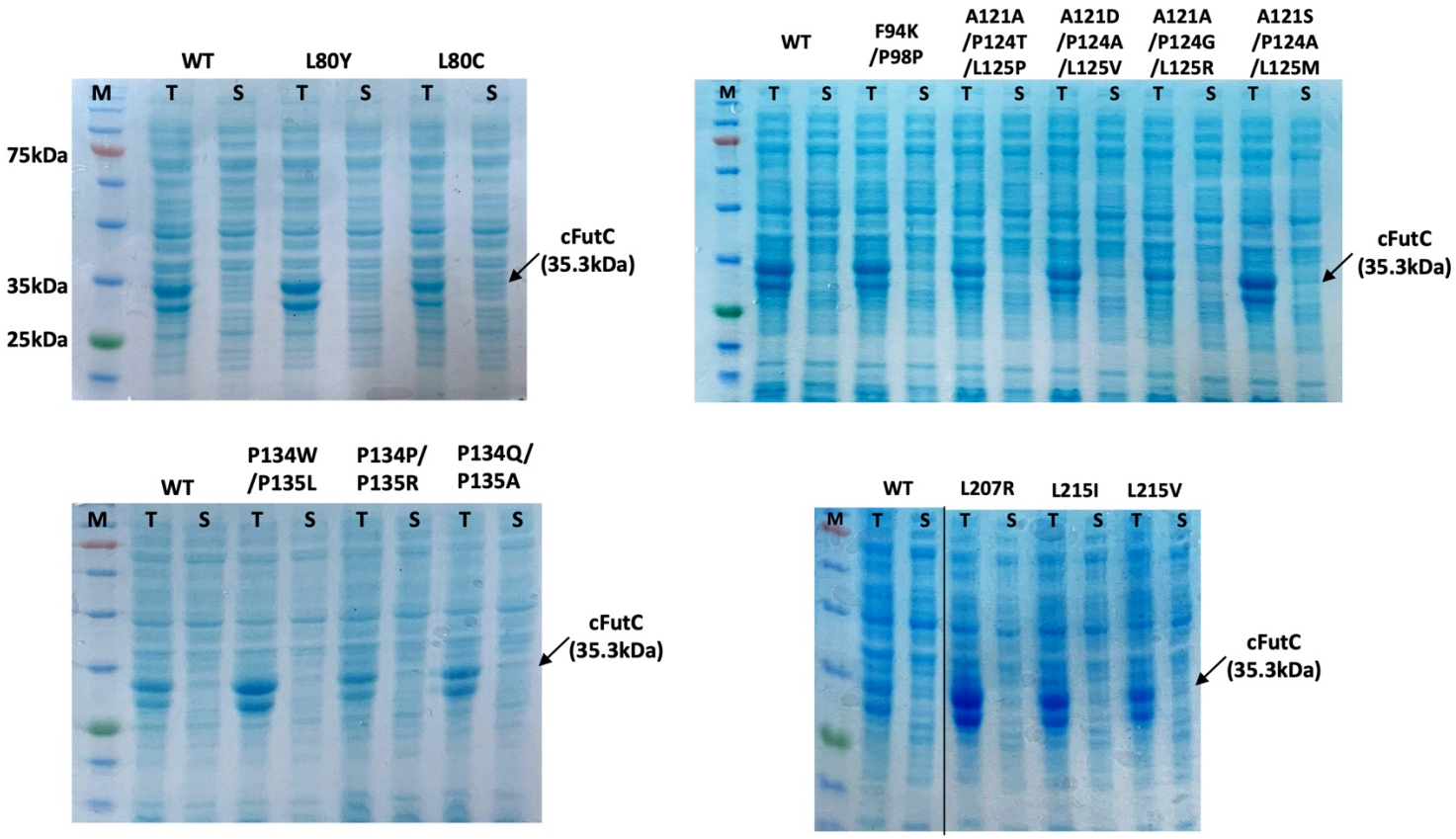


Figure 3.9 Protein expression for cFutC mutants with improved fluorescence/OD600 were analyzed by SDS-PAGE

M : protein marker; T: total; S : soluble proteins

3.6. *In vivo* 2'-FL production using α1,2-FucT mutants

The increase in 2'-FL titer of selected mutants, 13 mutants in which the fluorescence intensity/OD600 increased compared to the wild type, were cultured to produce 2'-FL in shake flask. Glucose was added as a C-source to synthesize GDP-L-fucose through the *de novo* pathway, and then lactose was added as another sugar backbone, to the culture medium.

After 16 hours of batch culture, 12 mutants except for two mutants, L80C and A121D/P124A/L125V, were not significant in the increases *in vivo* production of 2'-FL titers. (figure 3.10-3.13) But the L80C mutant produced 1.39g/L 2'-FL, which was 1.29-fold higher 2'-FL titer as compared to the wild type. (1.07±0.02 vs 1.39±0.17 g/L) (figure 3.10) Specific yield was 1.40-fold higher over the wild type. (0.139±0.002 vs 0.194±0.038 g/g DCW) (figure 3.12) Also, the A121D/P124A/L125V triple mutant produced 1.66g/L 2'-FL, which was 1.12-fold higher 2'-FL titer as compared to the wild type, after 16 hours of batch culture. (1.48±0.28 vs 1.66±0.32 g/L) (figure 3.10) Specific yield was 1.19-fold higher over the wild type. (0.129±0.002 vs 0.153±0.006g/g DCW) (figure 3.12)

3.7. 2'-FL synthesis using quadruple mutant

After that, batch fermentation was performed on three single mutants of the A121D/P124A/L125V mutants, A121D, P124A, and L125V mutants, to examine the effect of every single mutation. After 16 hours of batch fermentation, the A121D, P124A, and L125V mutants

produced 1.64g/L, 1.33g/L, 1.76g/L 2'-FL, which were comparable to 2'-FL titer of wild type. (1.48±0.28 vs 1.64±0.37, 1.33±0.03, 1.76±0.14 g/L) (figure 3.14) Each Specific yield was comparable or slightly decreased than that of the wild type. (0.164 ±0.002vs 0.161±0.010, 0.120 ±0.001, 0.148 ±0.012 g/g DCW) (figure 3.14) We conclude that each single mutation did not significantly affect the increase in 2'-FL production, but the triple mutant has a synergistic effect through each mutation.

After that, we generated a quadruple mutant (L80C/A121D/P124A/L125V) and shake flask batch fermentation was performed on that mutant. After 16 hours of batch culture, the quadruple mutants produced 3.21g/L 2'-FL, which was up to 1.73-fold higher 2'-FL titer than titer of the wild type. (1.86±0.23 vs 3.21±0.64 g/L) Specific yield was 1.83-fold higher over the wild type. (0.176±0.011 vs 0.322±0.059 g/g DCW) (figure 3.15)

In both single (L80C) and triple (A121D/P124A/L125V) mutants, mutations were occurred into amino acids with more hydrophilicity than existing amino acids. These mutants showed slightly increased 2'-FL titer and specific yield than wild type. Also, a quadruple mutant was assembled, which showed up to 1.73-fold enhancement in 2'-FL titer than that of the wild type. The results show us the possibility of application of high-throughput flow cytometry screening using a split-GFP for screening enhanced soluble expression of enzyme and target materials, mass production of 2'-FL in *E. coli* by microbial synthesis.

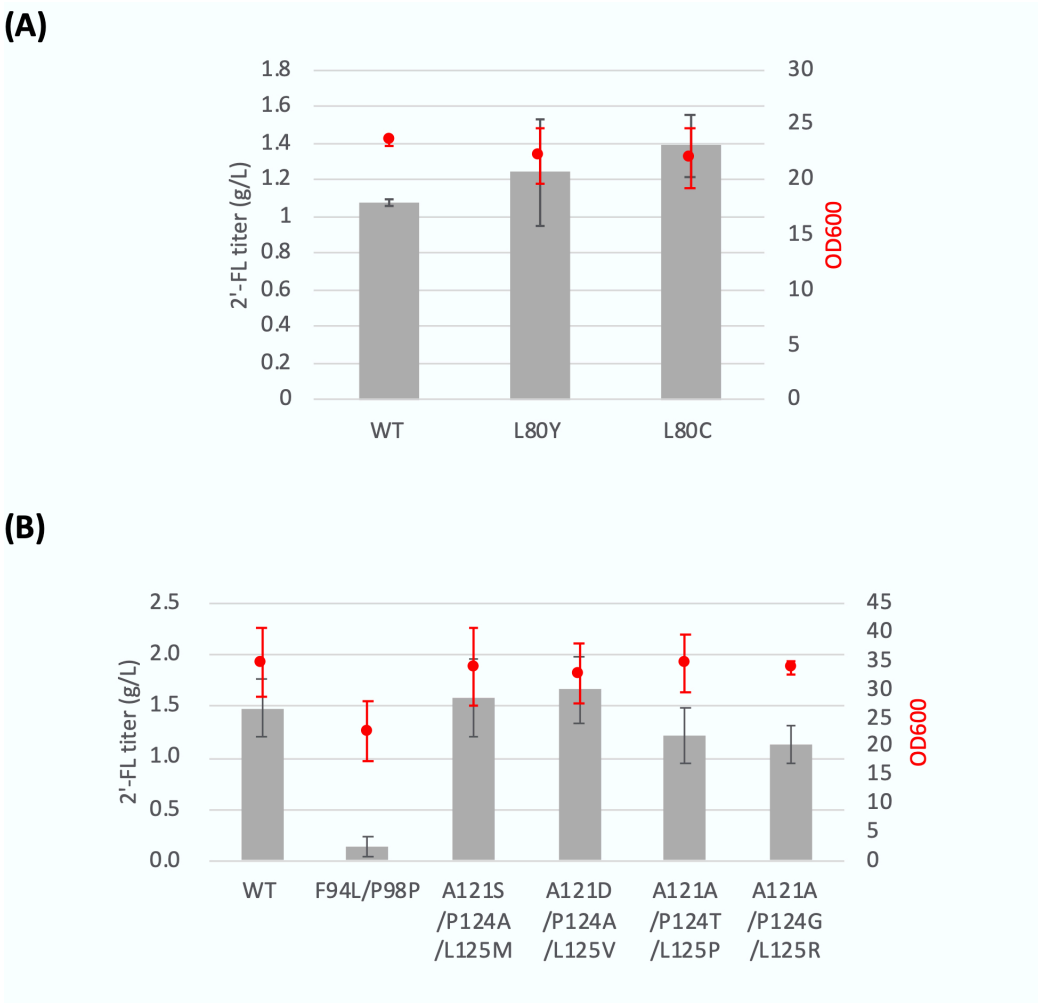


Figure 3.10 2'-FL production and OD600 by batch shake flask culture with selected mutants in alpha-helix group 1-3

(A) 2'-FL production and OD600 by batch shake flask culture with selected mutants of alpha-helix group 1

(B) 2'-FL production and OD600 by batch shake flask culture with selected mutants of alpha-helix group 2 and 3

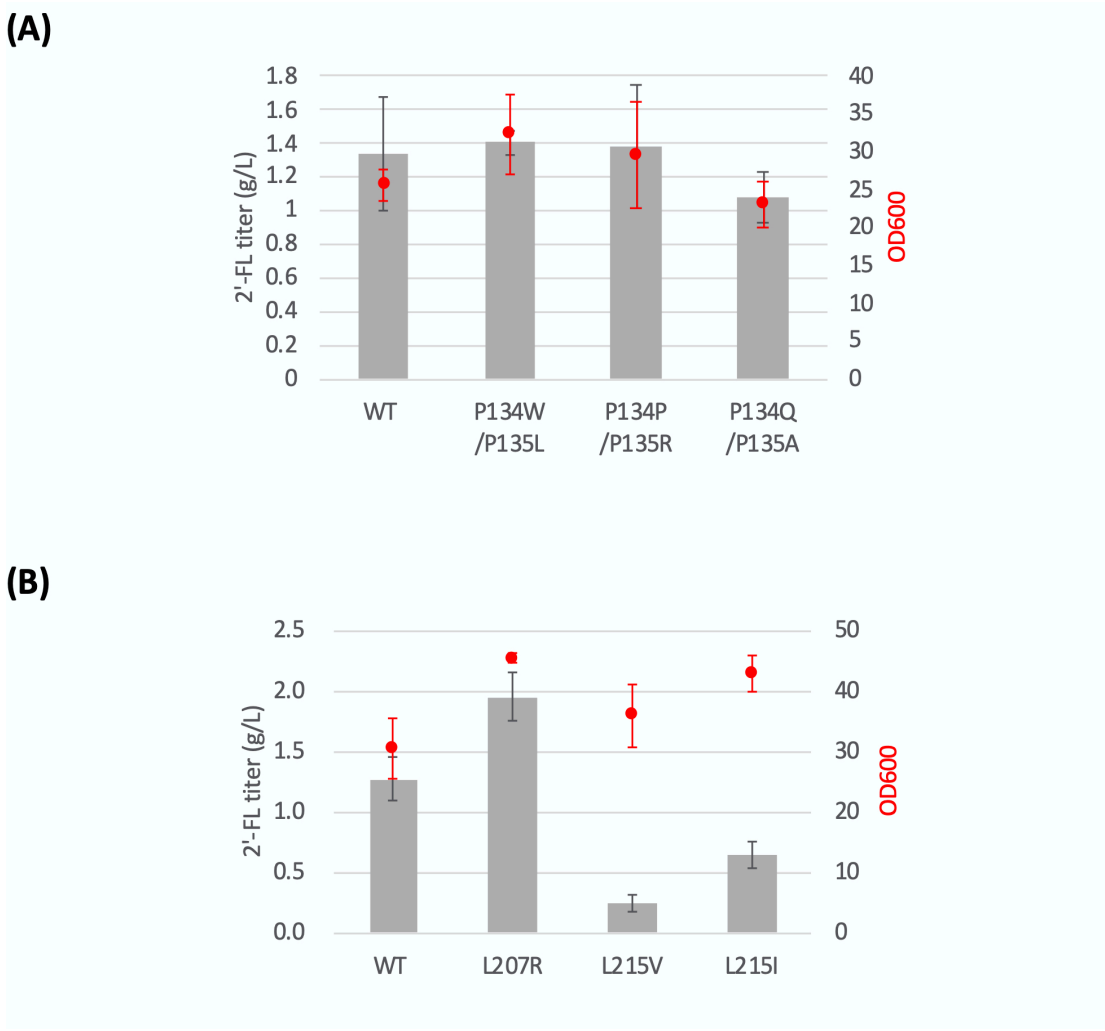


Figure 3.11 2'-FL production and OD600 by batch shake flask culture with selected mutants in alpha-helix group 4-6

(A) 2'-FL production and OD600 by batch shake flask culture with selected mutants of alpha-helix group 4

(B) 2'-FL production and OD600 by batch shake flask culture with selected mutants of alpha-helix group 5 and 6

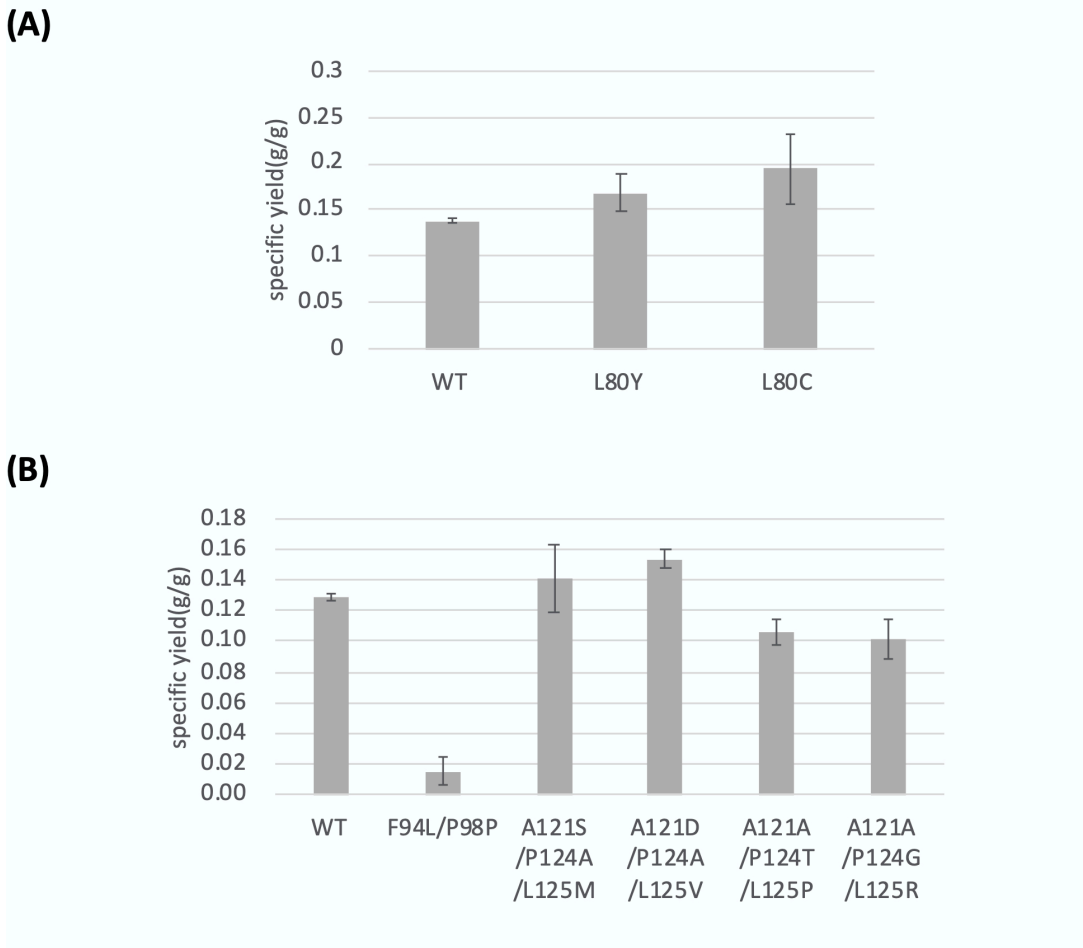


Figure 3.12 Specific yield of 2'-FL (g/g) by batch shake flask culture with selected mutants in alpha-helix group 1-3

(A) Specific yield of 2'-FL (g/g) by batch shake flask culture with selected mutants of alpha-helix group 1

(B) Specific yield of 2'-FL (g/g) by batch shake flask culture with selected mutants of alpha-helix group 2 and 3

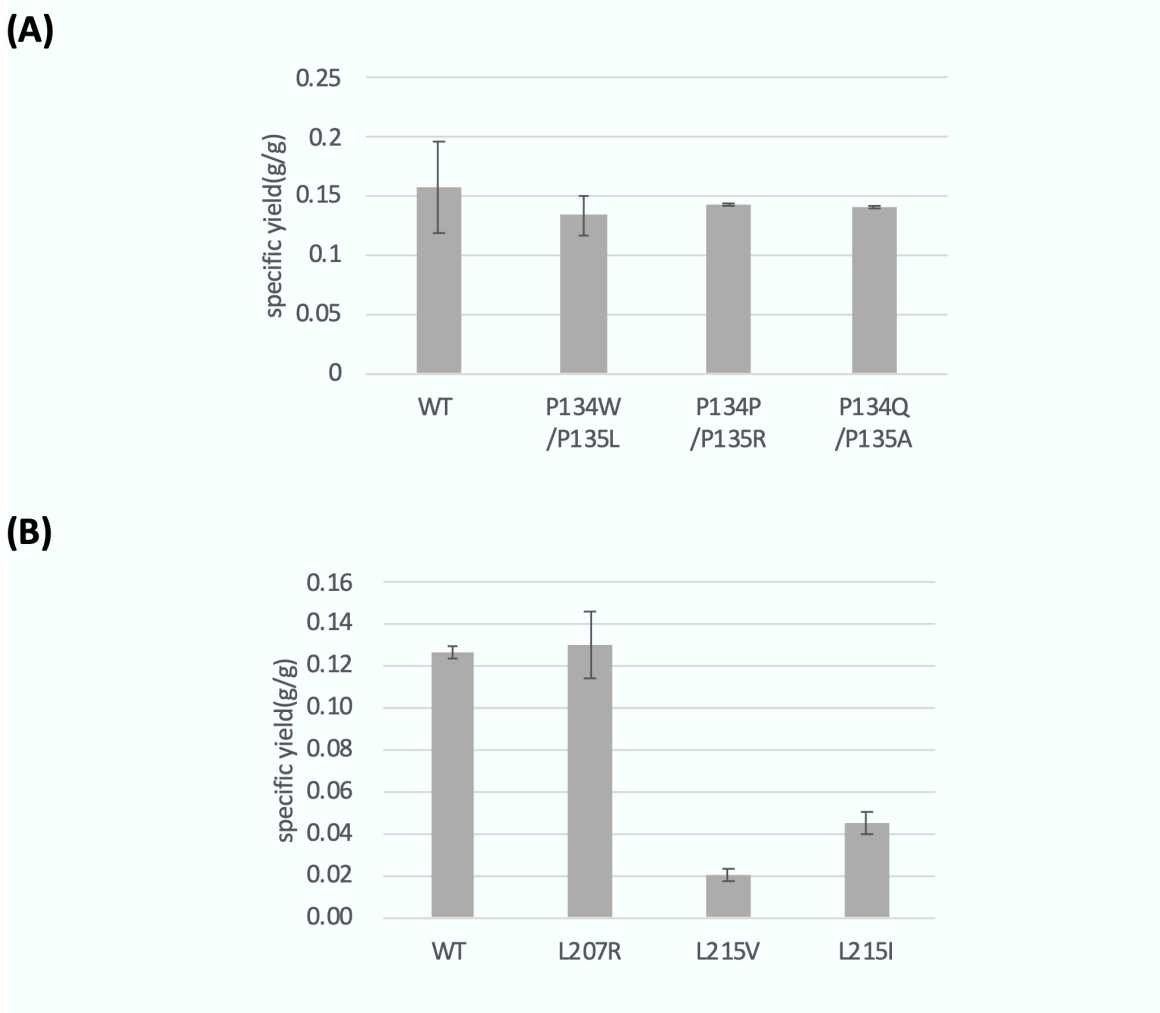


Figure 3.13 Specific yield of 2'-FL (g/g) by batch shake flask culture with selected mutants in alpha-helix group 4-6

(A) Specific yield of 2'-FL (g/g) by batch shake flask culture with selected mutants of alpha-helix group 4

(B) Specific yield of 2'-FL (g/g) by batch shake flask culture with selected mutants of alpha-helix group 6

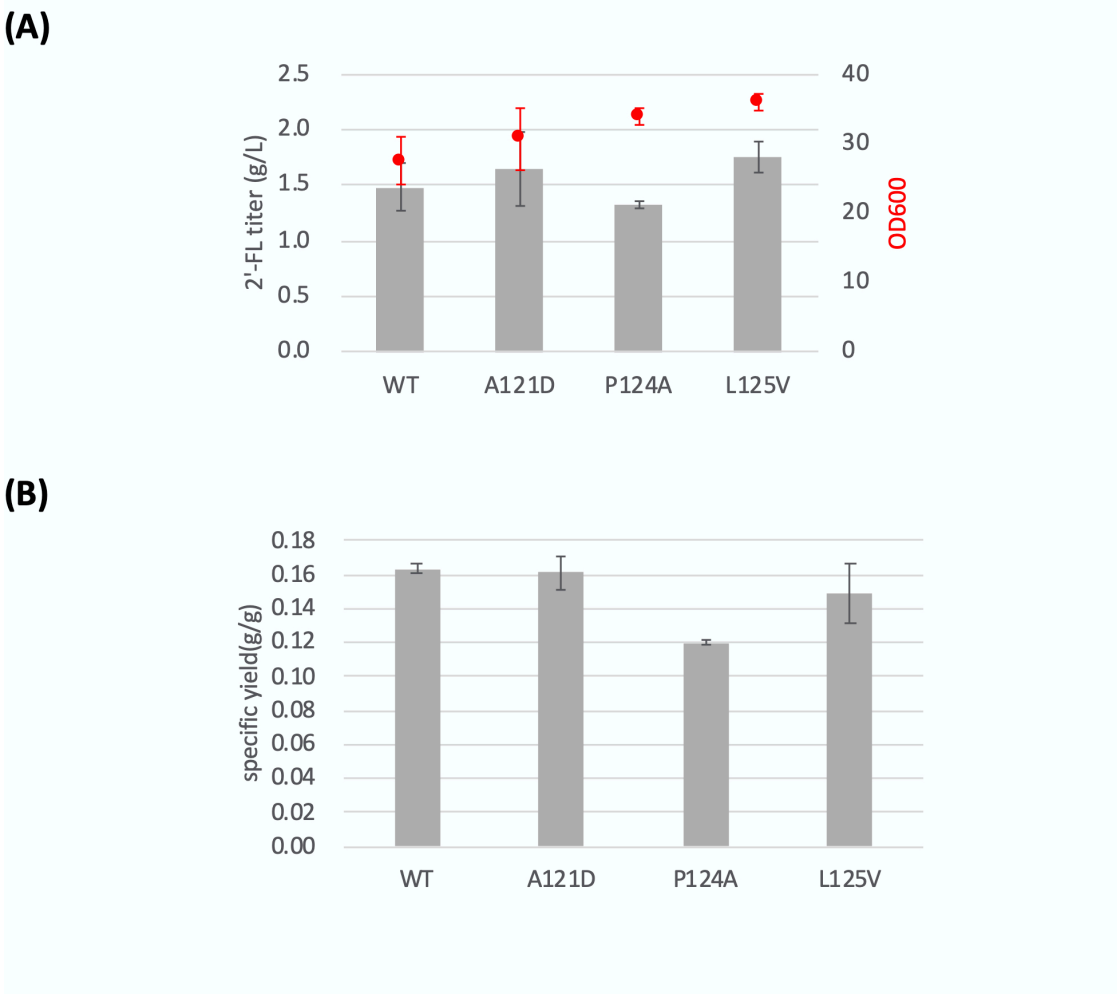


Figure 3.14 2'-FL production, OD600 and specific yield of 2'-FL (g/g) by batch shake flask culture with each single mutant of A121D/P124A/L125V

(A) 2'-FL production and OD600 by batch shake flask culture with each single mutant of A121D/P124A/L125V

(B) Specific yield of 2'-FL (g/g) by batch shake flask culture with each single mutant of A121D/P124A/L125V

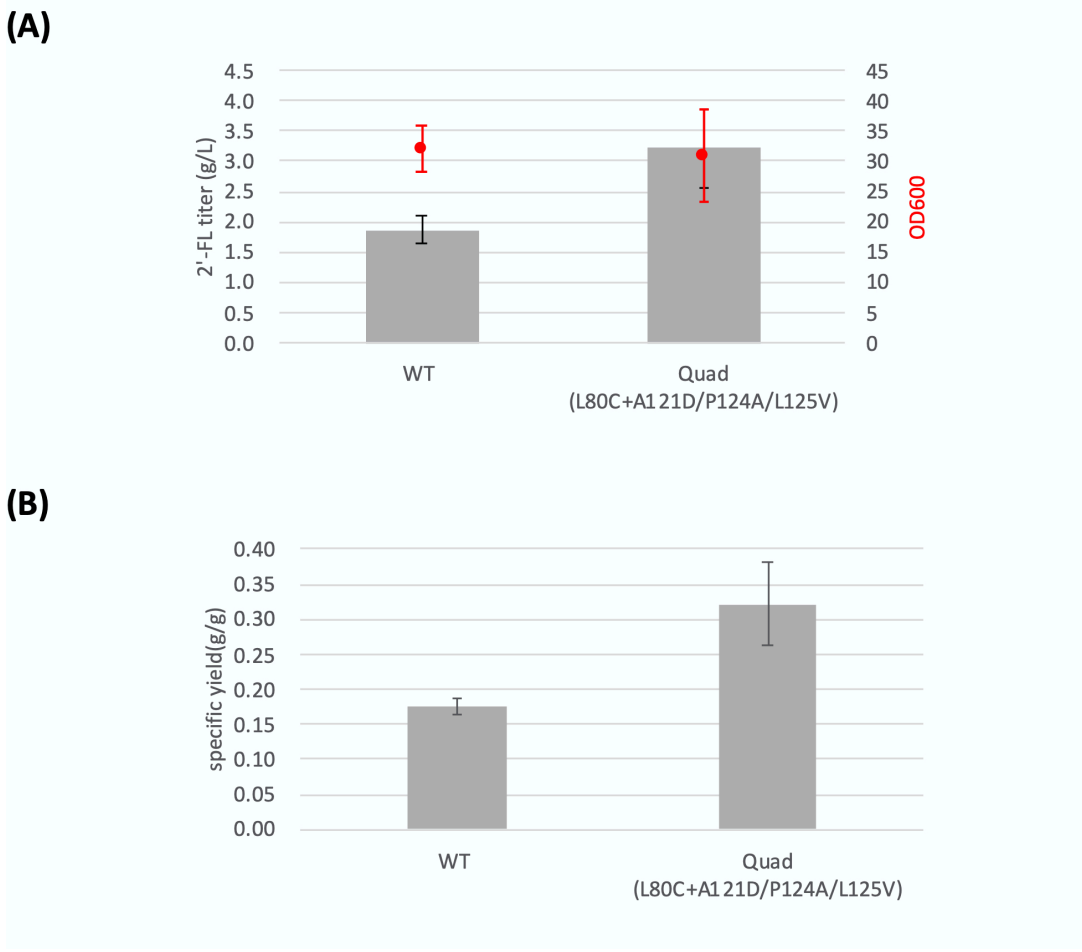


Figure 3.15 2'-FL production, OD600 and specific yield of 2'-FL (g/g) by batch shake flask culture with quadruple mutant (L80C/A121D/P124A/L125V)

(A) 2'-FL production and OD600 by batch shake flask culture with quadruple mutant (L80C/A121D/P124A/L125V)

(B) Specific yield of 2'-FL (g/g) by batch shake flask culture with quadruple mutant (L80C/A121D/P124A/L125V)

4. CONCLUSIONS

In this study, the soluble expression of α1,2-FucT which produces 2'-FL using substrates, GDP-L-fucose and lactose, was increased using site saturation mutagenesis through the α-helix rule. Also, the high-throughput screening using split-GFP system with flow cytometry was tested for large mutant libraries in *E. coli*. 2'-FL is important oligosaccharides in human milk and has potential advantages to affect the growth of gut microorganisms and the development of immunity in infants. However, the low soluble expression of α1,2-FucT acts as a limitation for industrial use of enzyme, for increasing the production of 2'-FL in *E. coli*, it is important to enhance the low solubility of α1,2-FucT.

To enhance the protein soluble expression of α1,2-FucT, we measured the feasibility of the split-GFP system for screening the α1,2-FucT solubility first and mutant that fused F-ePGK into the N-terminal of cFutC was used as a positive control. Each cFutC and F-ePGK-cFutC was fused with GFP11 to the C-terminal of the gene and transformed into *E. coli* BL21(DE3) cells with GFP1-10 gene to form a complete fluorescence. After induction, the fluorescence intensity of soluble mutant, F-ePGK-cFutC-GFP11 was more than 3-fold than the fluorescence intensity of cFutC-GFP11. Therefore, it was supported that split-GFP can be used for α1,2-FucT solubility screening. After that, for optimization of expression level, the cFutC-GFP11 gene was inserted into pBAD/HisA vector to regulate expression tightly. Next,

aggregation hotspots in cFutC were targeted using the α-helix rule, and site saturation mutagenesis was applied at hydrophobic amino acids present in the hydrophilic regions of each α-helix using degenerated codons.

Each mutant library was screened using Fluorescence-activated cell sorter (FACS) for powerful and high-throughput screening of large mutant libraries up to 10^{10}. After sorting the cells with the 10% highest fluorescence twice, whole-cell fluorescence/OD600 of the colonies measured by fluorescence spectrometer for further screening. We obtained 15 mutants that contained the mutations to more hydrophilic amino acids than the wild type, and SDS-PAGE analysis was performed to examine the increase in protein soluble expression of these mutants. Each mutant gene was inserted into pET24ma vector and was cultured to perform the SDS-PAGE analysis. The results showed that soluble protein expressions of the L80C and the A121D/P124A/L125V mutant were similar to the wild type or slightly increased.

The increase in 2'-FL titer of mutants with an enhancement of soluble protein expression was examined through shake flask culture. Glucose and another sugar backbone, lactose was fed for batch fermentation. After 16 hours of batch fermentation, the L80C mutant produced 1.39g/L 2'-FL, which was 1.29-fold higher 2'-FL titer than the wild type. And the A121D/P124A/L125V triple mutant produced 1.66g/L 2'-FL, which was 1.12-fold higher 2'-FL titer than wild type. As a result, a quadruple mutant(L80C/A121D/P124A/L125V) was

generated, which showed 1.73-fold improvement in productivity relative to wild-type.

In summary, each mutant library was prepared through site-saturation mutagenesis and hydrophobic amino acids present in the hydrophilic regions of each α-helix were targeted. Mutants with increased soluble expression were screened using a split-GFP system and high-throughput flow cytometry. As a result, mutants with an increase in soluble expression were screened, and increases in the titer and specific yield of 2'-FL were examined. The results show us the possibility of application of high-throughput flow cytometry screening using a split-GFP for screening enhanced soluble expression of enzyme, production of target materials, and industrial application of 2'-FL in *E. coli*. Also, if the combination of enzyme activity measurements and split-GFP system can reduce false negatives and false positives, it is expected that the efficiency of the high-throughput screening system can be improved.

Reference

1. Vandenplas, Y., et al., *Human milk oligosaccharides: 2′-fucosyllactose (2′-FL) and lacto-N-neotetraose (LNnT) in infant formula.* Nutrients, 2018. **10**(9): p. 1161.
2. Bode, L., *The functional biology of human milk oligosaccharides.* Early human development, 2015. **91**(11): p. 619-622.
3. Barile, D. and R.A. Rastall, *Human milk and related oligosaccharides as prebiotics.* Current opinion in biotechnology, 2013. **24**(2): p. 214-219.
4. Kunz, C., et al., *Oligosaccharides in human milk: structural, functional, and metabolic aspects.* Annual review of nutrition, 2000. **20**(1): p. 699-722.
5. Liu, J.-J., et al., *Biosynthesis of a functional human milk oligosaccharide, 2′-fucosyllactose, and l-fucose using engineered Saccharomyces Cerevisiae.* ACS synthetic biology, 2018. **7**(11): p. 2529-2536.
6. Chin, Y.-W., et al., *Improved production of 2′-fucosyllactose in engineered Escherichia coli by expressing putative α-1, 2-fucosyltransferase, WcfB from Bacteroides fragilis.* Journal of biotechnology, 2017. **257**: p. 192-198.
7. Eiwegger, T., et al., *Human milk-derived oligosaccharides and plant-derived oligosaccharides stimulate cytokine production*

of cord blood T-cells in vitro. Pediatric research, 2004. **56**(4): p. 536-540.

8. Baumgärtner, F., et al., *Construction of Escherichia coli strains with chromosomally integrated expression cassettes for the synthesis of 2′-fucosyllactose.* Microbial cell factories, 2013. **12**(1): p. 40.

9. Hollands, K., et al., *Engineering two species of yeast as cell factories for 2′-fucosyllactose.* Metabolic engineering, 2019. **52**: p. 232-242.

10. Parschat, K., et al., *High-Titer De Novo Biosynthesis of the Predominant Human Milk Oligosaccharide 2′-Fucosyllactose from Sucrose in Escherichia coli.* ACS Synthetic Biology, 2020. **9**(10): p. 2784-2796.

11. Yu, J., et al., *Engineering of α-1, 3-fucosyltransferases for production of 3-fucosyllactose in Escherichia coli.* Metabolic engineering, 2018. **48**: p. 269-278.

12. Paraskevopoulou, V. and F.H. Falcone, *Polyionic tags as enhancers of protein solubility in recombinant protein expression.* Microorganisms, 2018. **6**(2): p. 47.

13. Hayhurst, A. and W.J. Harris, *Escherichia coliskp chaperone coexpression improves solubility and phage display of single-chain antibody fragments.* Protein expression and purification, 1999. **15**(3): p. 336-343.

14. Matsui, D., et al., *Rational identification of aggregation hotspots based on secondary structure and amino acid hydrophobicity.* Scientific reports, 2017. **7**(1): p. 1-12.

15. Cabantous, S., T.C. Terwilliger, and G.S. Waldo, *Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein.* Nature biotechnology, 2005. **23**(1): p. 102-107.

16. Cabantous, S. and G.S. Waldo, *In vivo and in vitro protein solubility assays using split GFP.* Nature methods, 2006. **3**(10): p. 845-854.

17. Kakimoto, Y., et al., *Visualizing multiple inter-organelle contact sites using the organelle-targeted split-GFP system.* Scientific reports, 2018. **8**(1): p. 1-13.

18. Yang, G. and S.G. Withers, *Ultrahigh-throughput FACS-based screening for directed enzyme evolution.* ChemBioChem, 2009. **10**(17): p. 2704-2715.

19. Choi, Y.H., et al., *Biosynthesis of the human milk oligosaccharide 3-fucosyllactose in metabolically engineered Escherichia coli via the salvage pathway through increasing GTP synthesis and β-galactosidase modification.* Biotechnology and bioengineering, 2019. **116**(12): p. 3324-3332.

20. 최윤희, *Strategy for protein engineering of glycosyltransferase using hybrid approach and biosynthesis of sialyl and fucosyl oligosaccharides*. 2016, 서울대학교 대학원.

21. Choi, Y.H., et al., *Solubilization and Iterative Saturation Mutagenesis of a1, 3-fucosyltransferase from Helicobacter pylori to enhance its catalytic efficiency.* Biotechnology and bioengineering, 2016. **113**(8): p. 1666-1675.

22. Giacalone, M.J., et al., *Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system.* Biotechniques, 2006. **40**(3): p. 355-364.

국문초록

대장균에서의 α1,2-퓨코당 전이효소의 용해도 향상을 위한 split-GFP 및 유세포 분석기를 결합한 고속 대량 스크리닝 시스템의 적용

최근 모유 유래 올리고당이 생체 내에서 가지는 다양한 생물학적 기능에 관한 연구가 보고되고 있다. 모유 유래 올리고당은 장내 유용 미생물의 성장을 촉진하는 프리바이오틱스로 사용된다. 또한 인간의 병원균이 장 점막에 부착하는 것을 방지하는 등의 효과가 있어 면역 조절에도 도움을 준다. 모유 유래 올리고당의 한 종류인 2'-퓨코실 락토오스는 높은 생산성을 가지도록 설계된 대장균과 같은 다양한 미생물에서 생산될 수 있다. 일반적으로, α1,2-퓨코당 전이효소는 구아노신-5'-다이포스포-베타-L-퓨코즈 (GDP-Fuc)로부터 락토오스로 퓨코실기를 전이시켜 2'-퓨코실 락토오스를 합성한다. 하지만 α1,2-퓨코당 전이효소의 낮은 용해도는 반응에 있어 속도 결정 단계이며, 산업적으로 효소를 응용하기에 한계로 작용한다.

α1,2-퓨코당 전이효소의 용해도 문제를 해결하기 위해, 다음과 같이 α1,2-퓨코당 전이효소의 용해도를 향상시키기 위한 연구가 수행되었다. 우선 알파 헬릭스 룰에 따라 선정된 소수성 아미노산에 site saturation mutagenesis(SSM)를 진행하여 α1,2-퓨코당 전이효소의 변이주 라이브러리를 제작하였다. 효소의 표면에 위치하는 알파 헬릭스에 대한 헬리컬 휠을 도식화 하였다. 이후 알파 헬릭스의 친수성 영역에 존재하는 소수성 아미노산을 선정하고 그룹화 하였다. 이때 알파 헬릭스

별로 6 개의 그룹을 나누어 라이브러리를 제작하였으며, 제작된 라이브러리는 split-GFP 시스템을 적용하여 스크리닝하였다. split-GFP 중 GFP11 은 α1,2-퓨코당 전이효소 의 C-말단에 연결되어 α1,2-퓨코당 전이효소의 용해도에 따라 발현 효율에 영향을 받게 된다. 따라서 GFP1-10 과 GFP 11 의 결합에 따른 완전한 GFP 형광 세기에 따라 α1,2-퓨코당 전이효소 의 용해도를 구별할 수 있다. 제작된 라이브러리의 High-Throughput Screening 을 위하여 FACS 장비를 도입하여 상위 10%의 형광 세기를 가지는 세포를 2 단계로 분리하여 변이주를 선정하였다. 그 후 선정된 변이주에 대한 96-well plate 스크리닝을 통해 OD600 대비 형광값을 비교하여 최종 변이주 후보군을 선정하였다.

선정된 α1,2-퓨코당 전이효소 변이주는 대장균 내에서 락토오스와 글루코스를 기질로 하여 *de novo* pathway 를 통해 2'-퓨코실 락토오스를 생산하였다. 16 시간의 batch culture 를 통해 실제 2'-퓨코실 락토오스의 생산량 증가 및 enzyme specific activity 의 증가를 bio-LC 를 사용하여 분석하였다. 이와 같은 돌연변이를 통해 용해도가 증가한 완성된 4 개의 변이가 도입된 α1,2-퓨코당 전이효소는 , 72%의 2'-퓨코실 락토오스의 생산량 증가를 가져왔으며, 이는 2'-퓨코실 락토오스를 대량 생산하기 위한 산업적인 효소 적용에 기여할 것으로 생각한다. 결과적으로, split-GFP 시스템과 FACS 를 사용한 고속 대량 스크리닝은 효소의 용해도 증가와 이를 통한 물질의 생산량 증가를 통해 산업적인 대량 생산을 용이하게 하는 것에 있어 기여할 것이다.

주요어: 모유 유래 올리고당, 2'- 퓨코실 락토오스, α1,2-퓨코당 전이효소, 알파 헬릭스 룰, 유세포 분석기, 분할 녹색 형광 단백질, 단백질 공학

학번 : 2019-24630