Confidential Research Report
Zoey Jung-Yu Liu
(Student number: 7074824)

Identification and Characterization of novel L-xylulose reductase encoding genes in *Aspergillus niger*

Jung-Yu Liu, BSc

Major Research Project
Bio Inspired Innovation Programme
Graduate School of Life Sciences
Utrecht University

Daily Supervisor: Astrid Mueller
Second Reviewer: Mao Peng

Fungal Physiology – de Vries Group
Westerdijk Fungal Biodiversity Institute
Utrecht, the Netherlands

Examiner: Prof. Dr. Ir. Ronald P. de Vries

Molecular Plant Physiology
Environmental Biology
Utrecht University

26-09-2021

WESTERDIJK FUNGAL BIO DIVERSITY INSTITUTE

Utrecht University

Preface

This report represents the research that I performed in the Fungal Physiology group of Westerdijk Fungal Biodiversity Institute from February to September 2021.

First of all, I am grateful to Ronald de Vries for giving me the opportunity to do my major research project in the Fungal Physiology group and holding a full environment with a positive attitude to stimulate good research to happen. Furthermore, I truly appreciate all the help from my supervisors Astrid Mueller and Agata Terebieniec. As my daily supervisor, Astrid has been supporting me since the very beginning, from the guidance of matters in Westerdijk Institute to a lot of instructions on the laboratory work, and even the comfort for mitigating my stress. With Astrid's help, I had a moderate start at a critical time for me. Besides, I have learned a lot of details in research from Agata, who was the agent supervisor in the last phase of my project. The thinking between theory and practical work, the ways of finding resources, and the strictness of principles are what I would like to take with me in my following career. Furthermore, I am thankful to all the colleagues in Westerdijk Institute, including people in the Practical Room, subgroup- Metabolism, Fungal Natural Products group, and the Monday Seminar. A good work atmosphere has been created in this research environment.

This Major Research Project, being part of the integral training of the Graduate School of Life Sciences of Utrecht University, is a huge step for me on the path I have been chasing. It is my fortune to be at this institute to complete the essential learning. Because of being back in the scientific field after 3-4 years of training in other different domains, a hard time happened to me as expected. However, an interesting combination of thinking and doing has been shaped as well.

Utrecht, 26 September 2021

Zoey Jung-Yu Liu

List of Abbreviations

AE	ArcticExpress (DE3)
A. niger	*Aspergillus niger*
AraR	L-arabinose-responsive transcription activator
BCA	bicinchoninic acid
BL21*	BL21 Star™(DE3)
BSA	bovine serum albumin
CAZymes	carbohydrate-active enzymes
CIAP	calf intestinal alkaline phosphatase
cDNA	complementary DNA
DhA	D-lyxo-heptulosaric acid
DMSO	dimethyl sulfoxide
dNTP	deoxyribonucleotide triphosphate
E. coli	*Escherichia coli*
FPKM	fragments per kilobase per million
FPLC	fast protein liquid chromatography
GO	gene ontology
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	high fidelity
IPTG	isopropyl β-D-1-thiogalactopyranoside
JGI	Joint Genome Institute
KDO	D-manno-octulosonic acid
Lar	L-arabinose reductase
LA	Luria Agar
LB	Luria Broth
Lxr	L-xylulose reductase
Tm	melting temperature
MM	minimal medium
NAD/NADH	nicotinamide adenine dinucleotide
NADP/NADPH	nicotinamide adenine dinucleotide phosphate
PBS	phosphate-buffered saline
PCP	pentose catabolic pathway
PCR	polymerase chain reaction
PFAM	protein families (database)
PPP	pentose phosphate pathway
SDR	short-chain dehydrogenase/reductase and related enzymes
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
TRIS-HCL	tris (hydroxymethyl) aminomethane (THAM) hydrochloride
Xdh	xylitol dehydrogenase
XlnR	(hemi-)cellulolytic regulator
Xyr	D-xylose reductase

Abstract

Fungal biotechnology has served as a means for the transition to a bio-based economy. Being decomposers in the ecosystem, filamentous fungi are especially considered as a key in the biological cycle in the concept of a circular economy. To make better use of the natural resources, research related to sugar catabolism in fungi has been expanded. This project follows previous research about the pentose catabolic pathway (PCP) in *Aspergillus niger*, which is responsible for the conversion of D-xylose and L-arabinose. To date, most of the enzymes involved in this pathway have been identified. However, the understanding of L-xylulose conversion is still incomplete. In this study, we have analyzed the transcriptome data, gene sequencing and the enzyme activity of the putative new L-xylulose reductase(s). Ten genes encoding enzymes belonging to Pfam 00106 of *A. niger* NRRL3 were chose based on their higher expression values on sugars related to PCP. Although LxrA and LxrB also belong to the PF00106 family, all of them, including ten potential genes, are in separate places on the phylogenetic tree. Among them, NRRL3_09880 showed a similar pattern in the transcriptome data as LxrA, and were closer to LxrB on the phylogenetic tree, which indicates its potential in this complex mechanism of the fungal metabolic pathway. Furthermore, six of the candidate genes, NRRL3_00094, NRRL3_07289, NRRL3_07921, NRRL3_09880, NRRL3_09881, NRRL3_11712, were successfully cloned into pET28 a (+). The recombinant plasmids were then transformed into *Escherichia coli* expression strains: Arctic Express and BL21 Star to proceed heterologous protein production. Based on the results of the small-scale production, proteins were produced in the soluble form. Moreover, the initial enzyme activity assays showed the possibility of NAD dependence in the reaction of xylitol conversion, instead of NADP, which is required in all known Lxr catalyzed reactions in other filamentous fungi. This finding can respond to the assumption of the existence of NADH-dependent L-xylulose reductases in previous research. However, in big-scale protein production, the amount and concentration of proteins were an obstacle to gaining further characterization of potential enzymes. To improve this problem, multiple adjustments are recommended in further experiments of heterologous expression. For example, a softer induction with lactose as the inducer and a mixed feed system with lactose and glycerol. By understanding much more about the unknown L-xylulose reductases, advanced applications can be made within the cell factory and assist the manufacture of bio-based products.

Table of Content

Chapter 1 – Introduction

1.1 The mission of fungal biotechnology in the circular economy

In the paradigm shift from a fossil-based economy to a bio-based economy, petroleum is gradually replaced with all kinds of material produced by living organisms. Among them, fungi play an essential role in making influences.

Fungal research has offered a set of solutions for various aspects of the United Nation's sustainable development goals (Meyer et al., 2020). In recent decades, fungi have been used to turn the by-products and waste from forestry and agriculture into composite material and bioenergy. To make a further step from its original function as primary decomposers in the ecosystem, scientists envision fungi to serve as other roles in the bio-based ecosystem. First, fungi can be used as a kind of material directly. To illustrate, by knowing the fungal life cycle and fruiting body formation much more, we can make fungi be the objects that directly offer functions with its body, such as being as the building components for construction or an analogue for animal leather (Meyer et al., 2020). Second, fungi can be the producer. To fulfill its original role as a decomposer in the ecosystem, the fungus can produce a diverse array of metabolites and enzymes. However, on another scale of the system, it can also be a producer with these functions, which is the concept of microbial cell factories (Fisher et al., 2014). In other words, the various kinds of biomolecules that fungal cells produce can be harvested and used to replace the classic petroleum-derived or petroleum-related products. For example, organic acid itaconate generated from biomass with the help of *Aspergillus terreus* can be used in polymer manufacturing (Teleky & Vodnar, 2019).

With the diverse production by fungi, most of the aspects of human need have been taken care of, such as food, medicine, feed, and basic chemicals. To boost the transformation fast and to join forces, countries around the world have started to support innovations in bio-utilization. For example, the Bio-Based Innovations program (https://www.bbi.europa.eu) sponsored by the European Commission is one of the main driving forces in this trend and aggregate hundreds of these organizations. Furthermore, to achieve the goal of resource efficiency, an engineering mindset must be applied in this transition.

Therefore, detailed research that investigates the genetic make-up of the fungus is essential no matter in which ways the fungus contributes. Metabolic engineering makes it a vivid way to optimize the consumption of the input and the emission of the output in the cell factories. By this, many more values can be created with the use of bioresources, biofuels, and biomass.

1.2 Plant biomass degradation and composition

80% of the sum of the biomass on Earth is the plant, which is around 450 gigatons of carbon (Bar-On et al., 2018). As for the common categories of biomass, land plants dominate the primary composition of plant biomass (Bar-On et al., 2018). Besides terrestrial plants, aquatic plants, such as water hyacinth and algae, are also a promising source in the future (Gaurav et al., 2020). Nowadays, the groups of plant biomass that are used mainly are crops and agricultural residues. Plant biomass is a complex and heterogeneous material, primarily composed of cellulose, hemicellulose, and lignin **(Figure 1)**. Among them, cellulose is a long-chain polymer of glucose molecules, and hemicellulose is constituted by several pentoses and hexoses, including D-xylose, L-arabinose, D-mannose, and D-galactose **(Table 1)**.

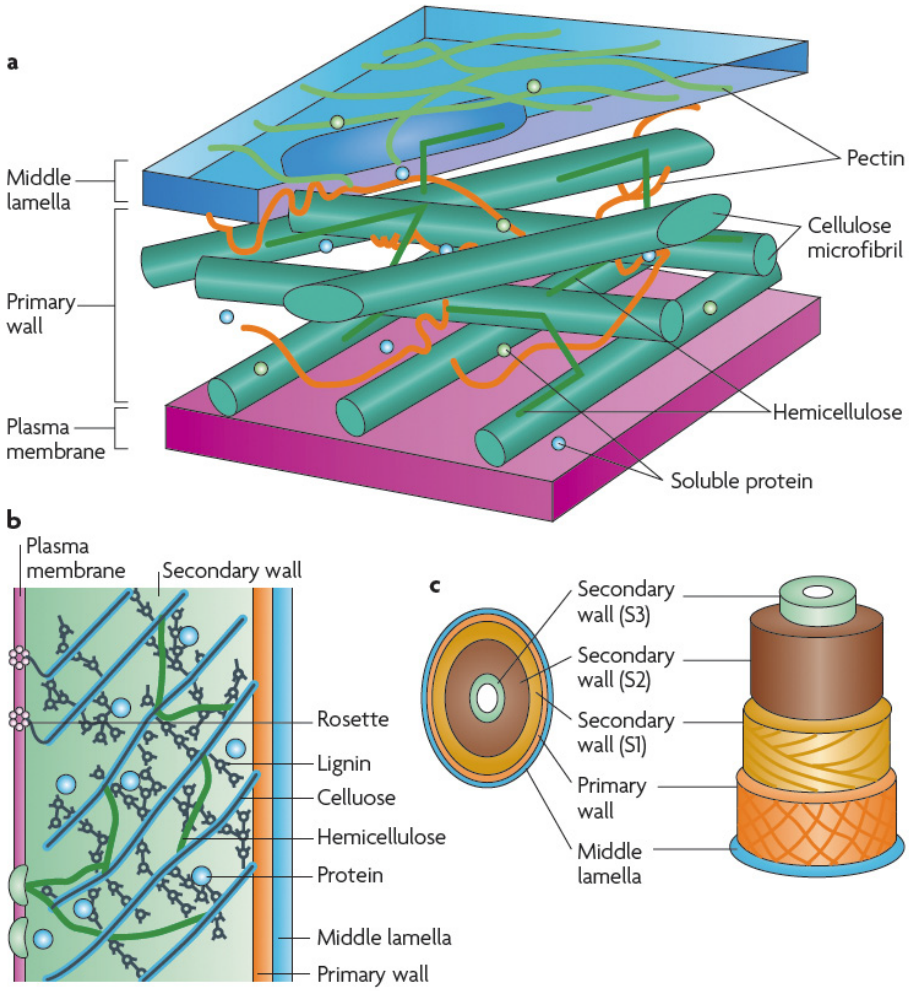


Figure 1. Schematic representation of the plant cell wall. a: Section view and structure of the primary cell wall. b: Section view and structure of secondary cell walls. c: Layers of plant cell walls (Klaubauf, 2015).

Table 1. Composition of plant polysaccharides. (Khosravi, 2017)

Polymer type	Polymer	Monomers
Cellulose	D-glucose	
Hemicellulose	Xylan	D-xylose
	Glucuronoxylan	D-glucuronic acid, D-xylose
	Arabinoglucuronoxylan	D-xylose, L-arabinose
	Arabinoxylan	D-xylose, L-arabinose
	Galacto(gluco)mannan	D-glucose, D-mannose, D-galactose
	Mannan/galactomannan	D-mannose, D-galactose
	Glucuronomannan	D-mannose, D-glucuronic acid, D-galactose, L-arabinose
	Xyloglucan	D-glucose, D-xylose, D-fructose, D-galactose
	Glucan	D-glucose
	Arabinogalactan	D-galactose, L-arabinose, D-glucuronic acid
Pectin	Homogalacturonan	D-galacturonic acid
	Xylogalacturonan	D-galacturonic acid, D-xylose
	Rhamnogalacturonan I	D-galacturonic acid, L-rhamnose, D-galactose, L-arabinose, ferulic acid, D-glucuronic-acid
	Rhamnogalacturonan II	D-galacturonic acid, L-rhamnose, D-galactose, L-arabinose, L-fucose, D-glucose, D-manno-octulosonic acid (KDO), D-lyxo-heptulosaric acid (DhA), D-xylose, D-apiose, L-acetic acid
Inulin	D-fructose, D-glucose	
Starch	Amylose	D-glucose
Amylopectin	D-glucose	
Lignin	Monolignols: p-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol	

These monomeric sugars have fixed carbon dioxide with photosynthesis and store chemical energy. As a result, carbon atoms and the energy here are the fundamental characteristics of the earth's carbon cycle **(Figure 2)**. At the same time, they provide the main functions that humans need for biomass, such as biofuels and biomaterials.

From converting the structure of plants to the carbon source of microbes, enzymes play a crucial role in the degradation of plant biomass. In this lignin-carbohydrate complex, cellulose makes up a substantial portion of the plant cell walls, offering a rigid, insoluble, and crystalline structure (Houfani et al., 2020). The recalcitrance property of lignocellulose makes the substances inside hard to be released. However, microbes, being the role of the decomposer in the ecosystem, can produce enzymes to obtain the nutrients they need from the degradation process of these carbohydrates. The enzymes that they encode to cleave the sugar linkages are called carbohydrate-active enzymes (CAZymes). With CAZymes and the other enzyme system in various pathways, microbes can efficiently make use of plant biomass (Boutard et al., 2014).

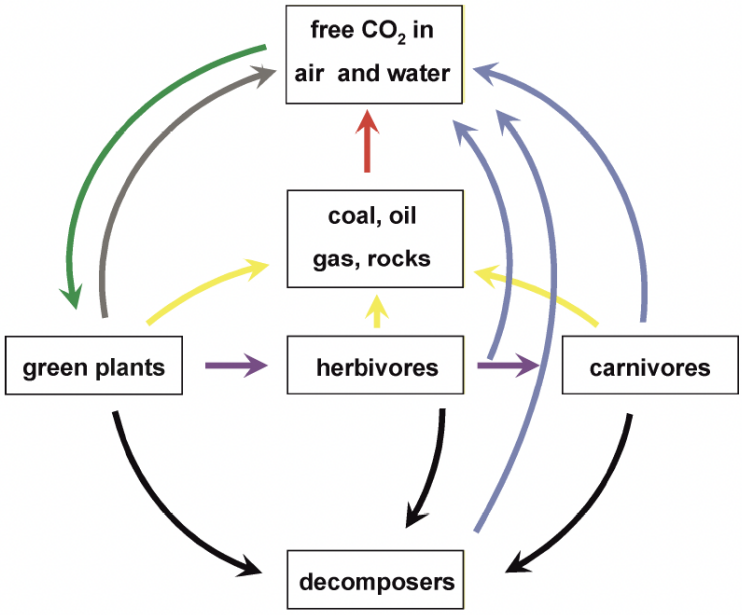


Figure 2. The earth's carbon cycle. Black arrows indicate the decay process; purple arrows indicate consumption; red arrow indicates combustion; yellow arrows indicate carbonification; blue arrows indicate respiration, and green arrow indicates photosynthesis (Battaglia, 2011a).

1.3 The carbon metabolism pathways in Aspergillus niger NRRL3

1.3.1 *Aspergillus niger* and the pentose catabolic pathway

Among microorganisms, filamentous fungi are considered efficient decomposers in nature, and the prominent helpers for degrading plant biomass in bioresource engineering. Being as one of the high-performance enzymes producing cell factories, *Aspergillus niger* has the potential to generate valuable organic acids, such as citric acid, itaconate, and galactarate, which can replace specific parts in current manufacturing process of the fossil fuel (Meyer et al., 2020).

The pentoses L-arabinose and D-xylose are the main constituents of hemicellulose xylan and xyloglucan except for glucose. They can be metabolized through the primary sugar catabolism of filamentous fungi. Pentose catabolic pathway (PCP) is the main pathway for D-xylose and L-arabinose conversion (Witteveen et al., 1989). In the PCP of *Aspergillus niger*, L-arabinose is reduced by L-arabinose reductase and then catalyzed by L-arabitol dehydrogenase to L-xylulose. Later, L-xylulose can be reduced to xylitol by L-xylulose reductase, while D-xylose can be converted to xylitol by D-xylose reductase. Xylitol is then catalyzed by xylitol dehydrogenase and converted by D-xylulose kinase to D-xylulose-5-phosphate, which enters the pentose phosphate pathway (PPP) for further processes (Horecker, 2002). To control this pathway, some essential regulators are necessary. XlnR regulates the first step of D-xylose conversion, AraR controls all the genes involved in the L-arabinose catabolism, and both co-regulate the steps from xylitol to D-xylulose-5-phosphate in *A. niger* (Battaglia et al., 2011b; Hasper et al., 2004). Since AraR is the specific L-arabinose regulator that controls the pathway and related degrading enzymes, the deletion of AraR would bring loss that XlnR is not able to compensate. Conversely, XlnR is involved in PCP of *A. niger* along with AraR, which means that the responses of related genes would not be significantly affected in the XlnR deletion. (Battaglia et al., 2011b; van Peij et al., 1998)

1.3.2 Unknown L-xylulose reductases

Due to the difficulty in cloning fungal L-xylulose reductases and the complexity in the metabolic pathway, it was the last step in PCP to find out all related enzymes (Chroumpi et al., 2021; Seiboth & Metz, 2011).

More than one enzyme has been described in each step of the pentose catabolic pathway (Chroumpi et al., 2021) For example, the reductases LarA, XyrA, and XyrB (Terebieniec et al., 2021). Likewise, multiple enzymes are involved in the catabolic way of L-arabinose. In previous studies, specific L-xylulose reductase genes were deleted to confirm the participation of additional genes, and the residual growth in L-arabitol shows the possibility of the existence of new L-xylulose reductase(s) (Chroumpi et al., 2021).

1.4 Aim and outline of this project

The aim of this project is to identify other L-xylulose reductases encoding genes in *A. niger*, and to biochemically characterizing these enzymes encoded by these genes. To tackle these unknowns, a workflow was constructed as shown in the following: First, candidate genes were selected based on the transcriptome and phylogeny analysis. Second, traditional cloning and sequencing check was conducted to acquire the recombinant plasmids carrying desired genes. Last, these putative enzymes were produced in the expression stains and purified for further tests. By this, more knowledge could supply potential for new fungal cell factories and evaluate the potential of these enzymes in the pathways.

Chapter 2 - Materials and Methods

2.1 Transcriptome analysis

In this study Aspergillus niger NRRL3 (ATCC 9029, CBS 120.49) was used as model strain and related gene list was retrieved from the database MycoCosm – the Fungal Genomics Resource (https://mycocosm.jgi.doe.gov/mycocosm/home) of Joint Genome Institute (JGI). Based on the Pfam database (http://pfam.xfam.org), PF00106 protein family was selected because it is classified as short chain dehydrogenase family.

The expression values, obtained from previously published transcriptome data (Aguilar-Pontes et al., 2018), were represented by Fragments Per Kilobase per Million (FPKM). They show the average expression value of genes under various substrate conditions. This transcriptome dataset was created by cultivating three *A.niger* strains, wild type (N402), AraR (dAraR) and XlnR (dXlnR) deletion strains on minimal medium for aspergillus (MM) agar plates, which contained 25mM monomeric sugars as carbon source for 5 days. The wildtype was cultured on monomeric sugars included L-arabinose (ara), D-xylose (xyl), D-galacturonic acid (galA), L-rhamnose (rha), D-galactose (gal), D-mannose (man), D-glucose (glc), and D-fructose (frc). The AraR and XlnR deletion strains were examined with the substrate L-arabinose and D-xylose, respectively. The average expression value was calculated from the duplicate samples in original data and were categorized into seven groups: less than 20, 20-40, 40-100, 100-250, 250-500, 500-1000, over 1000. Genes that have values more than 500 on L-arabinose or D-xylose and values less than 300 on D-glucose were picked. Gene that show relatively high expression on the ratio of L-arabinose or D-xylose to D-glucose was also chosen. The phylogenetic tree was built by using MEGA6 software, with parameters of Maximum Likelihood algorithms by using 500 bootstraps of the short chain dehydrogenase group (Pfam 00106) sequences.

2.2 Conventional cloning

Ten gene candidates of L-xylulose reductases were cloned into the vector pET28a and transformed into competent cells *Escherichia coli* DH5α strain, respectively. The putative genes were amplified from

soybean hull, xylose, and arabinose cDNA. Primers containing the specific restriction sites were designed (Integrated DNA Technologies) and used in a polymerase chain reaction (PCR) to amplify the genes of interest. Several restriction recognition sites, including NdeI, BamHI, EcoRI, SacI, HindIII, NotI, and XhoI, were examined in the gene sequences, which were retrieved from the JGI fungal genome database. The Restriction sites, BamHI and NotI, were selected because both did not exist in the sequence of the candidate genes **(Table 10)**. The primer melting temperature (Tm) are all around 73 °C to 77 °C **(Table 11)**, so the annealing temperature of PCR is set at 72 °C. Besides, the difference of the forward and reverse primer melting temperatures are less than 1 °C. The PCR reaction contained 15 µl of water, 5 µl 5x high fidelity (HF) Buffer, 1 µl 10 mM dNTP mix (Promega), 0.5 µl of each specific primers pair, 2 µl of DNA, 0.75 µl dimethyl sulfoxide (DMSO) and 0.25 µl Phusion Polymerase (ThermoFisherScientific). The reaction was performed in a PCR machine as follows: Initial denaturation 98 °C for 30 seconds; denaturation at 98 °C for 10 seconds, annealing at 70 °C for 30 seconds and elongation at 72 °C for 5 minutes, repeated for 30 cycles; extension at 72 °C for 5 minutes. The pET-28a (+) plasmid (Genscript Biotech, Leiden, the Netherlands), which carried the N-terminal histidine tag for detection and purification of the recombinant protein, was isolated and purified using DNA Miniprep and PCR-Purification Kits (ThermoFisherScientific) according to the manufacturer's instruction. Afterwards, the restriction digestion was done with the fast digestion enzymes BamHI and NotI (ThermoFisherScientific) to make the cleavage sites for directional cloning.

Table 2. Contents and conditions of PCR. In the reaction mix, cDNA was used for the first gene amplification and the reamplification was performed for further PCR reactions.

Reaction mix	**Amount (1x) (µl)**
H_2O	15
5x HF buffer	5
dNTPs	1
Primer forward	0.5
Primer reverse	0.5
cDNA/ PCR product	2
DMSO	0.75
Phusion Polymerase	0.25

PCR conditions		
98 °C	00:30	
98 °C	00:10	30 circles
70 °C	00:30	
72 °C	01:30	
72 °C	05:00	
10 °C	forever	

The linearized plasmids were treated by calf intestinal alkaline phosphatase (CIAP) to catalyze the dephosphorylation of the 5'and 3'end of the DNA. Then the digested genes about 120 ng were ligated with 100 ng of the digested and treated plasmids, 1 µl of 10 x ligation buffer, and 1 µl of ligase. The reaction solution was added up to 10 µl by water and then performed at 12 °C overnight in the PCR machine. After ligation, the recombinant pET-28a (+) plasmids were then transformed into the *E. coli* strain DH5α for further propagation. For each gene, 2 or 6 µl of ligation product was mixed with 50 µl of competent cells of DH5α. The tubes were incubated on ice for ten minutes, then in a 42 °C water bath for 45-60 seconds and cooled on ice for three minutes. Next, 500 µl of LB was added and the tubes were incubated for one hour at 37 °C with 200 rpm of shaking. Thereafter, the tubes were short spin down, and the supernatant was discarded. The obtained pellets were then resuspended and streaked on plates. *E. coli* DH5α was screened on the LA plates containing 50 µg/ml kanamycin as the selection marker. Grown colonies were verified by colony PCR. PCR reaction contains 5 µl 5x GoTaq® Flexi Green Buffer (Promega), 1.3 µl 10 mM dNTP mix (Promega), 1 µl of T7 promoter and terminator, 2.5 µl $MgCl_2$, 0.2 µl GoTaq® Flexi DNA Polymerase (Promega), 14 µl water, and the colony template, which is a pick of colony taken by a pipette tip. The reaction was performed in a PCR machine (Labcycler Basic, SensoQuest) as follows: initial denaturation 95 °C for 2 minutes; denaturation at 95 °C for 30 seconds, annealing at 58 °C for 30 seconds and elongation at 72 °C for 30 seconds, repeated for 30 cycles; final extension at 72 °C for 5 minutes. Then the results of the cloning process were confirmed by sequencing (Macrogen Europe B.V.) and positive clones were selected. The recombinant plasmids were then isolated and transformed as the method above into *E. coli* production strains: Arctic Express DE3 (Agilent Technologies) and BL21 Star (Invitrogen). The transformants were grown on LA plates supplemented with different antibiotics to select for the recombinant plasmid. For the growth requirement of transformants, Arctic Express DE3

needs 25 µg/ml of kanamycin and 25 µg/ml of gentamycin, and BL21 Star needs 50 µg/ml of kanamycin. After overnight cultivation, grown colonies were verified by colony PCR.

Table 3. Contents and conditions of colony PCR.

Reaction mix	Amount (1x) (µl)
5x Green Buffer	5
dNTPs	1.3
T7 promoter	1
T7 terminator	1
$MgCl_2$	2.5
DNA Polymerase	0.2
H_2O	14
Colony template	-

PCR conditions		
95 °C	02:00	
95 °C	00:30	30 circles
58 °C	00:30	
72 °C	00:30	
72 °C	05:00	
10 °C	forever	

2.3 Small-scale protein production and purification

E. coli strain Arctic Express DE3 and BL21 Star cells containing the recombinant plasmid were grown to an OD_{600} of 0.8 in LB medium with antibiotics stated above at 37 °C and shaking at 200 rpm. During the further incubation overnight at 12 °C and 200 rpm, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added in the culture to the final concentration of 0.1 mM, as IPTG can induce the expression of protein by mimicking allolactose to remove the repressor from the lac operon. To compare the effect of inducer, lactose was used to replace IPTG in one of the small-scale production experiments. After overnight induction, the cells were centrifuged at 4000 x g for 15 min. The harvested cell pellet was resuspended in 1.5 mL of BugBuster Protein Extraction Reagent with 1 µL Benzonase Nuclease (Merc Millipore, Darmstadt, Germany). To receive the purified protein, Benzonase Nuclease was used to reduce the

viscosity and the contamination by eliminating nucleic acid in cell lysis step. After the rotating incubation at 4 °C for 30 minutes, the cells were spined down at 11,000 x g for 20 minutes at 4 °C. Both supernatant and pellet were kept for the verification of solubility of desired protein by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE). The supernatant (crude extract) was filtered through 45 μm syringe filters (Whatman, GE Healthcare Life Sciences, Pittsburgh, PA, USA) and used in initial enzyme activity analysis.

2.4 Large-scale protein production and purification

Based on the results of small-scale protein production, 00094, 09880, 09881, and 07289 were selected to be produced in expression strain BL21 star. 8 mL of the pre-culture was inoculated in 400 mL of LB supplemented with 50 μg/mL of kanamycin. After incubated to OD_{600} of 0.8, 400 μL of 100 mM IPTG was added to proceed with induction overnight at 12 °C and 200 rpm. After overnight incubation, the culture was then transferred to 380 mL centrifuge bottles, centrifuging with the same weights at 8000 x g and 4 °C for 15 minutes, using High-speed centrifuge (Avanti J26S XP, Beckman Coulter) with the JA-10 rotor. The harvested cell pellet was resuspended in 30 mL of BugBuster Protein Extraction Reagent with 6 μL Benzonase Nuclease before the followed-up process as stated in *2.3 Small-scale protein production and purification.* Then the filtered supernatant was purified by the ÄKTA start system (GE Healthcare Life Sciences, Uppsala, Sweden) coupling with a HisTrap FF 1-ml column to perform the fast protein liquid chromatography (FPLC). The column was equilibrated with buffers- HisTrap Buffer A (20 mM HEPES, 20 mM Imidazole, 400 mM NaCl, pH 7.5) and HisTrap Buffer B (20 mM HEPES, 400 mM Imidazole, 400 mM NaCl, pH 7.5). The protein was eluted, and 1 mL fractions were collected by Frac360 fraction collector. Based on the absorbance peak shown in chromatograms, certain fractions were pooled and transferred to 10 kDa membrane tube (Sartorius AG) with the addition of 15 mL Buffer C (20 mM HEPES). After the solution in tube was centrifuged until 1 mL left, 15 m Buffer C was added again. By repeating this three times, the desalting was done and 1 mL of enzyme with 9mL of Buffer C was collected. The protein concentration was measured with BCA Protein Assay Kit (ThermoFisherScientific). Based on the colorimetric detection of the purple reaction product formed by the chelation of bicinchoninic acid (BCA) and a cuprous ion, protein concentrations within a range of 20-2000 μg/ml can be quantified by the linear relationship of absorbance at 562 nm. A series of nine dilutions of known concentrations were

prepared as Table 4. Three samples of unknown concentration, including one undiluted and two 10 x diluted samples, were assayed together in the microplate. 25 µL of standard or unknown sample was mixed with 200 µL of the Working Reagent, preparing by 50 parts of BCA Reagent A with 1 part of BCA Reagent B, in each well and the solution was incubated at 37°C for 30 minutes before measured. Then a standard curve **(Figure 3)** was introduced to calculate the unknown concentration of protein. The protein was then diluted to a final concentration of 0.2 mg/mL to ensure equal concentrations in further enzyme assays.

Table 4. Preparation of diluted albumin (BSA) standards according to the instructions of the manufacture. (Working Range = 20–2,000 µg/ml)

Vial	Volume of Diluent	Volume and Source of BSA	Final BSA Concentration
A	0	300 µl of Stock	2,000 µg/ml
B	125 µl	375 µl of Stock	1,500 µg/ml
C	325 µl	325 µl of Stock	1,000 µg/ml
D	175 µl	175 µl of vial B dilution	750 µg/ml
E	325 µl	325 µl of vial C dilution	500 µg/ml
F	325 µl	325 µl of vial E dilution	250 µg/ml
G	325 µl	325 µl of vial F dilution	125 µg/ml
H	400 µl	100 µl of vial G dilution	25 µg/ml
I	400 µl	0	0 µg/ml = Blank

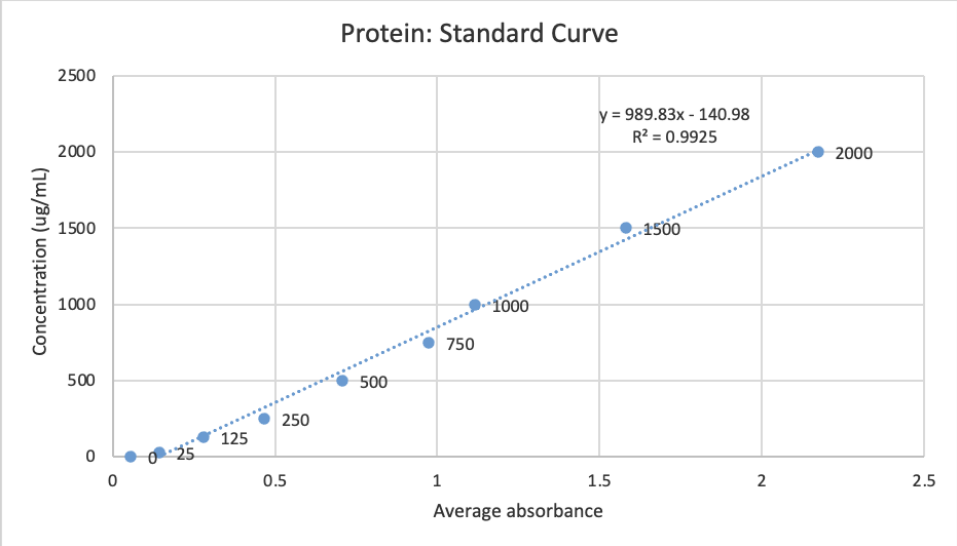


Figure 3. Stand curve of protein solutions of BCA assay. The concentration is calculated with the formula: y=989.83x-140.98, with an R^2 of 0.99.

2.5 Enzymatic assays

Initial enzyme activity was examined at 25 °C in the microplate reader (FLUOstar OPTIMA, BMG LABTECH, Ortenberg, Germany). A total of 1mL of crude extract from the small-scale production process was used for the enzyme assays. In each reaction, 60 µL of crude enzyme was tested in the conditions stated below. Reductase activities were measured in 50 mM PBS buffer, pH 7.0, with 0.2 mM NADPH cofactor and 100 mM substrate- L-arabinose or D-xylose. The reverse reactions, dehydrogenase activities, were measured in 100 mM TRIS-HCL, pH 8.5, with 1 mM NADP and 100 mM substrate- xylitol or arabitol. The consumption of a specific cofactor was evaluated by measuring the decrease in absorbance at 340 nm in flat-bottom microtiter plates (Grainer Bio-One, Kremsmünster, Austria) with the reaction mixture.

After the small-scale production, both reductase and dehydrogenase activity assays were conducted three times with a few adjustments on the cofactor and buffer. Both AE and BL21 star were used as the expression strain in activity assay I, and only BL21 star was used for protein production in the activity assay II and III. The condition in each test was shown in **Table 5** and **Table 6**. As for the large-scale production, seven substrates **(Table 7)** were used in the presence of cofactors NAD and NADP to test the dehydrogenase function, while eleven substrates were evaluated with the NADH and NADPH to check the reductase ability.

Table 5. Experimental condition of three dehydrogenase activity assays.

Test	***Substrate***	***Cofactor***	***Buffer***	***Strain***	***Figure***
Dehydrogenase Activities Assay I	Arabitol, Xylitol	NADP	pH 7 phosphate buffer or pH 9 TRIS-HCl	AE, BL21*	9
Dehydrogenase Activities Assay II	Arabitol, Xylitol	NAD	pH 7 phosphate buffer or pH 8.5 TRIS-HCl	BL21*	10
		NADP			11
Dehydrogenase Activities Assay III	Arabitol, Xylitol	NAD	pH 7 phosphate buffer or pH 8.5 TRIS-HCl	BL21*	12
		NADP			13

Table 6. Experimental condition of three reductase activity assays.

Test	***Substrate***	***Cofactor***	***Buffer***	***Strain***	***Figure***
Reductase Activities Assay I	L-arabinose, D-xylose	NADPH	pH 7 phosphate buffer	AE, BL21*	14
Reductase Activities Assay II	L-arabinose, D-xylose	NADH	pH 7 phosphate buffer	BL21*	15
Reductase Activities Assay III	L-arabinose, D-xylose	NADH	pH 7 phosphate buffer	BL21*	16
		NADPH			17

Table 7. Experimental condition of the dehydrogenase and reductase activity assays of 00094 in big-scale production.

Test	***Substrate***	***Cofactor***	***Buffer***	***Strain***	***Figure***
Dehydrogenase Activities Assay	Xylitol, L-arabitol, galactitol, sorbitol, glycerol, glyceraldehyde, L-rhamnose	NAD	pH 8.5 TRIS-HCl	BL21*	18-B
		NADP			18-C
Reductase Activities Assay	D-xylose, D-arabinose, L-xylose, L-arabinose, D-glucose, D-fructose, L-rhamnose, D-galactose, L-sorbose, D-ribose, D-mannose	NADH	pH 7 phosphate buffer	BL21*	19-B
		NADPH			19-C

Chapter 3 – Results

3.1 Gene selection of ten potential candidates

To select putative L-xylulose reductase encoding genes in *A. niger* NRRL3, the Pfam database was used to search in the PF00106 family (short-chain dehydrogenases/reductases family (SDR)) for potential candidates. According to previous research, the L-xylulose reductases in *Trichoderma reesei*, *Neurospora crassa*, *Ambrosiozyma monospora,* and the LxrA and LxrB in *Aspergillus niger* all belong to PF00106. Based on the fungal genomics resource, MycoCosm of JGI, there are 156 genes belonging to the PF00106 family in *Aspergillus niger* NRRL3. According to gene ontology (GO), all these genes have four GO terms, including the catalytic activity (GO:0003824), binding (GO:0005488), metabolic process (GO:0008152), and oxidoreductase activity (GO:0016491) (Attrill et al., 2019; "The Gene Ontology Resource: 20 Years and Still GOing Strong," 2019).

Based on previously published transcriptome data, the expression values of these 156 genes in *A. niger* NRRL3 were analyzed to genes of more potential candidates. By comparing the average values of expression in 11 conditions, 8 genes were selected out of 156 SDR genes. The candidate genes, NRRL3_00094, NRRL3_04005, and NRRL3_07921 have high expression values on D-xylose compared to other substrates, while NRRL3_00672, NRRL3_07451, NRRL3_09880, NRRL3_09881, and NRRL3_11712 have more than 500 or even 1000 FPKM values on both L-arabinose and D-xylose. Besides these eight genes, NRRL3_07289 and NRRL3_06236 were also added based on the results of our previous experiments (**Table 8, 9**).

Table 8. Transcriptome data of selected genes and the known L-xylulose reductase. All monomeric sugars are at 25 mM starting concentration and harvested after 2h. Substrates: ara= L-arabinose, xyl= D-xylose, galA= D-galacturonic acid, rha= L-rhamnose, gal= D-galactose, man= D-mannose, glc= D-glucose, frc= D-fructose. Strains: wild type (N402) unless indicated, dAraR= AraR deletion, dXlnR= XlnR deletion. NRRL3_04510= *lxrB*, NRRL3_10884= *lxrA*. FPKM values: less than 20, 20-40, 40-100, 100-250, 250-500, 500-1000, over 1000.

protein id	ara/glc ratio	xyl/glc ratio	ara	dAraR, ara	xyl	dXlnR, xyl	galA	rha	rha/gal	gal	man	glc	frc
NRRL3_00094	2.398	6.596	234.4	803.7	644.9	113.6	447.3	519.2	523.7	203.1	128.0	97.8	102.0
NRRL3_00672	4.864	4.520	540.4	2241.6	502.3	155.4	573.0	449.1	292.0	282.0	137.9	111.1	165.9
NRRL3_04005	0.666	3.702	189.4	242.3	1052.9	468.4	253.5	601.3	55.5	900.4	703.4	284.4	88.1
NRRL3_07451	16.645	15.810	1674.5	58.6	1590.6	539.3	112.2	134.3	149.8	266.1	164.3	100.6	123.6
NRRL3_07921	5.600	86.888	18.7	8.1	290.2	51.5	7.0	13.0	29.5	41.4	13.0	3.3	5.5
NRRL3_09880	11.263	7.398	2103.5	231.7	1381.7	1703.3	508.5	736.7	1454.5	47.4	191.6	186.8	204.2
NRRL3_09881	11.556	11.699	952.7	188.2	964.5	698.6	327.6	559.8	844.6	35.4	88.5	82.4	112.6
NRRL3_11712	23.556	42.154	828.4	50.3	1482.4	375.6	384.6	102.2	157.5	25.3	53.8	35.2	39.6
NRRL3_07289	1.661	1.286	28.9	18.4	22.4	16.3	15.9	10.1	25.1	2413.6	10.1	17.4	22.8
NRRL3_06236	2.086	0.782	2.0	0.8	0.8	1.7	0.8	0.8	1.9	1.0	2.0	1.0	1.2
NRRL3_04510	1.035	1.348	2325.5	1198.0	3028.7	4819.7	364.1	223.6	353.3	275.4	2339.6	2246.0	2327.8
NRRL3_10884	62.039	25.368	4234.5	69.3	1731.6	3640.0	532.1	39.1	240.7	47.0	98.7	68.3	54.4

Table 9. Gene description and annotation of selected genes and the known L-xylulose reductase. NRRL3_04510= *lxrB*, NRRL3_10884= *lxrA*.

PROTEIN ID	GENE SIZE	INTRONS	PFAM
NRRL3_00094	1029	3	PF00106
NRRL3_00672	3053	6	PF01575; PF00106
NRRL3_04005	801	0	PF00106
NRRL3_07451	786	0	PF00106
NRRL3_07921	753	0	PF00106
NRRL3_09880	1052	4	PF00106
NRRL3_09881	1112	3	PF00106
NRRL3_11712	1106	2	PF00106
NRRL3_07289	829	1	PF00106
NRRL3_06236	995	3	PF00106
NRRL3_04510	1002	2	PF00106
NRRL3_10884	972	1	PF00106

Besides the expression analysis, the phylogenetic tree was used to confirm the relationship of genes. In the phylogenetic tree of PF00106 family **(Supplementary Figure 1)**, LxrA (NRRL3_10884), LxrB (NRRL3_04510), previously putatively identified LxrC (NRRL3_11471), and 10 selected genes were marked with colors. In general, most of them are scattered in different clads. Two of them (NRRL3_04005, NRRL3_09880) are relatively closed to LxrB.

Based on the phylogenetic tree, there were no additional criteria to filter out genes, because of the distance between LxrA, LxrB, and previously identified LxrC is also far. As a result, after the expression and phylogeny, these 10 candidates were chosen for further analysis.

3.2 Recombinant plasmids obtained by conventional cloning

To acquire specific genes that were selected, primers were designed to proceed with gene cloning. All these ten genes do not contain BamHI and NotI sequence **(Table 10)**. As a result, oligonucleotides with restriction sites were designed **(Table 11)** and sticky ends would be formed.

Table 10. Restriction sites check and additional information of selected genes.

No.	Protein ID	Intron	Gene size	M.W. (kDa)	NdeI	BamHI	EcoRI	SacI	HindIII	NotI	XhoI	Putative activity
1	00094	3	1029	40.14	No	No	No	No	No	No	Yes	Short chain dehydrogenase
2	00672	6	3053	106.20	Yes	No	Yes	No	Yes	No	Yes	Short chain dehydrogenase MaoC like domain
3	04005	0	801	37.42	Yes	No	No	No	No	No	No	Short chain dehydrogenase
4	07451	0	786	37.48	No	No	No	No	Yes	No	No	Short chain dehydrogenase
5	07921	0	753	36.24	No	No	No	No	No	No	No	Short chain dehydrogenase
6	09880	4	1052	38.63	No	No	No	No	No	No	No	Short chain dehydrogenase
7	09881	3	1112	41.07	Yes	No	No	Yes	No	No	Yes	Short chain dehydrogenase
8	11712	2	1106	45.18	No	No	No	No	No	No	Yes	Short chain dehydrogenase
9	07289	1	829	36.28	No	No	No	No	No	No	No	Short chain dehydrogenase
10	06236	3	995	37.31	No	No	No	No	No	No	Yes	Short chain dehydrogenase

Table 11. Primers used in the study. Tm – melting temperature, recognition sites for restriction enzymes are indicated with red.

Primers	Sequence (5' → 3')	Tm (°C)
BamHI-00094fw	CGGAATTCCGGATCCATGTCGCCAACTTCTACAGATACTGCCTCCGC	74.7
NotI-stop-00094rv	CGGAATTCGCGGCCGCTCACCAAGCACCGACGATGCCTGCAAT	74.9
BamHI-00672fw	CGGAATTCCGGATCCATGTCTGAGCTGCGCTTCGACAACCAGACCGTG	77.6
NotI-stop-00672rv	CGGAATTCGCGGCCGCTCACAACTTGGCCTTGGCACCCTCCAACAGC	77.6
BamHI-04005fw	CGGAATTCCGGATCCATGCCTATCTCCATTCCCTCCGCCTCCAGC	76.4
NotI-stop-04005rv	CGGAATTCGCGGCCGCTCACCGCACGGTGTATCCTCCGTCAATGACC	76.2
BamHI-07451fw	CGGAATTCCGGATCCATGGCTCGCATTCTCATTACTGGCTCTGTC	72.4
NotI-stop-07451rv	CGGAATTCGCGGCCGCTCAAGCCGACAGTCTCAATCCAGTAACCTCTTC	72.8
BamHI-07921fw	CGGAATTCCGGATCCATGACTGTCATGAAACCATTAATCCTAATCACCGGTGC	73.1
NotI-stop-07921rv	CGGAATTCGCGGCCGCTCACCACCCAAACTCCACAAAATACCCCTCATC	73.6
BamHI-09880fw	CGGAATTCCGGATCCATGTCTATGAACGTAGCCGAGTGTGCCCC	73.6
NotI-stop-09880rv	CGGAATTCGCGGCCGCTCATGGTAGCACGTATCCTCCGTCAATGAGGAC	73.9
BamHI-09881fw	CGGAATTCCGGATCCATGGGTTCCAACTCAGAAACCCCCGCAATTCCT	76.5
NotI-stop-09881rv	CGGAATTCGCGGCCGCTCAGAAAACCAGGCCTCCACCATCGACCACAAT	76.1
BamHI-11712fw	CGGAATTCCGGATCCATGATCGCCCCTACTGAGCCCGTTCTCTCC	76.0
NotI-stop-11712rv	CGGAATTCGCGGCCGCTCATGCAGCCCAGGCATCTTTCAACACCTTCTG	75.5
BamHI-07289fw	CGGAATTCCGGATCCATGTCCCTCAAAGGTAAAGTCGCCATTGTCACT	73.3
NotI-stop-07289rv	CGGAATTCGCGGCCGCTCAGATATACAACATCCCACCATTAGCACATATCACATCGC	73.4
BamHI-06236fw	CGGAATTCCGGATCCATGCATTCCCCAACCTCTACTACAACCCAGC	73.7
NotI-stop-06236rv	CGGAATTCGCGGCCGCTCACAACAATAGCATTCCACCTGTACCACTAAGAGTACA	73.0

Ten pairs of designed primers were used to amplify the designated target genes. The efficacy of the PCR was checked by agarose electrophoresis **(Figure 4)**. Genes 04005 and 07451 were amplified from arabinose cDNA while 00094, 09880, 09881, and 11712 were from xylose cDNA. The other four genes, 672, 6236, 7921, and 7289, were amplified from soybean hull cDNA. DNA products were visible at the expected size of around 1000 bp. Among them, the amplification of 11712, 7451, and 672 was not as efficient. On the other side, pET28 a (+) plasmids were isolated from *E. coli* and purified **(Figure 5)**. The undigested DNA (lane 2-5) showed three bands, which represent three possible plasmid conformations. The upper two bands were relaxed and nicked DNA while the bottom ones were supercoiled DNA that ran smaller than the predicted sizes of around 5.3 kb **(Supplementary Figure 2)**. The last two lanes showed two digested and CIAP treated plasmids with an approximate concentration of 200 ng/µL, according to the Lambda DNA as a concentration marker. After the restriction digestion and DNA purification, the target gene, and the plasmid pET28a were ligated. The ligation mixture was then transformed into *E. coli* strain DH5α. After overnight incubation, the transformation results were verified by colony PCR, and the positive colonies were picked to isolate the recombinant plasmids for sequencing. Based on the alignment check of the sequencing results and the corresponding expected sequences, six of the genes, including 00094, 07289, 07921, 09880, 09881, and 11712, were cloned successfully **(Table 12)**.

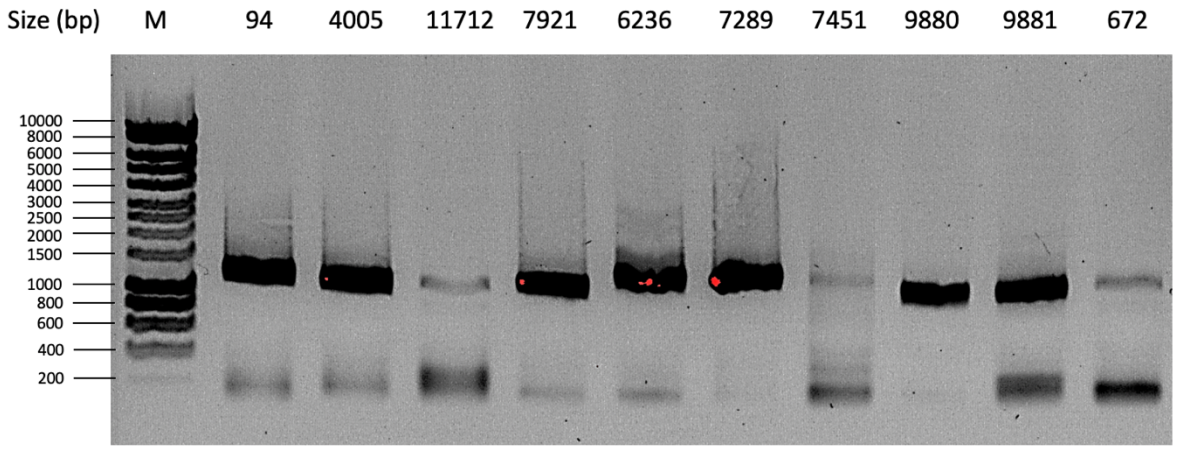


Figure 4. Genes amplification from cDNA (Purified and used for digestion). 04005 and 07451: from arabinose cDNA; 00094, 09880, 09881, and 11712: from xylose cDNA. 672, 6236, 7921, and 7289: from soybean hull cDNA. Expected DNA products at 1000 bp.

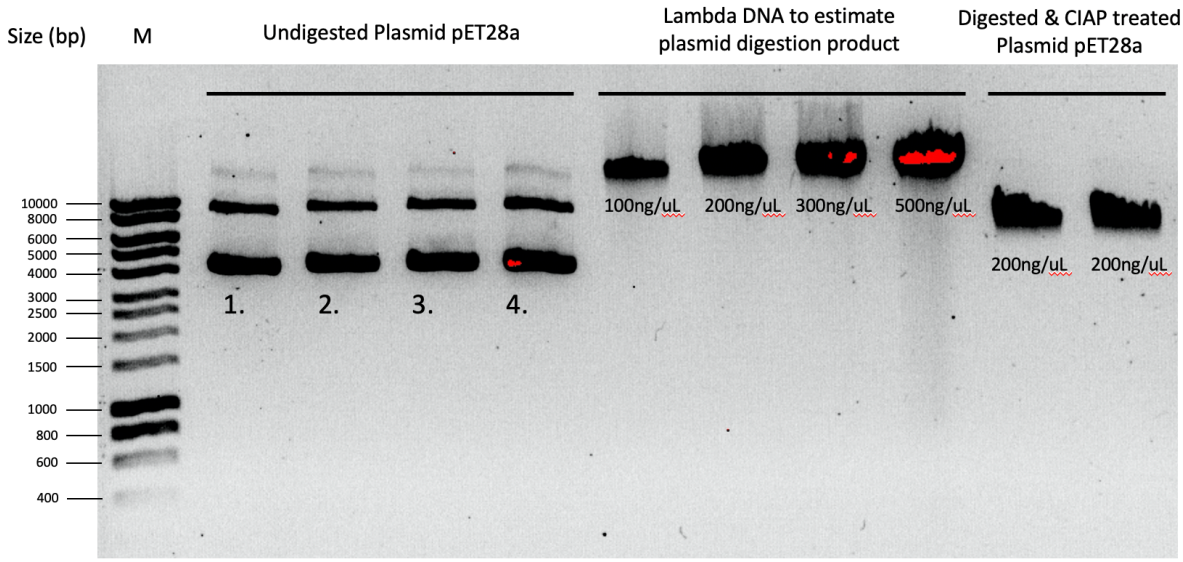


Figure 5. The estimated concentration of pET28a plasmid after digestion and CIAP treatment.

Table 12. Result of sequencing check for selected genes.

GENE ID	NUMBER OF PLASMIDS SENT FOR SEQUENCING	TRANSFORMATION RESULT
00094	2	Successful
00672	4	Failed
04005	1	Failed
06236	1	Failed
07289	8	Successful
07451	1	Failed
07921	3	Successful
09880	3	Successful
09881	2	Successful
11712	3	Successful

3.3 Small-scale protein production and initial activity assay

Recombinant plasmids carrying genes: NRRL3_00094, NRRL3_07289, NRRL3_07921, NRRL3_09880, NRRL3_09881, NRRL3_11712 were used for *E. coli* Arctic Express and BL21 Star transformation. The small-scale production proceeded as it was stated in Materials and Methods *2.3 Small-scale protein production and purification*, activity assay I, II, and III were conducted. **Figure 6** shows that all proteins in assay I with the expected sizes ranged from 36 to 40 kDa, were produced in the soluble form.

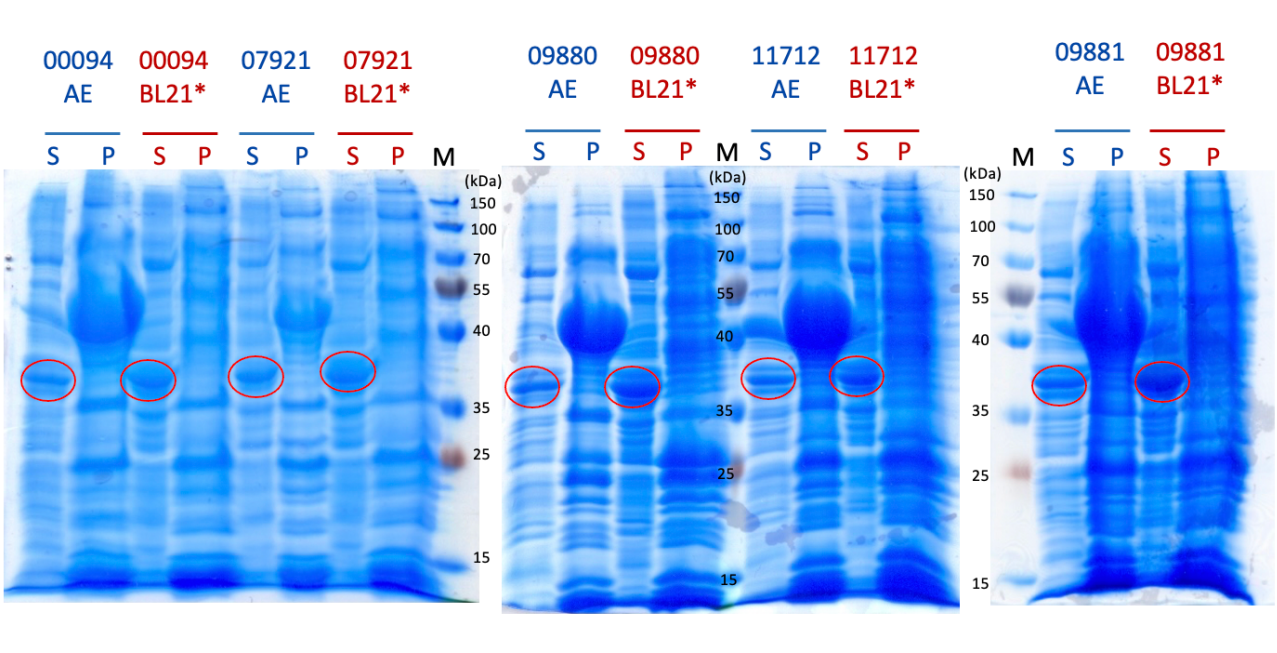


Figure 6. SDS-PAGE analysis of the location of produced proteins in activity assay I.

S: supernatant, P: pellet, M: marker. Red circles stand for the bands of target proteins.

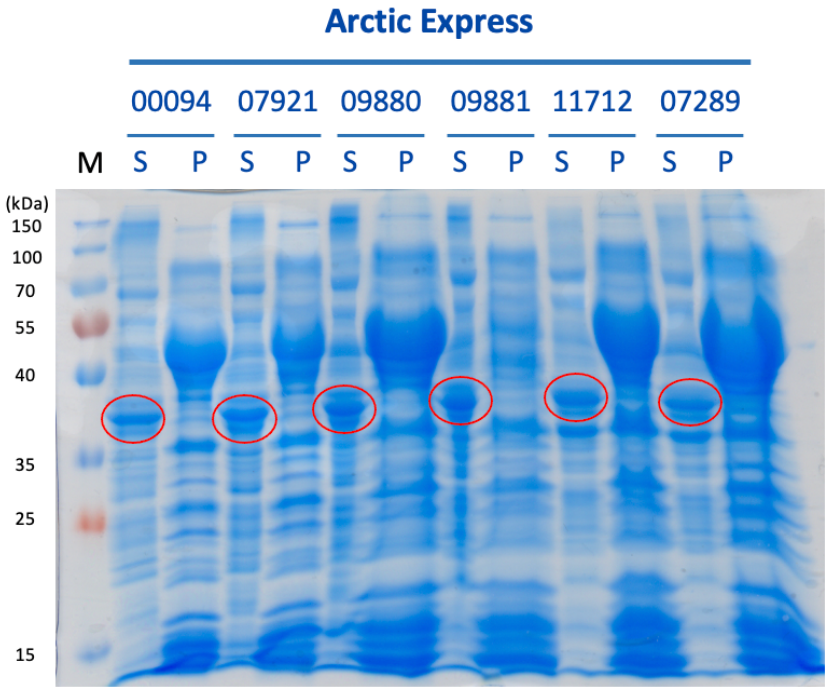


Figure 7. SDS-PAGE analysis of the location of produced proteins in activity assay II.

S: supernatant, P: pellet, M: marker. Red circles stand for the bands of target proteins.

Another test about inducing agents was also conducted. Both lactose and IPTG were used to induce protein expression in two expression strains. All the samples induced by lactose did not show activity on xylitol. Besides the activity test, 07289 and 09881 induced by lactose were compared by SDS-PAGE together with other IPTG-induced proteins, as shown in Figure 8. Protein 07289 was produced more in BL21 Star than in Arctic Express There was no difference for 09881 induced by lactose between two strains, while the comparison of 09881 induced by two agents was obvious, which showed that the protein production induced by IPTG in BL21 Star was higher **(Figure 8)**.

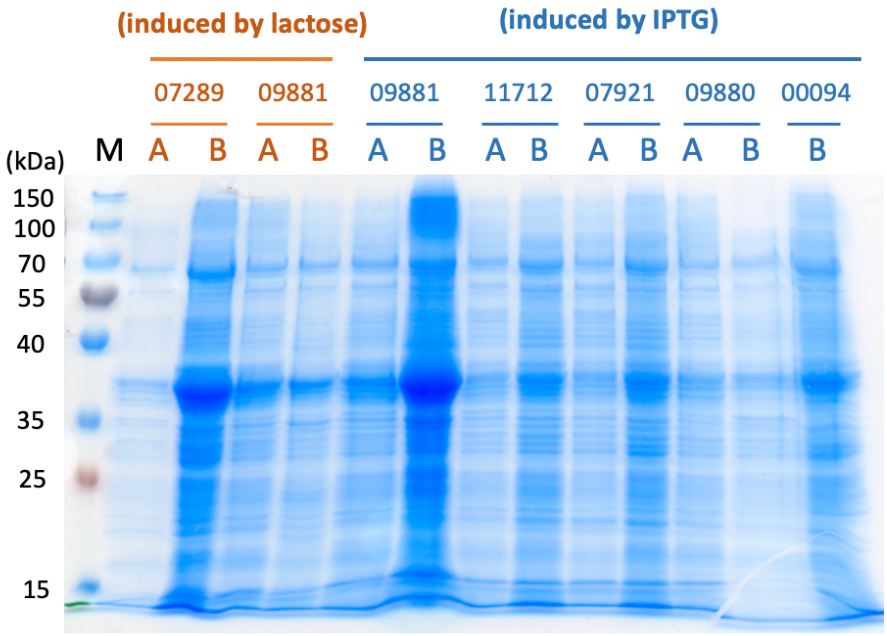


Figure 8. SDS-PAGE analysis of the comparison between lactose or IPTG as an inducer.
A: Arctic Express, B: BL21 star, M: marker.

In dehydrogenase activities assay I **(Figure 9)**, four proteins (00094, 07921, 09880, 11712) showed no activity in the pH 7 phosphate buffer, and so did the five proteins (00094, 07921, 09880, 09881, 11712) in the pH 9 TRIS-HCl. Among them, only one sample of 00094 produced in AE showed strong activity, but it was not consistent with other replicates. For these protein samples, there was a minor difference in activity between enzymes produced in AE and BL21 star. Therefore, in the subsequent experiments, only BL21 star would be used for protein production.

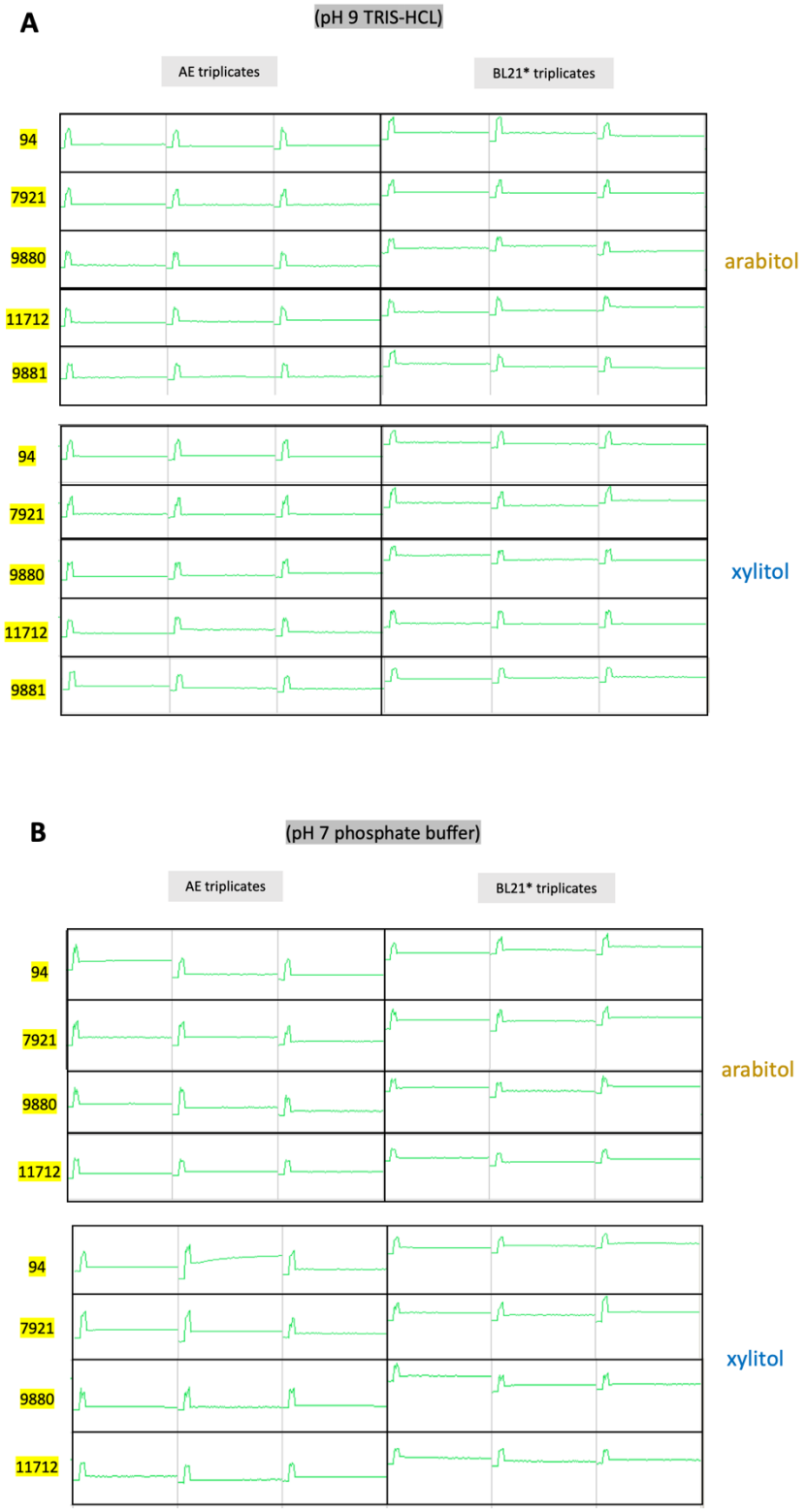


Figure 9. Overview of enzyme activity with NADP in dehydrogenase activities assay I.

In dehydrogenase activity assay II, another cofactor, NAD, was also tested **(Figure 10)**. As shown in **Figure 10-A**, enzymes were active in the buffer pH 8.5 TRIS-HCl. Especially, protein 00094, 09880, and 07289 had obviously strong activities on xylitol. However, in the pH 7 phosphate buffer **(Figure 10-B)**, except for the low activity of protein 00094, most of the samples' activities were not noticeable. **Figure 11** shows the results of the NADP test. Except for one of the duplicates in 00094 and 07921 that seemed to be active (showed ascending absorbance in reaction with arabitol), all the samples had no activity in both buffers.

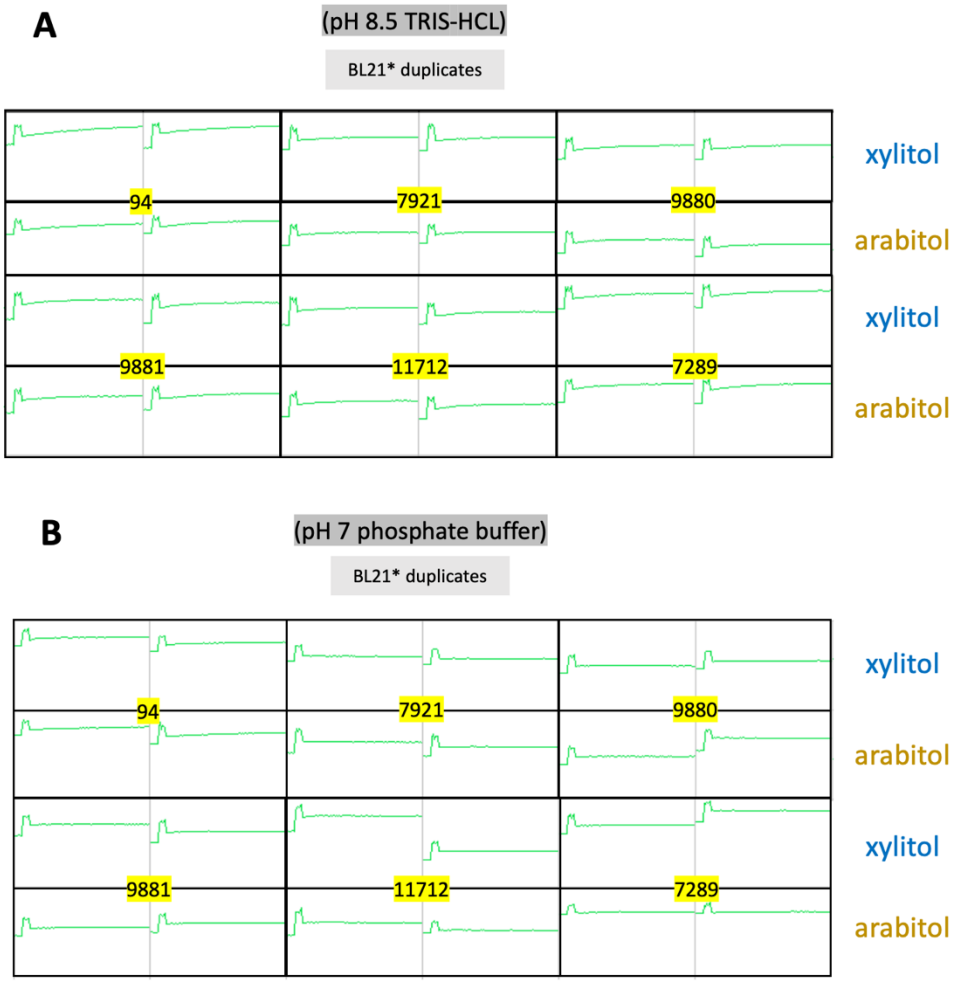


Figure 10. Overview of enzyme activity with NAD in dehydrogenase activity assay II.

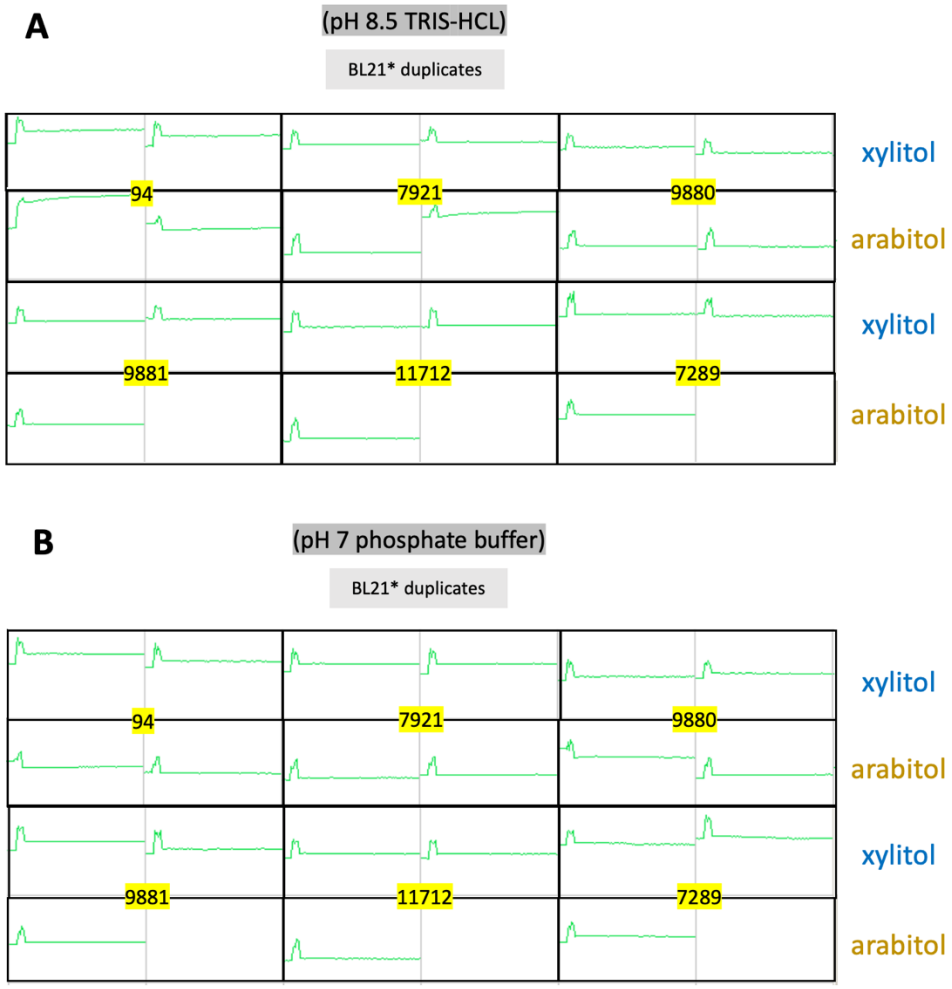


Figure 11. Overview of enzyme activity with NADP in dehydrogenase activity assay II.

In the dehydrogenase activities assay III **(Figure 13)**, every reaction with NADP showed no activity as the NADP test in dehydrogenase activities assay II. Aside from NADP, NAD was again tested in assay III. In comparison with the assay with NADP, samples in buffer pH 8.5 TRIS-HCl with NAD **(Figure 12)** had very slight activities. Nevertheless, they were still not active as the NAD test in dehydrogenase activities assay II.

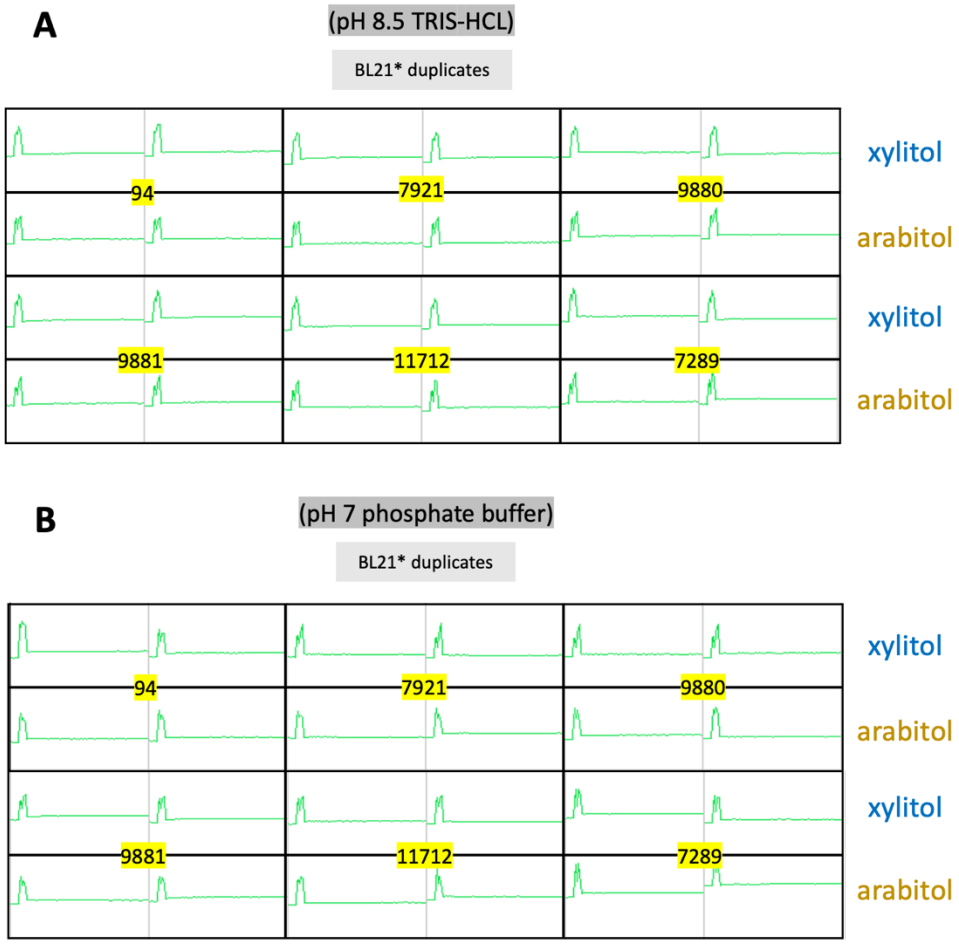


Figure 12. Overview of enzyme activity with NAD in dehydrogenase activity assay III.

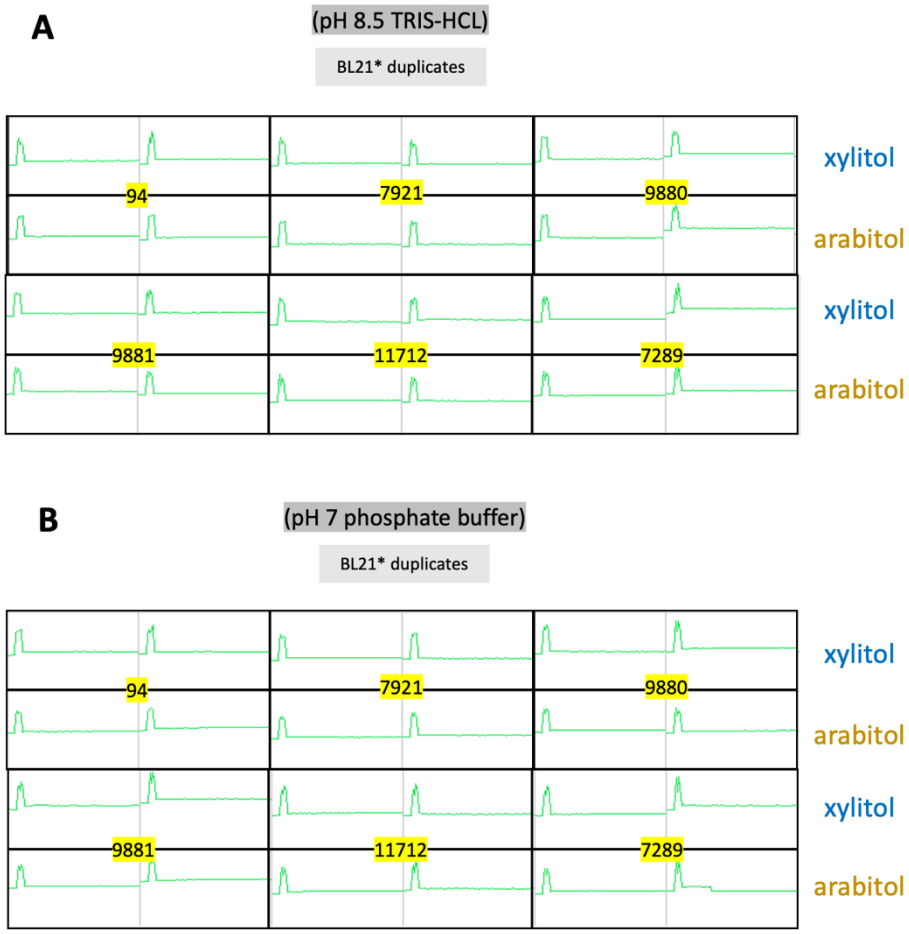


Figure 13. Overview of enzyme activity with NADP in dehydrogenase activity assay III.

In the reductase activities assay I **(Figure 14)**, the reaction mixture was tested with NADPH as a cofactor. In general, enzymes produced in BL21 star had higher activities on both substrates than the enzymes produced in AE. Therefore, only BL21 star was chosen to produce the proteins in the follow-up assays.

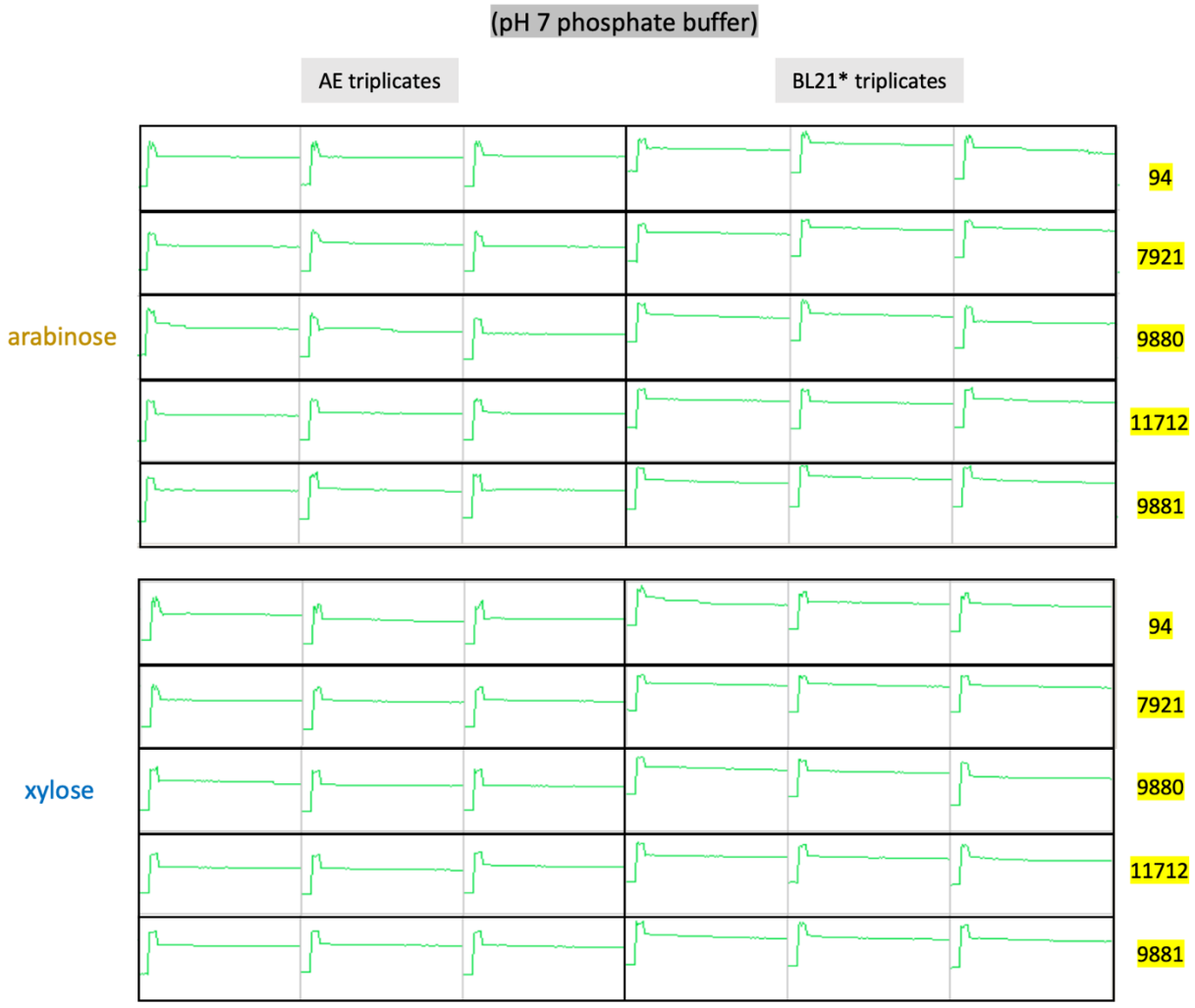


Figure 14. Overview of enzyme activity with NADPH in reductase activity assay I.

Another cofactor- NADH- was used in the reductase activities assay II. As shown in **Figure 15**, we can see most of the proteins show positive activities on xylose and arabinose, except for sample of 07289 on arabinose not showing activity, and samples of 00094 and 07921 on xylose showing unnormal signal.

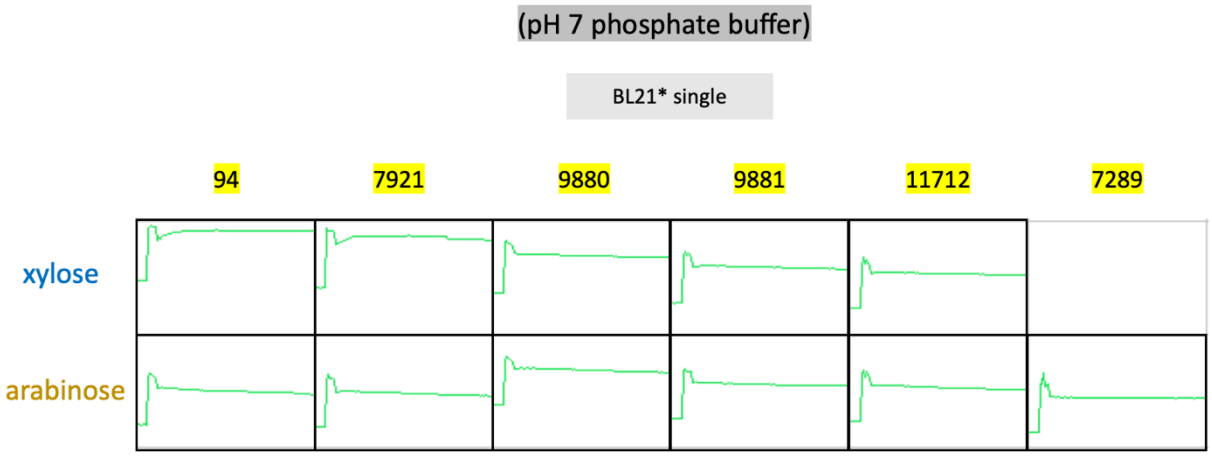


Figure 15. Overview of enzyme activity with NADH in reductase activities assay II.

In the reductase activities assay III, both NADH **(Figure 16)** and NADPH **(Figure 17)** were examined again with the reaction mixture. Currently, these six proteins showed more activities with NADPH. Especially on xylose, all proteins showed activity other than 09881.

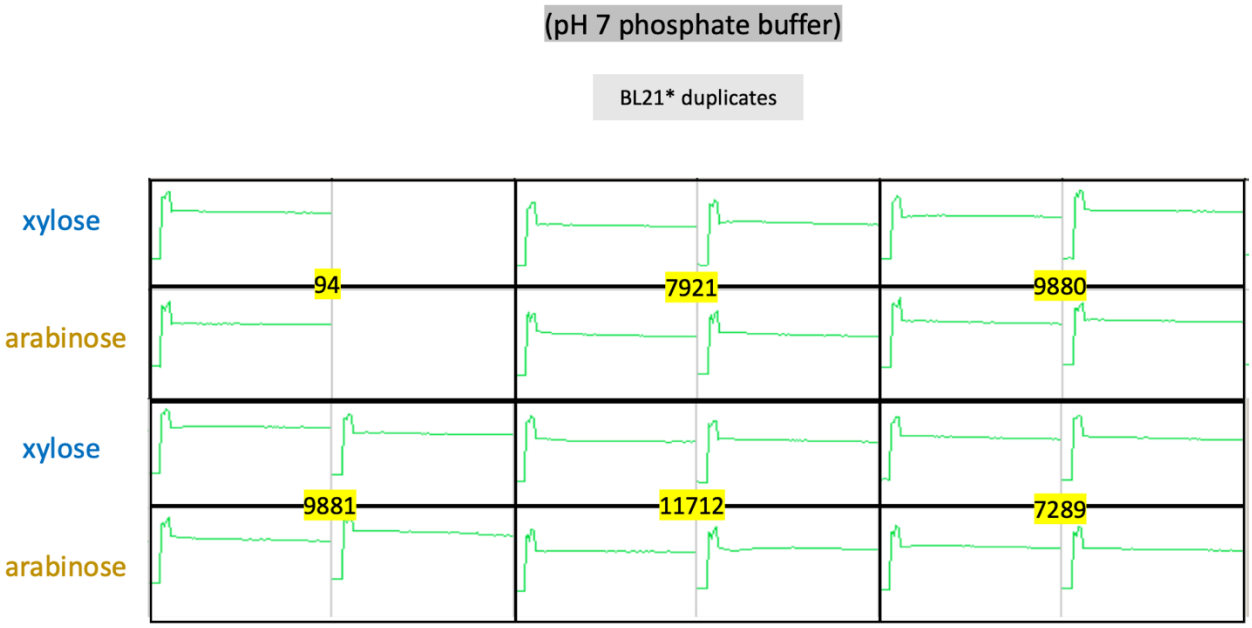


Figure 16. Overview of enzyme activity with NADH in reductase activities assay III.

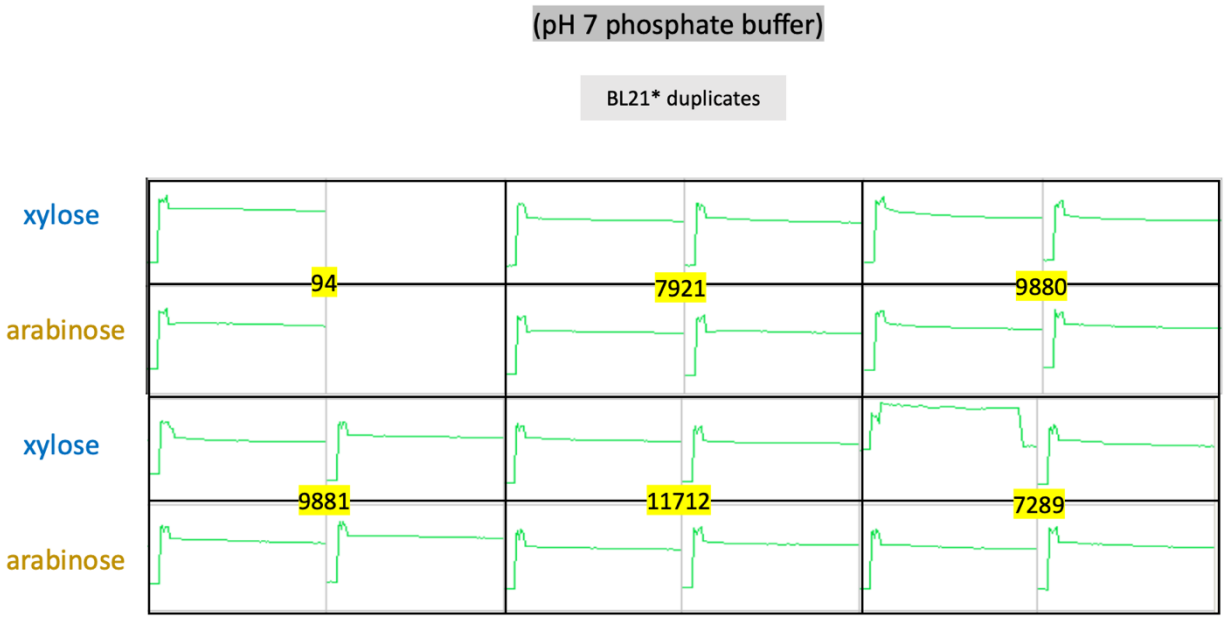


Figure 17. Overview of enzyme activity with NADPH in reductase activities assay III.

To sum up, the initial enzyme activity assays presented that these six potential proteins performed active when testing their dehydrogenase properties with cofactor NAD, while they were not active with NADP. In the test of reductase feature, these proteins could be active in the reaction with both NADH and NADPH. However, reductase activities assay II presented much more obvious results, while the outcomes from reductase activities assay III were less noticeable. According to **Figure 10**, **Figure 11**, and **Figure 15** to **Figure 17**, the overall initial activity of six samples was listed in **Table 13**.

Table 13. Overview of initial enzyme activity of six proteins in Activities Assays II and III. NAD and NADP were tested in pH 8.5 TRIS-HCl buffer while NADH and NADPH were tested in pH 7 phosphate buffer. +: positive, -: negative, -/+: undefined

	ACTIVITY ASSAY II						**ACTIVITY ASSAY III**			
PROTEIN ID	Xylitol /NAD	Arabitol /NAD	Xylitol /NADP	Arabitol /NADP	Xylose /NADH	Arabinose /NADH	Xylose /NADH	Arabinose /NADH	Xylose /NADPH	Arabinose /NADPH
00094	+	+	-	-/+	?	+	-/+	-	+	-/+
07921	-/+	-	-	-/+	?	+	-/+	+	+	-/+
09880	+	-/+	-	-	+	+	-	-/+	+	+
09881	-/+	+	-	-	+	+	-	+	-/+	+
11712	-/+	-/+	-	-	-/+	+	-/+	-	+	-/+
07289	+	+	-	-	N/A	-/+	+	-/+	+	+

3.4 Big-scale protein production and activity assay

Based on **Table 13**, proteins 00094, 09880, 09881, and 07289 were selected to continue the big-scale production because of their better performance in the dehydrogenase test. After being heterologous produced in expression strain BL21 star, the proteins were harvested, filtered, and purified. Then the target protein was eluted by ÄKTA, and FPLC chromatograms of each one was generated, which are shown in **Figure 18**.

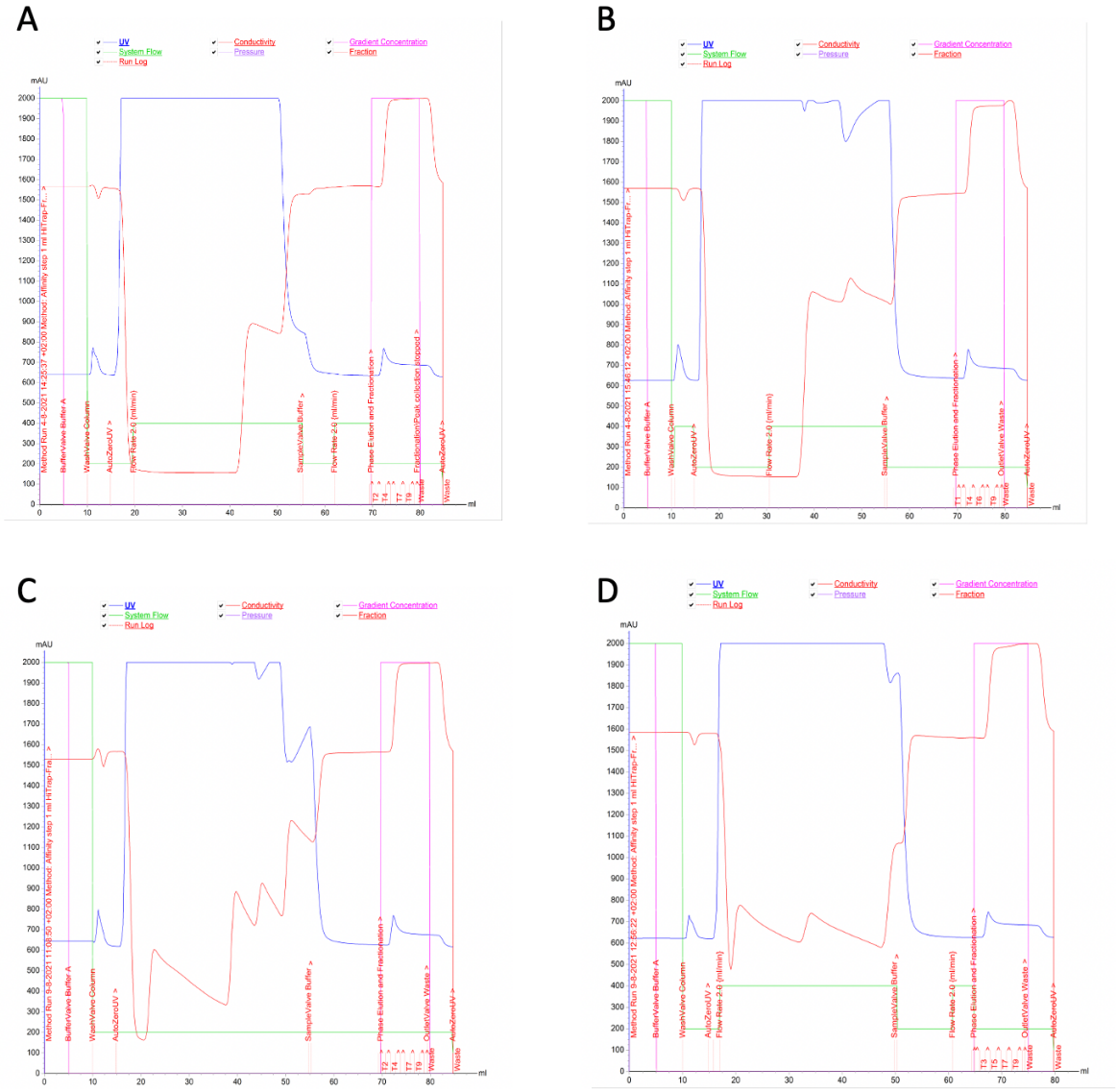


Figure 18. Elution profile of protein produced on a big scale. The red line indicates the conductivity, the pink line indicates gradient concentration, the blue line indicates UV signal. During the time of elution volume between 20 and 55 mL, sample application was proceeding with a flow rate of 2.0 mL/min. After the flow-through process, unbound proteins were washed out and then the elution and fractionation phase began when 70 mL of elution volume had been used. 09881 started the elution process earlier, at around the time of 65 mL. The peak between 70-80 mL indicates the target protein and the number of fraction tubes. A: 00094, B: 07289, C:09880, D:09881.

According to the absorbance peak in the FPLC chromatograms, the target purified protein was collected from Fractions 3 to Fractions 5 and combined. The absorbance of them was around 700-800 mAU. Among these four proteins, 00094 was the first to be pooled and desalted. According to the standard curve generated by BCA assay **(Figure 3)**, the concentration of 00094 is 421.24 ug/ml, calculated with the formula: y=989.83x-140.98, with an R^2 of 0.99. The enzyme activity assays were performed with a diluted protein concentration of 0.2 mg/ml. According to the result of two assays **(Figure 19 and 20)**, there was no enzyme activity on these substrates presenting in the large-scaled produced protein 00094.

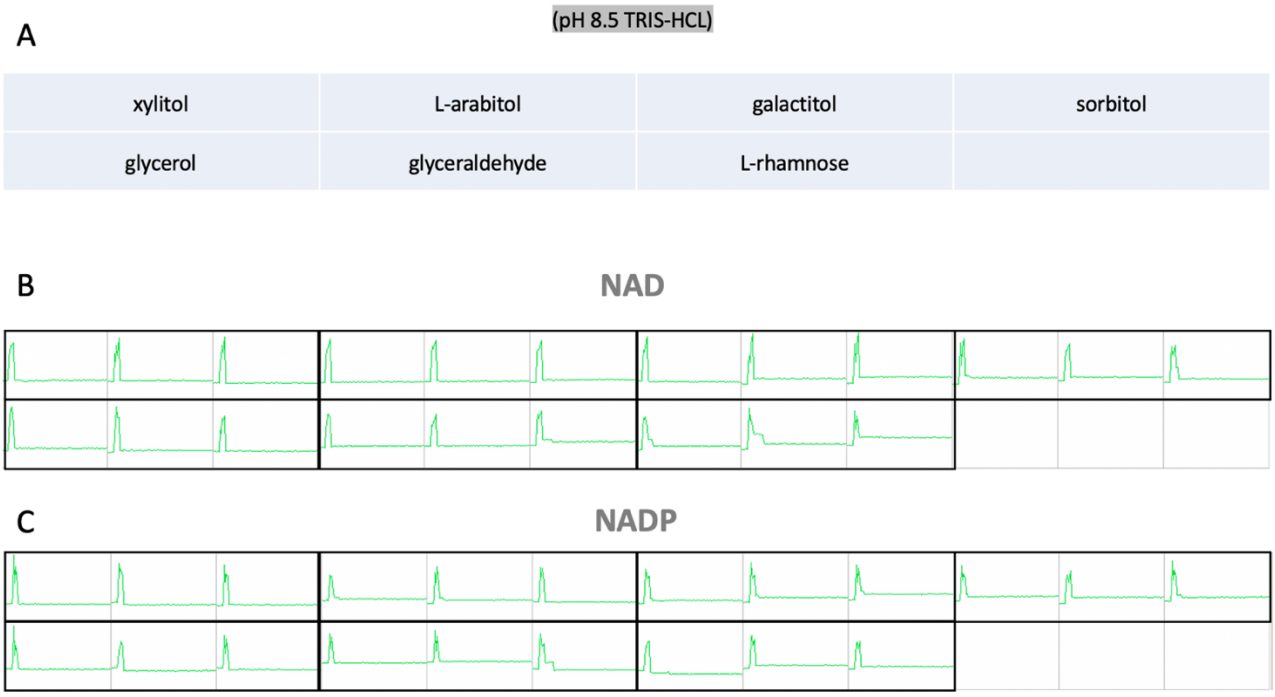


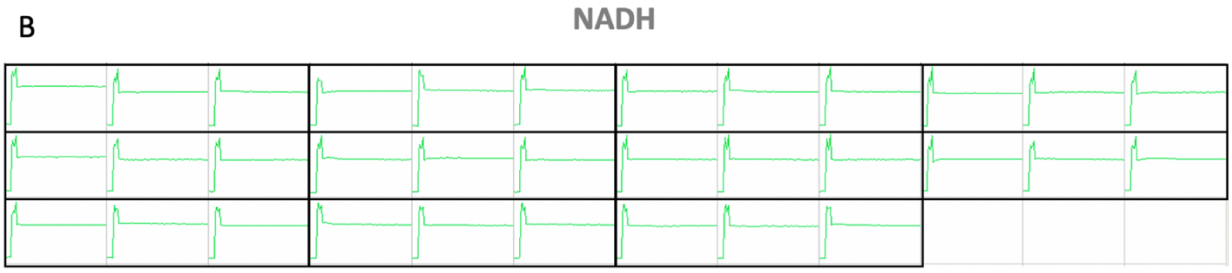
Figure 19. Overview of enzyme activity in dehydrogenase activities assay of big scale produced protein 00094. A: The order of substrates in plot B and C; B: Results of triplicates in NAD; C: Results of triplicates in NADP.

A

(pH 7 phosphate buffer)

D-xylose	D-arabinose	L-xylose	L-arabinose
D-glucose	D-fructose	L-rhamnose	D-galactose
L-sorbose	D-ribose	D-mannose	

B



C

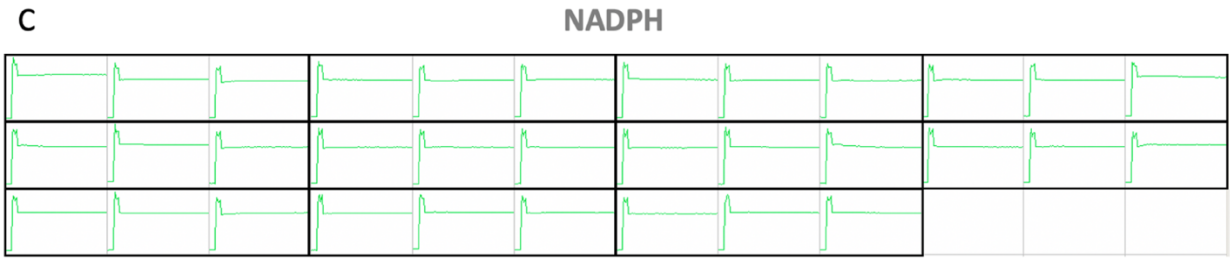


Figure 20. Overview of enzyme activity in reductase activities assay of big scale produced protein 00094. A: The order of substrates in plot B and C; B: Results of triplicates in NADH; C: Results of triplicates in NADPH.

Discussion

This project is aiming at finding the genes in *Aspergillus niger* that are related to the part of L-xylulose conversion in Pentose Catabolic Pathway. In the first selection step, Pfam 00106 family, which is also known as the short-chain dehydrogenases/reductases family (SDR), was chosen as LxrA and LxrB also belong to this family (Chroumpi et al., 2021). Based on the earlier research, the L-xylulose reductases in *Trichoderma reesei*, *Neurospora crassa*, *Aspergillus niger*, and yeast *Ambrosiozyma monospora* all belong to SDRs (Kallberg et al., 2009; Metz et al., 2013; Nair & Zhao, 2007; Verho et al., 2004). Although they are all in Pfam 00106 and the deletions of genes encoding these enzymes could result in reduced growth on L-arabinose and L-arabitol, they are not all phylogenetically close related to each other (Seiboth & Metz, 2011). Therefore, the second-round selection was according to the gene expression on various substrates.

SDR family represents a wide group of proteins that have at least two domains. One is the functional region that can bind the coenzyme, and the other is the catalytic region (Mistry et al., 2021). As for the substrates, it is expected that L-xylulose reductase in *A. niger* can be active on L-arabinose and D-xylose, which are the sugars involved in PCP (Witteveen et al., 1989). Subsequently, LxrA in *A. niger* was found, with the initial observation of transcription on L-arabinose and D-xylose as well, and no expression on D-glucose. Besides, in reference to the transcriptome analysis of Andersen et al., the L-arabinitol-4-dehydrogenase was upregulated on D-xylose although it does not contribute to D-xylose catabolism (Andersen et al., 2008; Mojzita et al., 2010). Based on this information, genes in Pfam 00106 family were then chosen by their comparative expression of 'on L-arabinose/on D-glucose' and/or 'on D-xylose/on D-glucose,' to consider more possibilities.

Among the candidate genes that had comparative higher values in L-arabinose and/or D-xylose, compared to glucose, NRRL3_09880 has a similar pattern to NRRL3_10884, which is the confirmed L-xylulose reductase LxrA. To illustrate, they are highly expressed on L-arabinose, D-xylose, and low on D-glucose, which is the control substrate. Moreover, they showed lower values on L-arabinose in the AraR deletion mutant, high values on D-xylose in the XlnR deletion mutant. And the values on D-xylose in XlnR

deletion mutant were higher compared to the wild type. XlnR is the D-xylose-related transcriptional activator and has interactions with AraR in specific steps in the PCP regulation system. Also, *araR* increased 3-fold more expression in the Δ*xlnR* on D-xylose than in the wild type (Battaglia et al., 2011b), and *xdhA* showed up-regulation on D-xylose in the Δ*xlnR* strain compared to the reference strain (Battaglia et al., 2011b; de Groot et al., 2007). These suggest that NRRL3_09880 might also take a similar role in the arabinanolytic pathway. Previous research showed that there was no further reduced phenotype when NRRL3_09880 was deleted in the Δ*lxrA*Δ*lxrB* strain. However, considering that *lxrA* had the same situation of phenotype on its single deletion and similar expression pattern on related conditions, NRRL3_09880 could still be expected to act as one of the essential enzymes in this metabolic mechanism with high complexity (Chroumpi et al., 2021).

As expected, the ten candidate genes were in quite separate places on the phylogenetic tree of Pfam 00106, except for NRRL3_04005 and NRRL3_09880 closer to *lxrB*. This proved again the challenge in finding the other enzymes involved in the L-xylulose conversion (Chroumpi et al., 2021). As LxrA and LxrB also belong in different clades of Pfam 00106, the extra enzymes could be barely identified with the aid of phylogenetic tree.

In this project, gene expression on D-xylose was considered because more possibilities could be included in the situation of complicated mechanisms in PCP existing. However, L-xylulose reductases are the exclusive enzymes in PCP that are not involved in both conversion pathways. This makes *lxr* different from other PCP genes that could be induced on both pentoses and even have different responses depending on species (Mojzita et al., 2010). Concerning that the co-induction of D-xylose and L-arabinose for fungus to be able to break down the plant cell wall (Seiboth & Metz, 2011), and the tiny amounts of arabinose in the commercial xylose preparations that can make an artefact (Andersen et al., 2008), D-xylose may not be a decisive factor to help with the choice. Besides, transcription of LxrA on D-xylose only showed in the first 2h, which differs a lot from the continuing transcription on L-arabinose. Since the L-arabitol is also a crucial and main inducer apart from L-arabinose of the arabinanolytic system and the extracellular arabinanolytic enzymes are co-regulated with the genes of L-arabinose pathway (de Vries, 1994, 2003), more transcriptome data on various comparisons might be useful. For example, the

gene expressions on L-arabitol, or on each substrate after 2h, or relationship of values between the extracellular arabinanolytic enzymes and Pfam 00106 family on L-arabitol and L-arabinose.

As for the protein production part, the main problem is the deficiency and inequality of the yields. The yield shortage was shown in both small-scale and large-scale productions. In the SDS-PAGE analysis of small-scale production **(Figure 6, 7)**, the bands between 35 kDa and 40 kDa were considered as the target proteins and they showed that no overproduction happened. Also, with the BCA assay in big-scale production, the calculated concentration of undiluted protein 00094 was ten times lower than expected, which subsequently influenced the degree of activity that could be detected in the enzyme tests. Furthermore, when comparing three enzyme activity tests, there were notably high-density growth of cells (Based on the visual assessment) and obvious activity outcomes shown in the activity assay II **(Figure 10, 11, and 15)** than activity assay I and III. Because all the materials including substrates, cofactors, and buffers in three tests were the same, the most different part that might make the effects in the experimental process was the inoculation source. In activity assay II, 200 µL of glycerol stock was added to incubate overnight while only a small picking of colony or glycerol stock was used in the pre-culture for overnight incubation in activity assay I and III, respectively.

The addition of glycerol impacts on the production host because the high substrate concentration minimizes the growth reduction that is supposed to happen in the heterologous cells through the *lac* expression with IPTG induction (Dekel & Alon, 2005; Malakar & Venkatesh, 2012). That is a phenomenon known as 'metabolic burden' or 'product burden.' It describes that stress reductions occur to optimize the capacity to grow because limited cellular resources are also needed for the formation of essential proteins, which are ribosomes. Also, the uptake of IPTG is related to the proton pumps (Krzewinski et al., 1996; Seol & Shatkin, 1992). As a result, cell growth rate decreases or even leads to cell lysis when IPTG being the inducer in the case of Lac protein expression (Kopp et al., 2017; Malakar & Venkatesh, 2012; Vind et al., 1993). Notably, the research of Malakar and Venkatesh showed that two phases of the reduction in growth rates happened when the expression cells grew at a fixed IPTG concentration with varying glycerol concentration. When the glycerol concentration was between 0 and 2 g/L, the reduction in growth rate increased due to the burden, while the reduction in growth rate decreased and even

saturated to zero when the glycerol concentration was beyond 2 g/L. This experiment might support the higher cells amount and protein activity in activity assay II, compared to the other two assays.

Apart from IPTG, lactose was also tested as the inducer to compare the production differences. It has been shown to be able to lighten the toxicity caused by IPTG without efficiency reduction (Dvorak et al., 2015). Lactose is a natural sugar effector that can be metabolized by cells and is considered as an alternative to be used in cells with *lac* operon. Although there were not sufficient samples in this experiment **(Figure 8)**, an obvious distinction could be seen on protein 09881. Furthermore, the solubility of these recombinant proteins was seen. According to the SDS-PAGE analysis **(Figure 6-8)**, they were soluble in the supernatant of cell extract but with a remarkably low amount. Inclusion bodies are possible to form because the adding of histidine tags, which are used to assist the purification of target proteins, can affect the protein solubility (Zhu et al., 2013). However, insoluble protein aggregating in the inclusion body indicates that the recombinant protein might be misfolded, which causes a huge problem in heterologous expression (Baneyx, 1999; Upadhyay et al., 2012; Villaverde & Mar Carrió, 2003; Vuillemin et al., 2014). The use of lactose is beneficial to heterologous protein formation. With lactose, a softer induction can be conducted in the production host, especially coupling with glycerol and glucose in the auto-induction systems (Kopp et al., 2017). Besides, glycerol can also be a protective agent to prevent proteins from aggregating (Lin, 1976).

Two main observations were involved in the initial enzyme activity tests of small-scale production. First, in the activity assay II, the buffer was one of the factors that affect the degree of detectability. By comparing **Figure 10-A and 10-B**, a clear difference was given out. This could be expected according to the explanation by Richard et al. It was considered that phosphate had an inhibitory effect on Lxr1 from *T. reesei* (*H. jecorina*) while the enzyme is like Lxr in *A. niger*; as a result, a similar effect was shown (Mojzita et al., 2010; Richard et al., 2002). Second, NAD is the effective cofactor in the dehydrogenase test in contrast to NADP. In the case of effective experiments (activity assay II), candidate proteins were highly active with NAD in the buffer environment of pH 8.5 TRIS-HCl, which are shown in **Figure 10, 11, and Table 13**. This indicates that these proteins show NAD dependence in the context of sufficient

concentration and amount. On the reverse reaction, obvious activity on NADH for the reductase test was also shown in activity assay II (**Figure 15 and Table 13**).

This finding can prove the inference that was mentioned in earlier research. To illustrate, L-xylulose reductase activity with both NADH and NADPH had already been described in *A. niger* previously (Mojzita et al., 2010). In that research, reductase activity with NADPH in the LxrA deletion strain had almost disappeared while the activity with NADH remained high. This was considered as the reason that the LxrA deletion strain could still grow on L-arabinose and L-arabitol. Additionally, the complete loss of NADPH activity also indicated that the additional enzymes in this step could be a non-NADPH- but NADH-dependent reductase (Metz et al., 2013; Mojzita et al., 2010; Seiboth & Metz, 2011). After the NADPH-dependent LxrB was found, residual growth and reductase activity were still found in the double deletion strain on L-arabinose and L-arabitol (Chroumpi et al., 2021). This indicates that the previous suggestions linked to the presence of NADH-dependent reductase are worth validating.

Overall, in this project, we were able to not only show that a few selected genes based on transcriptome were able to respond to the possibility of NAD-dependence, but also present the difference in expression strain, production amount, cofactor, and buffer. Although the yeast *Ambrosiozyma monospora* has an NADH-dependent L-xylulose reductase Alx1 and uses the same pathway as filamentous fungi (Verho et al., 2004; Fonseca et al., 2007), all discovered L-xylulose reductases in filamentous fungi to date were described as NADPH-dependent enzymes, such as Lxr3 in *T. reesei* (Metz et al., 2013) and Lxr in *Neurospora crassa* (Nair & Zhao, 2007). Also, it is considered that every reduction step in *A. niger* is NADPH-dependent to convert the sugars to the corresponding polyols (Witteveen et al., 1989). As the results are shown in this project, *A. niger* is a potential filamentous fungus that NADH-dependent L-xylulose reductases can be found in its metabolic pathways. Therefore, extended experiments should be conducted to gain more information aiming at the genes and corresponding cofactors. For example, the L-xylulose reductase activity with NADH of the double deletion of *lxrA* and *lxrB*.

As for protein production, solubility is still a challenge for recombinant proteins in some cases. Many factors are involved in the soluble enzyme yields and can be improved, such as expression strains, vectors,

the position of affinity tags, the temperature of incubation, culture duration, type of inducer, glucose, and glycerol concentration (Vuillemin et al., 2014). In this project, the concentration of glycerol brings out a notable change. Although the protein amounts shown on SDS-PAGE were not notable, enzyme activities were still shown in activity assay II. The impact of glycerol on the viability of cells and the production of soluble recombinant proteins was also proved by previous research (Al-Samarrai et al., 2013; Kopp et al., 2017; Vuillemin et al., 2014). That can be explained by enough energy that a high concentration of substrate supply for synthesizing essential and unneeded proteins (Malakar & Venkatesh, 2012). Apart from glycerol itself, the auto-induction system can be used for improvement as it leads to a diauxic growth that can control induction well in the host metabolism (Blommel et al., 2008; Tyler et al., 2005; Vuillemin et al., 2014), and the mixed feed system was also reported to result in considerable amounts of correctly folded protein (Kittler et al., 2020). The other production suggestion would be a softer induction with lactose, as the concentration of both substrate and IPTG is important for T7-based plasmids to express Lac protein (Kittler et al., 2020; Malakar & Venkatesh, 2012).

References

Aguilar-Pontes, M. V., Brandl, J., McDonnell, E., Strasser, K., Nguyen, T. T. M., Riley, R., Mondo, S., Salamov, A., Nybo, J. L., Vesth, T. C., Grigoriev, I. V., Andersen, M. R., Tsang, A., & de Vries, R. P. (2018). The gold-standard genome of *Aspergillus niger* NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi. *Studies in Mycology*, *91*. https://doi.org/10.1016/j.simyco.2018.10.001

Al-Samarrai, T. H., Jones, W. T., Harvey, D., Kirk, C. A., & Templtone, M. (2013). Effect of 4% glycerol and low aeration on result of expression in *Escherichia coli* of Cin3 and three *Venturia inaequalis* EST's recombinant proteins. *American Journal of Molecular Biology*, *03*(01). https://doi.org/10.4236/ajmb.2013.31001

Andersen, M. R., Vongsangnak, W., Panagiotou, G., Salazar, M. P., Lehmann, L., & Nielsen, J. (2008). A trispecies *Aspergillus* microarray: Comparative transcriptomics of three *Aspergillus* species. *Proceedings of the National Academy of Sciences*, *105*(11). https://doi.org/10.1073/pnas.0709964105

Attrill, H., Gaudet, P., Huntley, R. P., Lovering, R. C., Engel, S. R., Poux, S., van Auken, K. M., Georghiou, G., Chibucos, M. C., Berardini, T. Z., Wood, V., Drabkin, H., Fey, P., Garmiri, P., Harris, M. A., Sawford, T., Reiser, L., Tauber, R., & Toro, S. (2019). Annotation of gene product function from high-throughput studies using the Gene Ontology. *Database*, *2019*. https://doi.org/10.1093/database/baz007

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, *10*(5). https://doi.org/10.1016/S0958-1669(99)00003-8

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, *115*(25). https://doi.org/10.1073/pnas.1711842115

Battaglia, E. (2011). Regulation and diversity of plant polysaccharide utilisation in fungi (ISBN: 978-90-8891-287-0) [PhD thesis, Utrecht University]. Utrecht University Repository. https://dspace.library.uu.nl/handle/1874/205435

Battaglia, E., Visser, L., Nijssen, A., van Veluw, G. J., Wösten, H. A. B., & de Vries, R. P. (2011b). Analysis of regulation of pentose utilisation in *Aspergillus niger* reveals evolutionary adaptations in Eurotiales. *Studies in Mycology*, *69*. https://doi.org/10.3114/sim.2011.69.03

Blommel, P. G., Becker, K. J., Duvnjak, P., & Fox, B. G. (2008). Enhanced bacterial protein expression during auto-induction obtained by alteration of Lac repressor dosage and medium composition. *Biotechnology Progress*, *23*(3). https://doi.org/10.1021/bp070011x

Boutard, M., Cerisy, T., Nogue, P.-Y., Alberti, A., Weissenbach, J., Salanoubat, M., & Tolonen, A. C. (2014). Functional diversity of carbohydrate-active enzymes enabling a bacterium to ferment plant biomass. *PLoS Genetics*, *10*(11). https://doi.org/10.1371/journal.pgen.1004773

Chroumpi, T., Peng, M., Aguilar-Pontes, M. V., Müller, A., Wang, M., Yan, J., Lipzen, A., Ng, V., Grigoriev, I. v., Mäkelä, M. R., & de Vries, R. P. (2021). Revisiting a 'simple' fungal metabolic pathway reveals redundancy, complexity and diversity. *Microbial Biotechnology*. https://doi.org/10.1111/1751-7915.13790

de Groot, M. J. L., van den Dool, C., Wösten, H. A. B., Levisson, M., vanKuyk, P. A., Ruijter, G. J. G., & de Vries, R. P. (2007). Regulation of pentose catabolic pathway genes of *Aspergillus niger*. *Food Technology and Biotechnology*, *45*(2), 134–138.

de Vries, R. P. (2003). Regulation of *Aspergillus* genes encoding plant cell wall polysaccharide-degrading enzymes; relevance for industrial production. *Applied Microbiology and Biotechnology*, *61*(1). https://doi.org/10.1007/s00253-002-1171-9

de Vries, R. P., Flipphi, M. J. A., Witteveen, C. F. B., & Visser, J. (1994). Characterization of an *Aspergillus nidulans* L-arabitol dehydrogenase mutant. *FEMS Microbiology Letters*, *123*(1–2). https://doi.org/10.1111/j.1574-6968.1994.tb07205.x

Dekel, E., & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, *436*(7050). https://doi.org/10.1038/nature03842

Dvorak, P., Chrast, L., Nikel, P. I., Fedr, R., Soucek, K., Sedlackova, M., Chaloupkova, R., de Lorenzo, V., Prokop, Z., & Damborsky, J. (2015). Exacerbation of substrate toxicity by IPTG in *Escherichia coli* BL21(DE3) carrying a synthetic metabolic pathway. *Microbial Cell Factories*, *14*(1). https://doi.org/10.1186/s12934-015-0393-3

Fisher, A. K., Freedman, B. G., Bevan, D. R., & Senger, R. S. (2014). A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *Computational and Structural Biotechnology Journal*, *11*(18). https://doi.org/10.1016/j.csbj.2014.08.010

Gaurav, G. K., Mehmood, T., Cheng, L., Klemeš, J. J., & Shrivastava, D. K. (2020). Water hyacinth as a biomass: A review. *Journal of Cleaner Production*, *277*. https://doi.org/10.1016/j.jclepro.2020.122214

Hasper, A. A., Trindade, L. M., van der Veen, D., van Ooyen, A. J. J., & de Graaff, L. H. (2004). Functional analysis of the transcriptional activator XlnR from *Aspergillus niger*. *Microbiology*, *150*(5). https://doi.org/10.1099/mic.0.26557-0

Horecker, B. L. (2002). The pentose phosphate pathway. *Journal of Biological Chemistry*, *277*(50). https://doi.org/10.1074/jbc.X200007200

Houfani, A. A., Anders, N., Spiess, A. C., Baldrian, P., & Benallaoua, S. (2020). Insights from enzymatic degradation of cellulose and hemicellulose to fermentable sugars– a review. *Biomass and Bioenergy*, *134*. https://doi.org/10.1016/j.biombioe.2020.105481

Kallberg, Y., Oppermann, U., Jörnvall, H., & Persson, B. (2009). Short-chain dehydrogenase/reductase (SDR) relationships: A large family with eight clusters common to human, animal, and plant genomes. *Protein Science*, *11*(3). https://doi.org/10.1110/ps.26902

Khosravi, C. (2017). Sugar catabolism during growth on plant biomass in *Aspergillus* (ISBN: 978-94-6295-773-2) [PhD thesis, Utrecht University]. Utrecht University Repository. https://dspace.library.uu.nl/handle/1874/357322

Kittler, S., Kopp, J., Veelenturf, P. G., Spadiut, O., Delvigne, F., Herwig, C., & Slouka, C. (2020). The Lazarus *Escherichia coli* effect: Recovery of productivity on glycerol/lactose mixed feed in continuous biomanufacturing. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.00993

Klaubauf, S. (2015). Regulation of plant polysaccharide utilisation in *Magnaporthe oryzae* and other ascomycetous fungi (ISBN: 978-94-6108-940-3) [PhD thesis, Utrecht University]. Utrecht University Repository. https://pure.knaw.nl/portal/en/publications/62297c28-73ae-43ac-9904-7dccf1c92ba0

Kopp, J., Slouka, C., Ulonska, S., Kager, J., Fricke, J., Spadiut, O., & Herwig, C. (2017). Impact of glycerol as carbon source onto specific sugar and inducer uptake rates and inclusion body productivity in *E. coli* BL21(DE3). *Bioengineering*, *5*(1). https://doi.org/10.3390/bioengineering5010001

Krzewinski, F., Brassart, C., Gavini, F., & Bouquelet, S. (1996). Characterization of the lactose transport system in the strain *Bifidobacterium bifidum* DSM 20082. *Current Microbiology*, *32*(6). https://doi.org/10.1007/s002849900054

Lin, E. C. C. (1976). Glycerol dissimilation and its regulation in bacteria. *Annual Review of Microbiology*, *30*(1). https://doi.org/10.1146/annurev.mi.30.100176.002535

Malakar, P., & Venkatesh, K. v. (2012). Effect of substrate and IPTG concentrations on the burden to growth of *Escherichia coli* on glycerol due to the expression of Lac proteins. *Applied Microbiology and Biotechnology*, *93*(6). https://doi.org/10.1007/s00253-011-3642-3

Metz, B., Mojzita, D., Herold, S., Kubicek, C. P., Richard, P., & Seiboth, B. (2013). A novel L-xylulose reductase essential for L-arabinose catabolism in *Trichoderma reesei*. *Biochemistry*, *52*(14). https://doi.org/10.1021/bi301583u

Meyer, V., Basenko, E. Y., Benz, J. P., Braus, G. H., Caddick, M. X., Csukai, M., de Vries, R. P., Endy, D., Frisvad, J. C., Gunde-Cimerman, N., Haarmann, T., Hadar, Y., Hansen, K., Johnson, R. I., Keller, N. P., Kraševec, N., Mortensen, U. H., Perez, R., Ram, A. F. J., … Wösten, H. A. B. (2020). Growing a circular economy with fungal biotechnology: A white paper. *Fungal Biology and Biotechnology*, *7*(1). https://doi.org/10.1186/s40694-020-00095-z

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1). https://doi.org/10.1093/nar/gkaa913

Mojzita, D., Vuoristo, K., Koivistoinen, O. M., Penttilä, M., & Richard, P. (2010). The 'true' L-xylulose reductase of filamentous fungi identified in *Aspergillus niger*. *FEBS Letters*, *584*(16). https://doi.org/10.1016/j.febslet.2010.06.037

Nair, N., & Zhao, H. (2007). Biochemical characterization of an L-xylulose reductase from *Neurospora crassa*. *Applied and Environmental Microbiology*, *73*(6). https://doi.org/10.1128/AEM.02515-06

Richard, P., Putkonen, M., Väänänen, R., Londesborough, J., & Penttilä, M. (2002). The missing link in the fungal L-arabinose catabolic pathway, Identification of the L-Xylulose Reductase Gene. *Biochemistry*, *41*(20). https://doi.org/10.1021/bi025529i

Seiboth, B., & Metz, B. (2011). Fungal arabinan and L-arabinose metabolism. *Applied Microbiology and Biotechnology*, *89*(6). https://doi.org/10.1007/s00253-010-3071-8

Seol, W., & Shatkin, A. J. (1992). *Escherichia coli* alpha-ketoglutarate permease is a constitutively expressed proton symporter. *Journal of Biological Chemistry*, *267*(9). https://doi.org/10.1016/S0021-9258(18)42710-X

SnapGene (n.d.). pET-28a(+) Sequence and Map. Retrieved from https://www.snapgene.com/resources/plasmid-files/?set=pet_and_duet_vectors_(novagen)&plasmid=pET-28a(%2B)

Teleky, B.-E., & Vodnar, D. (2019). Biomass-derived production of itaconic acid as a building block in specialty polymers. *Polymers*, *11*(6). https://doi.org/10.3390/polym11061035

Terebieniec, A., Chroumpi, T., Dilokpimol, A., Aguilar-Pontes, M. V., Mäkelä, M. R., & de Vries, R. P. (2021). Characterization of D-xylose reductase, XyrB, from *Aspergillus niger*. *Biotechnology Reports*, *30*. https://doi.org/10.1016/j.btre.2021.e00610

The Gene Ontology Resource: 20 years and still GOing strong. (2019). *Nucleic Acids Research*, *47*(D1). https://doi.org/10.1093/nar/gky1055

Tyler, R. C., Sreenath, H. K., Singh, S., Aceti, D. J., Bingman, C. A., Markley, J. L., & Fox, B. G. (2005). Auto-induction medium for the production of [U-^{15}N]- and [U-^{13}C, U-^{15}N]-labeled proteins for NMR screening and structure determination. *Protein Expression and Purification*, *40*(2). https://doi.org/10.1016/j.pep.2004.12.024

Upadhyay, A. K., Murmu, A., Singh, A., & Panda, A. K. (2012). Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in *Escherichia coli*. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033951

van Peij, N. N. M. E., Visser, J., & de Graaff, L. H. (1998). Isolation and analysis of *xln*R, encoding a transcriptional activator co-ordinating xylanolytic expression in *Aspergillus niger*. *Molecular Microbiology*, *27*(1). https://doi.org/10.1046/j.1365-2958.1998.00666.x

Verho, R., Putkonen, M., Londesborough, J., Penttilä, M., & Richard, P. (2004). A novel NADH-linked L-xylulose reductase in the L-arabinose catabolic pathway of yeast. *Journal of Biological Chemistry*, *279*(15). https://doi.org/10.1074/jbc.M312533200

Villaverde, A., & Mar Carrió, M. (2003). Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnology Letters*, *25*(17). https://doi.org/10.1023/A:1025024104862

Vind, J., Sørensen, M. A., Rasmussen, M. D., & Pedersen, S. (1993). Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. *Journal of Molecular Biology*, *231*(3). https://doi.org/10.1006/jmbi.1993.1319

Vuillemin, M., Malbert, Y., Laguerre, S., Remaud-Siméon, M., & Moulis, C. (2014). Optimizing the production of an α-(1→2) branching sucrase in *Escherichia coli* using statistical design. *Applied Microbiology and Biotechnology*, *98*(11). https://doi.org/10.1007/s00253-014-5627-5

Witteveen, C. F. B., Busink, R., van de Vondervoort, P., Dijkema, C., Swart, K., & Visser, J. (1989). L-arabinose and D-xylose catabolism in *Aspergillus niger*. *Microbiology*, *135*(8). https://doi.org/10.1099/00221287-135-8-2163

Zhu, S., Gong, C., Ren, L., Li, X., Song, D., & Zheng, G. (2013). A simple and effective strategy for solving the problem of inclusion bodies in recombinant protein technology: His-tag deletions enhance soluble expression. *Applied Microbiology and Biotechnology*, *97*(2). https://doi.org/10.1007/s00253-012-4630-y

Supplementary Figures

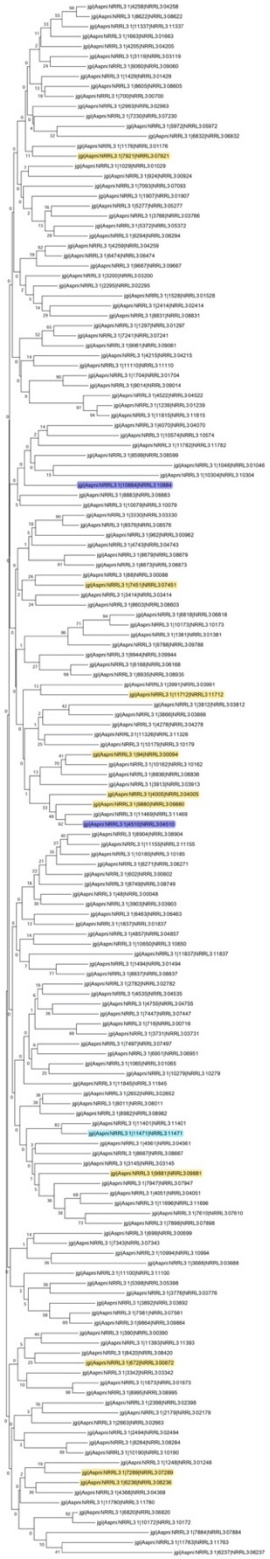


Figure S1. Phylogenetic tree of Pfam 00106. NRRL3_04510= lxrB, NRRL3_10884= lxrA. Others marked with yellow: ten selected genes. Details are shown in a separate file.

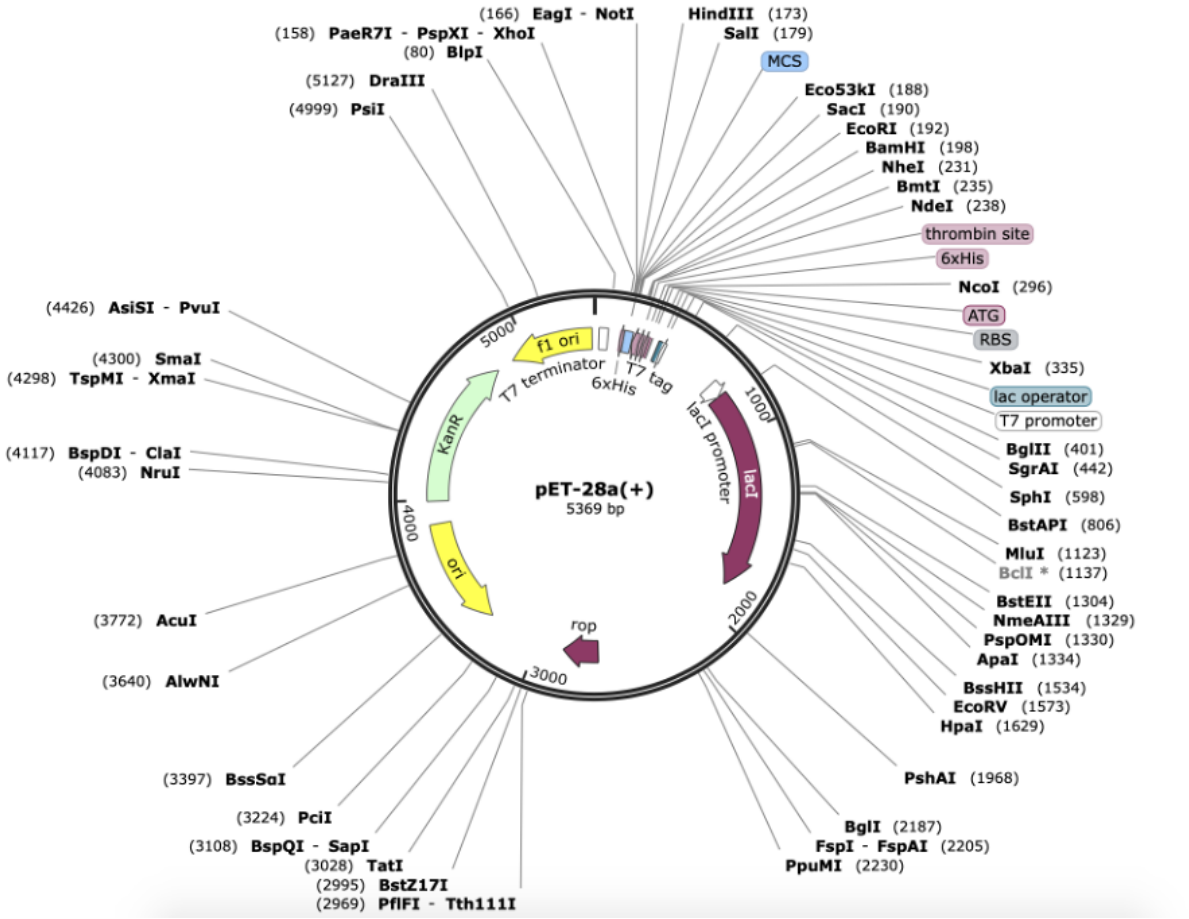


Figure S2. Gene map of pET-28a(+) plasmid. (SnapGene)