Bioproduction of D-Tagatose in Escherichia coli by Harnessing the Reverse Leloir Pathway

Bioproduction of D-Tagatose in *Escherichia coli* by Harnessing the Reverse Leloir Pathway

Christopher Toomey

A Thesis in the Field of Biotechnology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

March 2023

Abstract

D-Tagatose is a natural, rare monosaccharide containing 38% of the calories of sucrose. As an alternative sweetener, it offers numerous benefits including a low glycemic index, prebiotic properties, and the ability to lower blood sugar levels. Currently, tagatose is largely produced by industrial isomerization of galactose using both chemical and enzymatic catalysis by a class of enzymes called L-arabinose isomerases (LAIs). While these processes are effective, they are laborious and use expensive feedstocks. This study proposes and characterizes a novel, fully *in vivo* pathway to D-tagatose from a minimal-cost feedstock, glucose, using an *Escherichia coli* microbial bioproduction process. This process is possible due to the discovery and characterization of a phosphatase from, *Dictyostelium discoideum*, which exhibits selective dephosphorylation of galactose-1-phosphate to make free galactose using *E.coli* cells. By expressing this phosphatase, along with an LAI in a modified strain background that eliminates the cell's galactose catabolism potential, this system demonstrates the ability to make tagatose from glucose. The highest yields of tagatose from this project are currently 5.4%. However, it is also demonstrated that the system can be further optimized to achieve up to 77% increases in galactose, and 23% increases in tagatose. Additionally, this project shows unprecedented yields of galactose made from glucose *in vivo* (31.5%). Showing the potential of a novel system such as this demonstrates the attainability of a cheap, simple, and efficient process to make rare sugars. This could not only lead to healthier products but ones more broadly accessible to the public.

Acknowledgments

- The Manus Bio Team
- My Thesis Co-Directors Aaron Love and Christine Santos
- Professor Nikhil Nair (Tufts University Department of Chemical and Biological Engineering)

Table of Contents

List of Tables

List of Figures

Chapter I.

Introduction

This chapter contains background information on the properties and health benefits of tagatose, the current ways it is made, and this project's novel design to make it.

Glossary of Acronyms Used

GRAS……...…….Generally Recognized As Safe

FDA……...…..Federal Drug Administration

GI…….…….Glycemic Index

HDL……..……High Density Lipoprotein

LAI……..…….……L-Arabinose Isomerase

F6P or Fru6P…….……..Fructose-6-Phosphate

FBA……..…Fructose-1,6-Bisphosphate Aldolase

DHAP……..…Dihydroxyacetone Phosphate

T6P……..……...……Tagatose-6-Phosphate

XR……..……...Xylitol Reductase

GDH……..……...…Galactitol Dehydrogenase

G1P or Glu1P……..…….……Glucose-1-Phosphate

G6P or Glu6P……..…….……Glucose-6-Phosphate

Gal1P……..……...…Galactose-1-Phosphate

G6PI…….…Galactose-6-Phosphate Isomerase

TP……..…..Tagatose Phosphorylase

UDP……..Uridine-Diphosphate

G1PP……..…Galactose-1-Phosphate Phosphatase

UMP……..…..Uridine-Monophosphate

LC-MS……...……..Liquid Chromatography Mass Spectrometry

BAC……...…Bacterial Artificial Chromosome

PCR…….…...…Polymerase Chain Reaction

UPLC-RID…...…Ultra Performance Liquid Chromatography-Refractive Index Detection

LC-QQQ……...……Liquid Chromatography-Triple Quadrupole Mass Spectrometry

BLAST…….Basic Local Alignment Search Tool

ATP…….……Adenosine Triphosphate

ADP……...…..…Adenosine Diphosphate

PPi……..……...Inorganic Diphosphate

UTP……...……..Uridine-Triphosphate

Tagatose Background Information

D-tagatose is a naturally occurring, low-calorie sweetener that is an epimer of fructose, differing only in the position of the hydroxyl group on carbon 4 (**see Figure 1**).

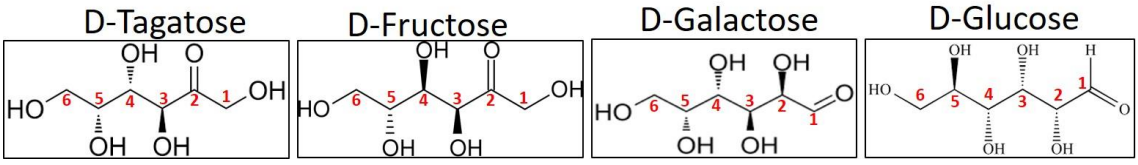


Figure 1. A Depiction of Different Sugar Molecules Related to D-Tagatose Production.

D-tagatose has 38% of the calories and 92% of the sweetness as common table sugar (sucrose) and is GRAS certified by the FDA, meaning that it is generally recognized as safe (Walsh, 2007). Tagatose is a rare sugar in nature but can be found at low concentrations in dairy products and various fruits including pineapples and oranges (Vera & Illanes, 2016). Because it is a natural sweetener with low caloric properties, tagatose is gaining interest as an alternative sweetener in consumer products. While it has not yet taken off as a widespread sweetener alternative, it can be found in a few consumer products on the market now such as 7-Eleven's Diet Pepsi Slurpee and a meal replacement shake called Therasweeet, produced by the company Living Fuel (*Tagatose*, n.d.).

Health Benefits to Using Tagatose as a Sweetener

There are numerous health benefits to using tagatose as an alternative sweetener other than its low caloric value. One of these benefits is its glycemic index of 3, which is much lower than sucrose's, which has a glycemic index of 65 (Guerrero-Wyss et al., 2018; The FDHA [@TheFDHA], 2018). For some other references, **Figure 2** depicts the glycemic indexes of some other commonly used sweeteners, both natural and unnatural.

Sweeteners	GI
Stevia (natural)	0
Fructose	15
Agave Syrup	15
Coconut Palm Sugar	35
Maple Syrup	54
Carmel	60
Honey	61
Sucrose	65
Splenda (artificial)	80
Glucose	100

Figure 2. Glycemic Indexes of Commonly Used Sweeteners.

(The FDHA [@TheFDHA], 2018). "GI" = Glycemic Index

Because tagatose is considered to have a very low glycemic index, it will minimally affect blood sugar, making it a safer and healthier sweetener for those with diabetes and other health issues.

Furthermore, in a 2015 phase 3 trial, tagatose's ability to treat hyperglycemia in type-2 diabetics was assessed. The results of this trial indicated that there was a statistically significant decrease in glycated hemoglobin, hba1c, in the trial patients when compared to the placebo control (Ensor et al., 2015). While the mechanism of action of how tagatose can lower high blood sugar levels is unknown, early research suggests that it has to do with tagatose blocking the uptake of glucose in the intestinal wall (Guerrero-Wyss et al., 2018). This effect of lowering blood sugar also has implications for using tagatose as a weight loss supplement.

In addition to tagatose's ability to lower blood sugar, there are many other benefits to using tagatose for human health. For example, in a 2010 study, subjects with

type 2 diabetes were subjected to a 12-month treatment using tagatose. These subjects were reported to have highly significant increases in high-density lipoprotein (HDL) cholesterol, which is also known as “good” cholesterol as it can scavenge your bloodstream for the “bad” types of cholesterol (Donner et al., 2010).

Secondly, in 2002, the FDA made a definitive ruling based on the data at hand about the nature of tagatose and tooth decay. Their ruling stated that tagatose cannot be fermented by the microorganisms living in your mouth to an extent that would break down tooth enamel. Because of this, they stated that tagatose could be advertised to not promote the risk of tooth decay (Dotzel, 2002).

Lastly, tagatose is gaining interest in the nutritional community due to its potential prebiotic properties. A prebiotic serves as an energy source for probiotics, in this case, the microflora inhabiting the gut. In a 1999 study performed in pigs, it was determined that tagatose is only 15-20% absorbed by the small intestine, while the rest of the intake is metabolized by the microflora living in the colon (Bertelsen et al., 1999).

Physical Properties of Tagatose

D-Tagatose is a hexose monosaccharide, meaning that it is a single unit sugar that has 6 carbons. Because it is a hexose, tagatose is comparable in molecular weight to many other common monosaccharides including glucose, galactose, and fructose (**see Figure 1**). Tagatose is a member of the ketohexose family, meaning that it has a ketone group on carbon 2 instead of a hydroxyl group (**see Figure 1**) (*14.2 Classes of Monosaccharides*, 2014). Additionally, both the D and L enantiomers of tagatose can be found in nature, however, the D form of tagatose is the one that is relevant as a sweetener

and can be found in milk and fruits. The L form of tagatose only exists as an uncommon microbial catabolite (Axelsen, 2017).

While tagatose is largely of different chemical composition than sucrose, it happens to be very physically similar both in its appearance (**see Figure 3**) and tactile properties. This makes it an excellent substitution for sucrose as it can be formulated in similar ways and yields similar tasting products with healthier nutritional properties.

Figure 3. A Depiction of Tagatose in its Crystallized Form.

*(*Synthesis of Tagatose_Chemicalbook, *2021)*

For example, like sucrose and most other sugars, tagatose has an extremely high solubility in water (1600g/L at 20°C), making it possible to be formulated into beverages and medicines (Fry, 2012). Additionally, it has excellent stability at acidic pH levels

found in most drinks. It is stable for up to 8 months in acidic drinks, although it is less stable in alkaline conditions (Bell, 2015). Tagatose is also stable at elevated temperatures, making it possible to pasteurize beverages containing tagatose. It has been shown that tagatose degrades a minimal 1% in the pasteurization process (Bell, 2015).

Lastly, because tagatose is a reducing sugar, it can react with amino acids and undergo the Maillard reaction. This reaction is relevant to human cuisine and is responsible for the browning of food when it is cooked. This is especially important in baked goods, where tagatose can be used to replace the usual sucrose. In fact, in a 2020 study, the replacement of sucrose in a cocoa bean biscuit food was proposed and proved to be possible. These types of innovations are important for individuals with diabetes and other health issues to be able to consume these types of products (Rojo-Poveda et al., 2020).

Current Strategies Being used for Tagatose Production

Tagatose has been a target of interest for both bioproduction and chemical production for years, first being described in 1984 (P. Kim, 2004). Many strategies have been explored to make tagatose. Most of these strategies involve using either a lactose or galactose starting material (**see Figure 4**.).

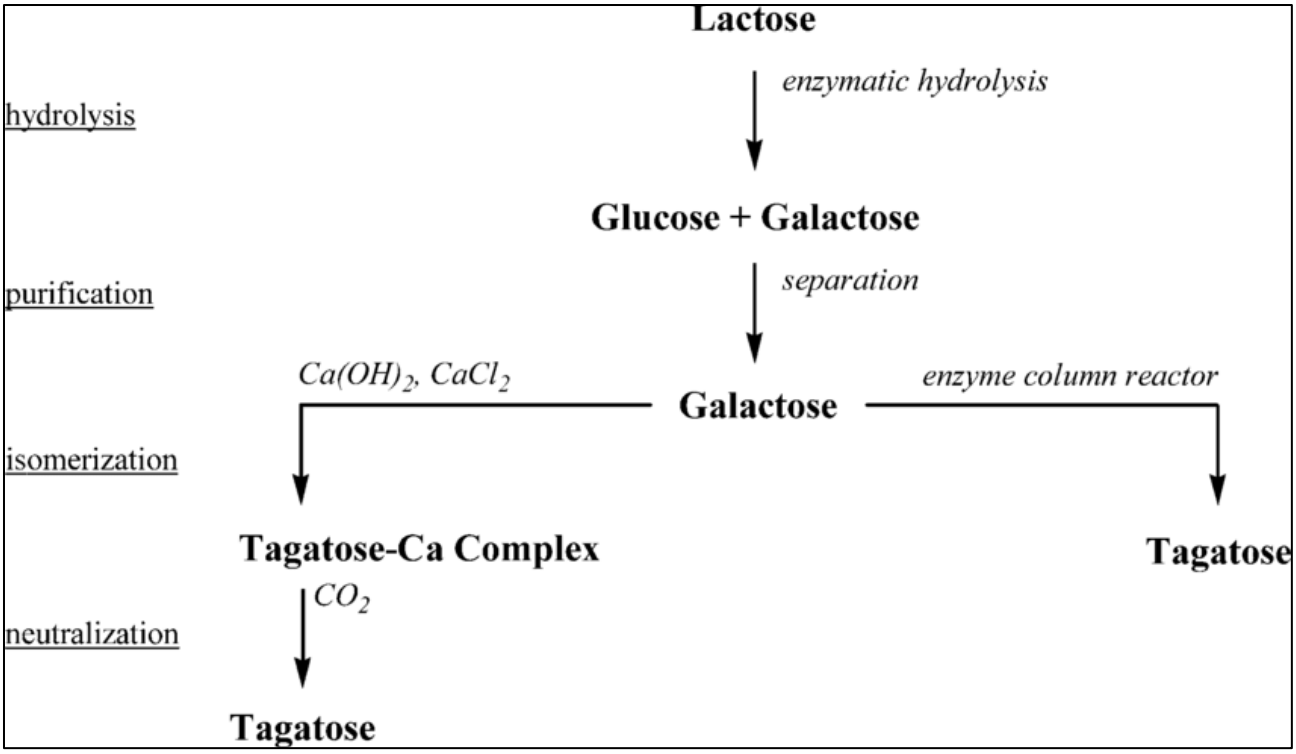


Figure 4. A Visual Summarization of the Main Strategies to Make Tagatose.

(P. Kim, 2004)

As shown above, the production schemes today use either an inorganic calcium catalyst to convert galactose to tagatose or biocatalysts, in this case, an immobilized enzyme catalyst (enzyme column reactor). Galactose can either be supplied directly into the system or separated from lactose, a disaccharide comprised of glucose and galactose (P. Kim, 2004). The chemical process of making tagatose from galactose is low yield and has been reported to give ~29% conversion of galactose to tagatose (Shintani, 2019). Additionally, it requires extensive downstream processing to purify the final tagatose product which leads to higher expenses.

Synthesis of Tagatose Using Enzymatic Biocatalysts

Contrary to the chemical synthesis route, tagatose can also be made using biocatalysts. This is shown above (**see Figure 4**) in the form of an immobilized enzyme column reactor, however, this is also done in various ways using recombinant enzymes and microbial hosts. There are 3 main methods for accomplishing this currently described in the literature, the fructose-1,6-bisphosphate aldolase method, the L-Arabinose Isomerase (LAI) method, and lastly, a galactitol dehydrogenase method.

Fructose-1,6-Bisphosphate Aldolase Method. Fructose-6-phosphate (F6P) is a key intermediate in glycolysis and is the product of an isomerization reaction from glucose-6-phosphate. Enzymes called fructose-1,6-bisphosphate aldolases (FBA), which natively convert fructose-1,6-bisphosphate into glyceraldehyde-3-phosphate and DHAP in glycolysis have been characterized to convert F6P into tagatose-6-phosphate (T6P). A 2017 study characterized an enzymatic application of converting fructose into tagatose using an FBA via a 3-step enzymatic process. This process is outlined below.

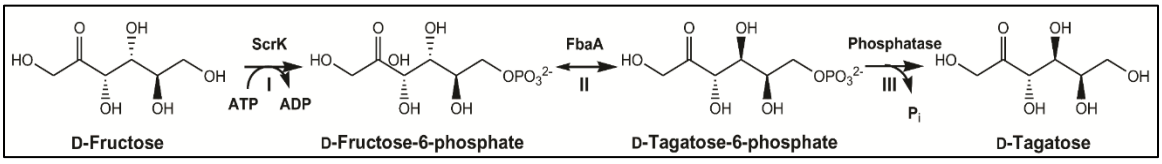


Figure 5. An Enzymatic Conversion Scheme from Fructose to Tagatose.

Lee et al. Demonstrated the conversion of fructose to tagatose using the above enzymes, ScrK, FbaA, and T6P(Phosphatase). Fruktokinase, Fructose-1,6-bisphosphate aldolase, and tagatose-6-phosphatase respectively.

The process was carried out *in vitro* with a ratio of these specific 3 purified enzymes in a reaction mixture. The group was able to get 77% of the fed fructose substrate converted to tagatose (Lee et al., 2017).

There have been other demonstrations of tagatose production using similar tactics as well. For example, in 2021 a group used maltodextrin as a feed source for purified recombinant *E.coli* cells expressing the necessary enzymes to convert the maltodextrin to F6P. They then used an FBA called GatZ to convert the F6P to T6P, which was then dephosphorylated to make the final tagatose product. The addition of this terminal phosphatase step is highly thermodynamically favorable as it releases a large amount of energy. Additionally, because the F6P to T6P step is in equilibrium, removing the T6P product by dephosphorylation drives the reaction forward in the desired direction. The group was able to make 3.2g/L tagatose from a 20g/L maltodextrin feed, achieving a 16% yield (Dai et al., 2021). This type of process is referred to as a "whole cell catalyst" process and is practically an *in vitro* process. The group speculates that low yields are due to the non-specific nature of the terminal tagatose-6-phosphatase. This causes the phosphatase to work on preceding intermediates in the pathway and divert carbon flux that could have otherwise been converted to tagatose (Dai et al., 2021).

While the FBA method works, it's difficult to demonstrate using *in vivo* biosynthesis as there is always competing fructose-1,6-bisphosphate in the cells which is the native substrate for the fructose-1,6-bisphosphate aldolases. The conversion rate of F6P to T6P using an FBA is also very slow by nature as this is not the normal function of the enzyme (Dai et al., 2021).

L-Arabinose Isomerase Method. Tagatose is very closely related to galactose as it is the same sugar structurally, with the exception of the functional ketone group. Tagatose could be described as the ketohexose form of the aldohexose galactose. Because of this close relationship, the L-Arabinose Isomerase (LAI) enzyme class has the ability to convert galactose to tagatose through a simple isomerization reaction.

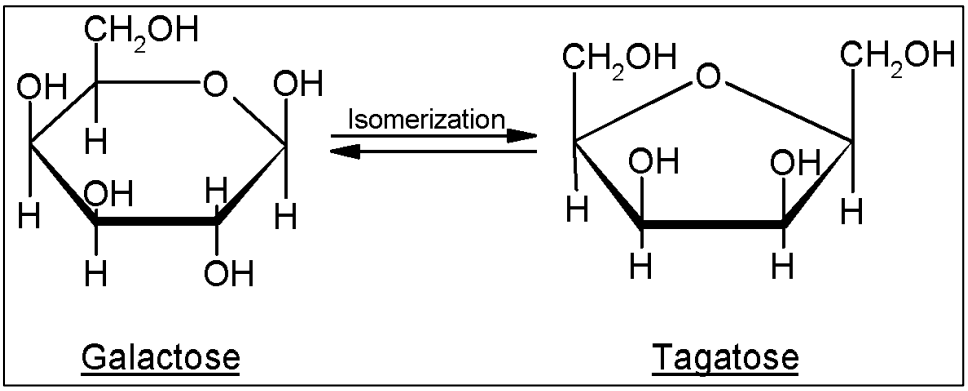


Figure 6. The Isomerization Reaction Between Galactose and Tagatose.

(Xu, 2011)

The isomerization of galactose to tagatose is the method that is currently used in the chemical *in vitro* production strategies today that were previously outlined above (**See Figure 4**). While these methods do work, they are inherently low yield processes as the reaction is in equilibrium and therefore theoretically cannot reach a conversion of over 50%. *In vitro* enzyme column reactor production aims circumvent this by recycling the unconverted galactose back into the enzyme column reactor. Bioproduction has also worked around this in several different ways.

For example, in 2019, a group published their work on another whole cell catalyst process converting galactose to tagatose using a LAI from *Lactobacillus sakei*. The group showed that by feeding galactose to a suspension of *Lactobacillus plantarum* cells that are expressing the LAI the conversion rates of galactose to tagatose can be increased, presumably because of the differential partitioning of galactose and tagatose across the cell membrane. This differential partitioning influences the equilibrium of the isomerization reaction and assists in driving the reaction toward the tagatose product. *Ls*LAI showed conversion rates of ~39% at 37° purified on its own but then showed conversion rates of ~47% when expressed in the *Lactobacillus plantarum* cells in a PBS buffer suspension at 37°. (Bober & Nair, 2019). Additionally, the group showed that the LAI favors tagatose production at higher temperatures and reported ~80% conversion when expressed in a PBS buffer suspension of *Lactobacillus plantarum* cells at 50°C. This elevation in temperature would normally affect the enzyme negatively. However, when expressed within the *Lactobacillus plantarum* cell wall the enzyme appeared to be more thermostable, as the enzyme purified by itself gave only a 16% conversion at 50°C (Bober & Nair, 2019)

Expanding on this concept, a group published in 2008 on a similar phenomenon. This group used araA, the native *E.coli* LAI, to convert galactose to tagatose in a whole cell catalyst process. They reported that araA was able to convert 36% of the fed galactose to tagatose when using the purified enzyme, but its conversion percentage increased to 68% when expressed in purified *E.coli* cells. To investigate why the isomerization equilibrium may be influenced by cell presence, they performed sequential knockouts of various transporters in *E.coli* to investigate if the tagatose product may be

exported out of the cell, allowing the equilibrium to shift toward tagatose. They found that when they knocked out the mglB subunit of the mglBAC transport system, the tagatose production went down by ~60% (J.-H. Kim et al., 2008). They concluded that the mglBAC transporter may not only be importing galactose into the cell, but also exporting the converted tagatose from the cell. This would create a natural partition between tagatose and galactose and allow the isomerization to favor the direction of tagatose. This concept is outlined below.

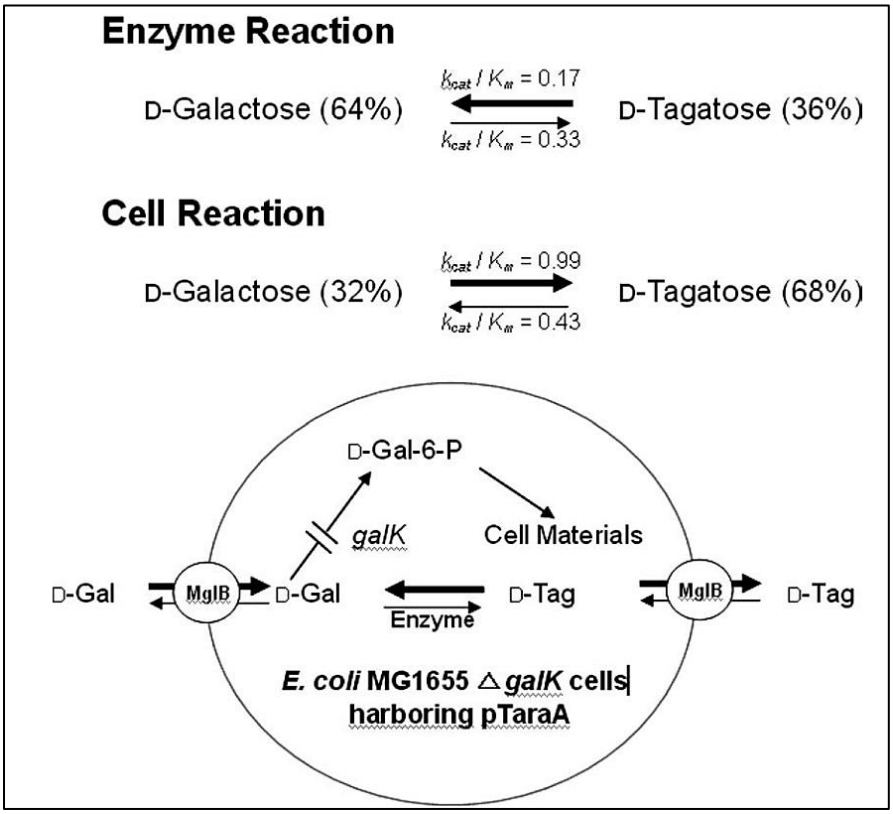


Figure 7. The Implicated Role of mglB in Tagatose Export in *E.coli*. *(J.-H. Kim et al., 2008)*

Galactitol Dehydrogenase Method. Another method of tagatose production in the literature circumvents the equilibrium limitations of the LAI isomerization in a different way. This *in vivo* method uses *Saccharomyces cerevisiae* expressing a xylitol reductase (XR) to convert galactose into galactitol and then a galactitol dehydrogenase (GDH) to convert the galactitol to tagatose. The original galactose comes from a lactose feed which is broken down into glucose and galactose by the cell. The cell then uses the glucose for cell growth and the galactose is converted over to tagatose. The galactose is not able to be metabolized by the cell because the conversion is carried out in a yeast strain that has the galactokinase knocked out. This process is outlined below.

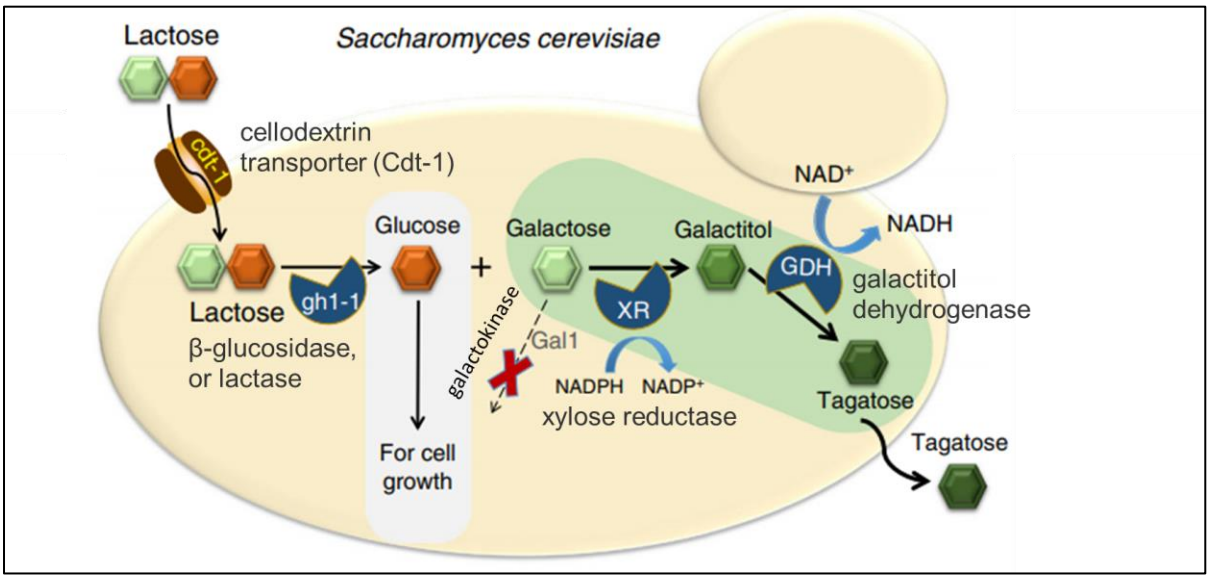


Figure 8. Production of Tagatose Using a Redox Potential Approach in *Saccharomyces cerevisiae*.

(Liu et al., 2019).

By using a redox coupled system, the Liu et al. group does not have an isomerization reaction to consider and was able to achieve ~37.7g/L of tagatose from ~114g/L of lactose feed, giving them ~33% yield (Liu et al., 2019).

Pitfalls to the Current Enzymatic Catalysis Methods

While some of the previously described methods for tagatose biosynthesis seem to be promising candidates for large-scale production of tagatose they all have their pitfalls. Firstly, many of these processes have a low yields. This includes the chemical production of tagatose, the method of using maltodextrin as a feed source to make F6P and then converting it to tagatose via the FBA, GatZ, and lastly, the galactitol dehydrogenase method. These processes give tagatose yields of 29%, 16%, and 33% respectively.

Secondly, most of these processes previously described are *in vitro* processes, with the exception of the galactitol dehydrogenase method carried out in *Saccharomyces cerevisiae*. Because of their *in vitro* nature, these processes are long and labor intensive and include many steps. These include the growth of recombinant cells and then subsequent induction, harvest, and purification of the desired enzymes for the *in vitro* enzymatic conversions. Additionally, for the whole cell biocatalyst processes, steps include the growth of the recombinant *E.coli* cells, the purification of those cells out of their growth mediums, and then their subsequent resuspension into the working buffer before the substrate can be added for conversion. Methods like this, although they have the potential to produce high yields, are difficult to use in large-scale production as they require large purification steps and increased material handling, which leads to higher costs and larger risks of contamination.

Lastly, one of the main issues with these processes is their use of more expensive carbon sources. Most of these processes are either using lactose or galactose with the exceptions of the processes using fructose and maltodextrin. The problem with these substrates is their cost, which is outlined below.

Bulk sweeteners	£/kg
Sucrose	0·44
Glucose	0·26
Fructose	0·70
Lactose	0·35
Sorbitol	0·96
Mannitol	1·40–1·57
Xylitol	2·36–2·62
Maltitol	1·40
Isomalt	1·40
Lactitol	1·31

Figure 9. A Cost Comparison of Common Sugars.

Sugar arranged from lowest to highest priced: Glucose<Lactose<Sucrose<Fructose. (Zumbé et al., 2001) While galactose is not listed here, it is usually produced from lactose commercially, putting it at a price point slightly higher than lactose due to production costs (Shintani, 2019).

Using feed sources other than glucose significantly increases the cost of the production process. If lactose is used, the raw material cost is increased by ~35%. Additionally, methods using lactose as a source of galactose for an LAI are already limited to a 50% yield, as lactose is a disaccharide of glucose and galactose. If fructose is used as a raw material over glucose, the raw material cost goes up by ~169%. Therefore, it is much more economical to use glucose as the carbon source in a production process.

Tagatose Production Strategies Used in this Study

Due of the pitfalls that currently surround the tagatose production strategies previously described in the literature, this study aims to develop a novel, fully *in vivo* pathway from glucose to tagatose using *E.coli* as the host. Two routes by which to accomplish this are proposed and evaluated by experimentation in this study.

Route 1: The Glucose-1-Phosphate Epimerase Route

This proposed route involves 6 steps total from glucose to get to tagatose and also proposes 2 novel reactions that are currently unknown in the literature. These novel reactions include the epimerization of glucose-1-phosphate (Glucose-1-P) to galactose-1-phosphate (galactose-1-P) and a subsequent mutase reaction from galactose-1-phosphate to galactose-6-phosphate (galactose-6-P). The route would then implement a galactose-6-phosphate isomerase (G6PI) and a tagatose phosphorylase (TP) to complete the conversion to tagatose. Both the G6PI and TP enzymes have been previously reported in the literature to perform these reactions. Route 1 is outlined in detail below.

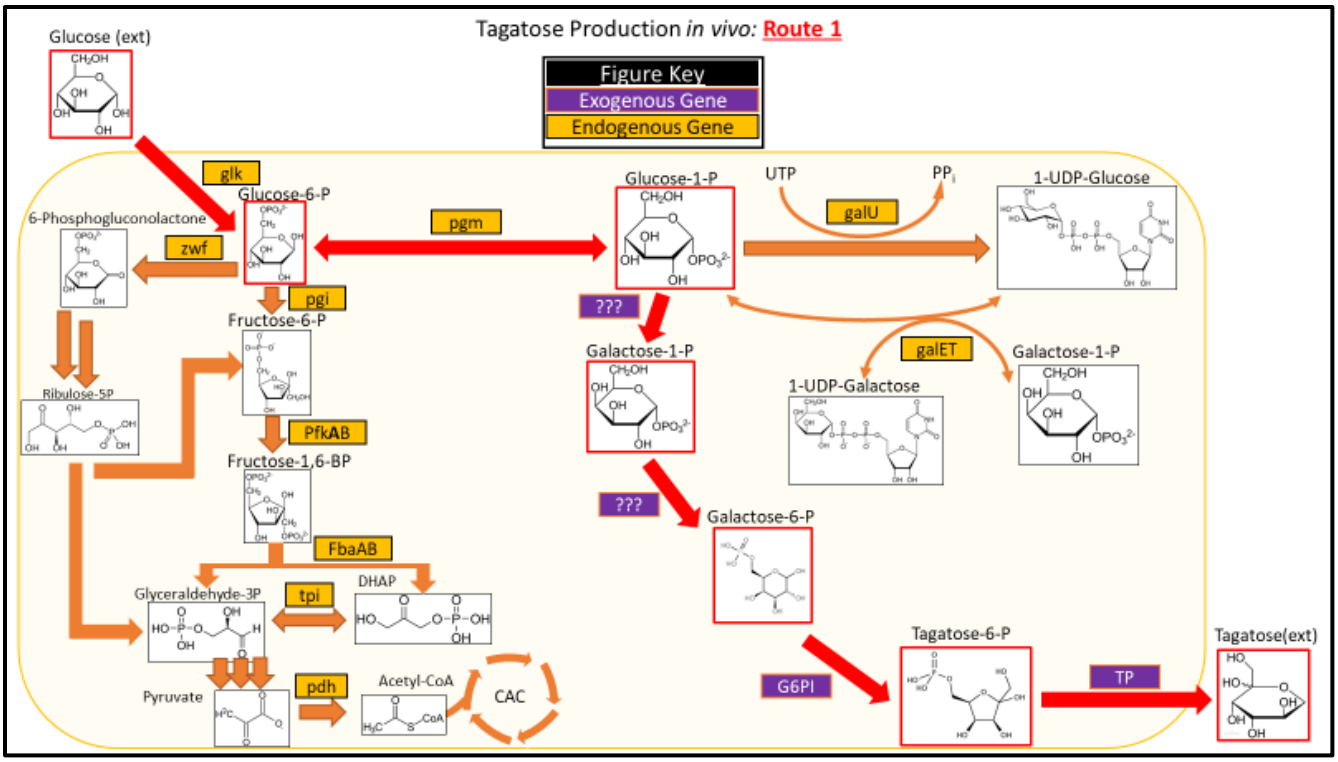


Figure 10. Reactions Required to Make Tagatose from Glucose Using Route 1.

The reactions for route 1 are shown in red arrows and the intermediates are outlined in red. Enzymes that perform the reactions are abbreviated in boxes next to the reaction arrow. The enzymes in the pathway for route 1 are as follows: Glucokinase(glk)>Phosphoglucomutase(pgm)>Unknown epimerase(???)>Unknown Mutase(???)>Galactose-6-Phosphate Isomerase(G6PI)>Tagatose Phosphorylase(TP) All other reactions, intermediates, and enzymes depicted are important reactions that are endogenous to E.coli. These reactions either contribute to cell growth or directly interact with the proposed pathway.

Because there were 2 novel reactions to characterize for route 1, this route posed as a larger challenge. However, both of these novel reactions had both literature and biological based precedence.

The first of these precedented reactions is epimerization of UDP-glucose to UDP-galactose by the *E.coli* galE enzyme. It was hypothesized that galE may be able to

perform the glucose-1-phosphate to galactose-1-phosphate reaction as the UDP group is attached to the 1 carbon on glucose, making UDP-glucose a similar substrate to glucose-1-phosphate in terms of its structure, see below.

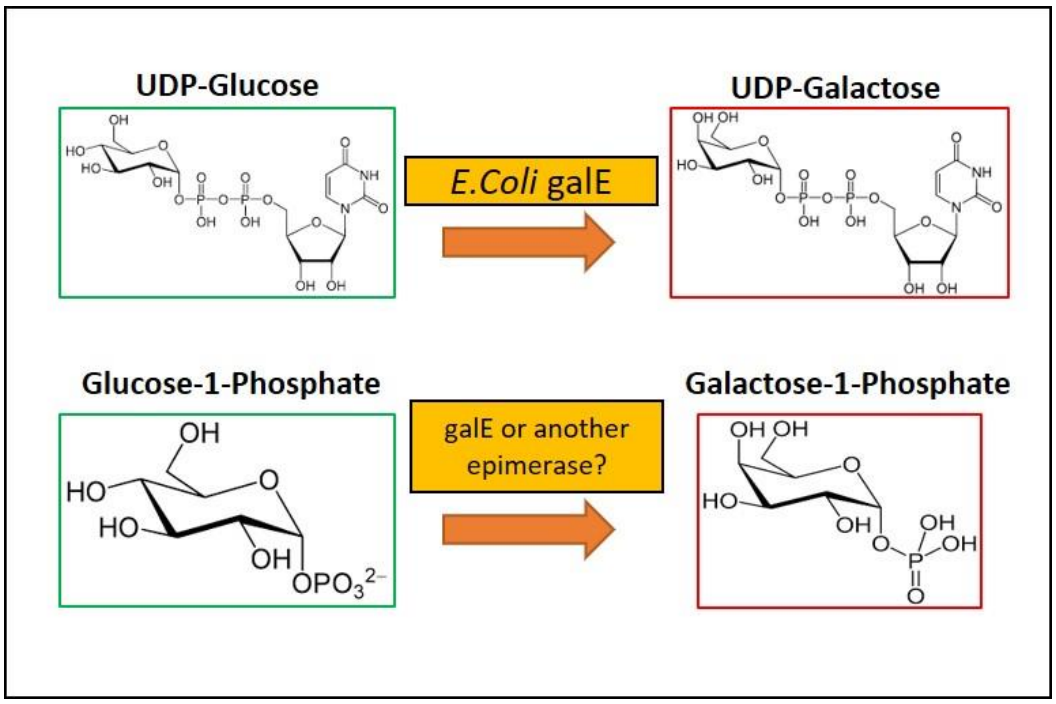


Figure 11. GalE Epimerization and UDP-Glucose Compared to Glucose-1-Phosphate

Glucose containing compounds are highlighted in green and galactose containing compounds are highlighted in red.

Furthermore, the reaction mechanism for galE characterized by the Y, Nam et al. group shows that the isomerization of UDP-glucose to UDP-galactose is performed independently of the UDP group (Nam et al., 2019). This adds confidence to the ability of galE or another epimerase to be able to do the glucose-1-phosphate to galactose-1-phosphate epimerization.

In addition to using galE to test for the epimerization reaction, homologs of galE that could potentially do the reaction as well were screened for their ability to epimerize glucose-1-P. There is an extensive tree of galE homologs highlighted in Y, Nam *et al.* publication that were all potential targets (Nam et al., 2019). A select number of epimerases were tested from this tree. Specifics on the rationale for choosing them is in the following materials and methods chapter.

The second of these 2 novel reactions for route 1 is the mutase reaction, which is analogous to the conversion of glucose-6-phosphate to glucose-1-phosphate by the *E.coli* pgm enzyme. Due to pgm performing this reaction in both directions, it was predicted that enzymes of this class may be able to convert galactose-1-phosphate to galactose-6-phosphate. Because these types of reactions already existed in *E.coli*, there was some confidence that it would be possible to find enzymes that could do these novel reactions, galE and pgm being prime candidates to begin with.

Route 2: The Galactose-1-Phosphate Phosphatase Route

This route aimed to make galactose intracellularly from glucose and then isomerize it to tagatose using the already well-studied LAI technology. Route 2 aimed to accomplish this by harnessing the already established Leloir pathway in *E.coli* which is normally activated to convert galactose to glucose for cellular energy. The Leloir pathway is shown in detail in Appendix 2, **Figure 30**. This route involves 5 steps total from glucose which are outlined below in **Figure 12**, and proposes only 1 novel reaction that was not previously known in the literature. This reaction involves a phosphatase with the ability to selectively dephosphorylate galactose-1-phosphate to make galactose. It

also involved characterizing an LAI that would be well expressed and well functional in *E.coli*. These LAIs had already been reported in the literature and some were discussed in the previous tagatose production strategies. This route is outlined in detail below.

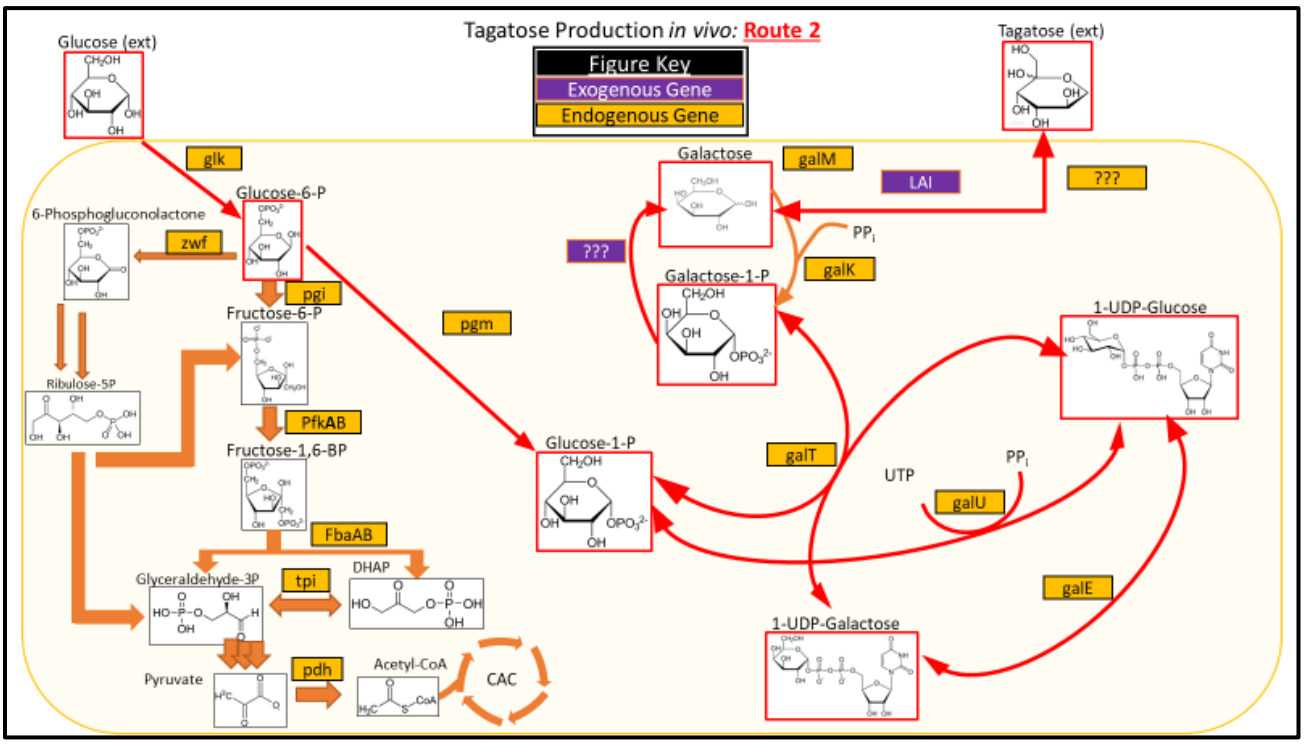


Figure 12. Reactions Required to Make Tagatose from Glucose Using Route 2.

The reactions for Route 2 are shown in red arrows and the intermediates are outlined in red. Enzymes that perform the reactions are abbreviated in boxes next to the reaction arrow. The enzymes in the pathway for Route 1 are as follows: Glucokinase(glk)>Phosphoglucomutase(pgm)> (galactose-1-phosphate uridylyltransferase(galT)>Unknown Galactose-1-Phosphate Phosphatase(???)>Unknown L-Arabinose Isomerase(LAI). There are also some additional reactions highlighted in red that are direct supporting reactions for galT. These reactions are as follows: UTP-glucose-1-phosphate uridylyltransferase(galU)> UDP-glucose 4-epimerase(galE). All other reactions, intermediates, and enzymes depicted are important reactions that are endogenous to E.coli. These reactions either contribute to cell growth or directly interact with the proposed pathway.

The largest challenge route 2 posed was finding a galactose-1-phosphate phosphatase (G1PP) that would be specific enough to galactose-1-phosphate. If a phosphatase were to be overexpressed that was non-specific, it could also begin to dephosphorylate other important intermediates in the pathway such as Glu6P and Glu1P. This would result in a low tagatose yield, or other unwanted metabolic dysfunction, similar to the issues that the Dai. group had seen with their terminal tagatose phosphatase (Dai et al., 2021).

Chapter II:

Materials and Methods

This chapter contains information on the methodology and materials used to perform the experiments for this project, as well as an explanation of naming conventions used for the results and discussion chapters.

Enzyme Discovery

Enzyme Selection

16 epimerases, 9 phosphatases, and 2 LAIs were selected to be cloned and tested for this project for a total of 27 enzymes. Epimerases were all selected from a specific publication from the Nam, et al. group. This paper analyzes and predicts the UDP-galactose epimerase mechanism from *Bifidobacterium longum* and proposes 40 other epimerases and their predicted substrate specificities (Nam et al., 2019). Enzymes were chosen from this paper if they were non-pathogenic and not from animals to avoid issues with food product regulations. One of the selected epimerases was the native *E.coli* epimerase, galE to be used as a control.

Phosphatases were chosen from a search on UniPort.org for “galactose-1-phosphate phosphatase.” Enzymes were selected for the project as long as they were non-pathogenic, not from animals, and had reviewed UniProtKB entries. Some of these entries documented activities on certain substrates that were backed up by published data.

Other entries documented potential activities that the enzymes may perform based off their sequence similarity to other experimentally verified entries.

Lastly, only 2 LAIs were selected for this project as it has been reported in the literature that most of them work for the isomerization of galactose to tagatose. For this project, the LAI from *Lactobacillus sakei* was selected, which was used in the Bober & Nair study, previously described in the introduction chapter. The second LAI was also selected from a publication that showed the *Bacillus coagulans NL01* LAI could achieve conversion rates of galactose to tagatose of up to 40.8% in a galactose bioconversion study. (Mei et al., 2016).

Cloning for Enzyme Discovery

All enzyme sequences were taken from Uniprot.org and recoded for *E.coli* expression using the codon optimization tool from Azenta Life Sciences (Boston, MA). Sequences were ordered from Twist Biosciences (San Francisco, CA) and were designed to be cloned directly into a pET-28a(+) vectors using an NcoI and HindIII double digest (New England Biolabs Ipswich, MA). Vectors were cloned using *E. coli* (Dh5α) cells and sequence verified. **See Appendix 1** for a full list of the 27 enzymes cloned for this project. **See Appendix 4** for a more extensive list of all enzymes and genomic parts used in this project, including all sources and sequences.

Cell Culture and Recombinant Enzyme Harvest

Sequence verified pET-28a(+) vectors as well as one additional control "blank" plasmid, which contained no cloned enzyme, were expressed in *E.coli* BL21(DE3) cells. Cells were grown up in 50mL tubes with 10mL Lauria Bertani (LB) medium + 50ug/mL

kanamycin at 37°C, 250RPM for ~16 hrs. The following day, the cultures were subcultured to OD=.05 in 50mL tubes with 10mL LB + 50ug/mL kanamycin and grown at 37°C for 2hrs. Cultures were then induced with 20uL of 0.5M IPTG (1 mM final concentration) and left to grow at 37°C, 250RPM, for an additional 4 hours to allow for protein expression.

All steps after the cell culture previously described were performed on ice, chilled metal beads, or in centrifuges cooled to 4°C to prevent the enzymes of interest from crashing out of solution. Cells were cold shocked on ice for ~20min and then spun down for 10min at 4000RPM, the supernatant was then decanted and the cell pellets were snap frozen in liquid nitrogen and stored at -80°C overnight. Cell pellets were thawed on ice the following day and then resuspended in 5mL of 0.25M, pH=7 HEPES buffer in the same 50mL tube. Samples were then sonicated with a Branson Sonifier 250 (Danbury, CT). Used 5 seconds on 20 seconds off cycles for 24 cycles, with a 40% amplitude protocol giving 2min total of sonication on chilled, ice cold metal beads. Protein extracts were clarified by spinning down samples at 20,000rcf for 10min. 1mL aliquots were taken in microcentrifuge tubes and stored at -20°C. Cell free extracts were then evaluated for protein expression using a total protein Bradford assay and an SDS-PAGE gel. Enzymes 2, 3, 5, 8, 17, 23, and 25 were removed from the project due to their lack of expression in *E.coli*. **See Appendix 1, Figure 29** for images of these gel results.

In Vitro Enzyme Screening Assays

Enzymes in this project were screened using the cell free extracts from the previous section with the desired reaction substrate. Both the phosphatases and the epimerases were screened on 10mM galactose-1-phosphate with the desired products

being galactose for the phosphatases and glucose-1-phosphate for the epimerases. The epimerases were also controlled with an additional native substrate UDP-galactose feed reaction to confirm that the enzymes were indeed active. Additionally, the epimerases were tested with and without the addition of 10mM UMP (uridine monophosphate) alongside the 10mM galactose-1-phosphate. The rationale for this being that because the native substrate for the epimerases is UDP-galactose (uridine diphosphate galactose), the addition of UMP would allow for enzyme function by binding the UMP and the desired galactose-1-phosphate substrate. This would allow it to have the 2 phosphate groups available, such as when binding UDP galactose. The L-Arabinose Isomerases were tested on galactose with the desired product being tagatose. Reactions were carried out in PCR tubes in 50uL final volume HEPES buffer (pH=7) suspensions. These reactions were set up with 30-60ug/mL of total protein so that all total protein concentration was normalized from sample to sample. Phosphatase reactions were supplemented with 1mM $MgSO_4$, the epimerase and LAI reactions did not contain any cofactors. All in vitro enzyme screens were controlled with a "biological blank" sample which consisted of the cell free extract from the empty pET28a(+) vector to control for background enzymatic activity. All reaction components were added first and then the substrate was added at time 0 and the plate was allowed to incubate at 37°C for 8-22 hours. The reaction was stopped with the addition of 150uL LC-MS grade acetonitrile. Reactions were then assessed for their effectiveness using qualitative mass spectrometry which is discussed in a following analytical instrumentation section of this chapter.

In Vivo Fermentation Assay Enzyme Screening

Once the lead enzymes were characterized using the *in vitro* enzyme screening assays they were subcloned into bacterial artificial chromosome (BAC) plasmids and/or integrated into the genome to be tested with an *in vivo* fermentation.

BAC Cloning and Strain Engineering

BAC cloning was done by PCR amplification out of the pET vectors followed by an AarI restriction digest and subsequent ligation using a type IIs Golden Gate assembly tactic. The BACs were then confirmed to be sequence correct by sanger sequencing. Integrations of lead enzymes and supporting chassis modifications were performed using the classical sacB-kanamycin positive/negative selection method to achieve clean, scarless knock ins or knock outs of the genes of interest. Enzyme screens were then carried out in 96 well 2mL polypropylene deep well plates (96DWP) in 400uL working volumes. These i*n vivo* fermentation screens consisted of a seed culture, a production culture, and an extraction.

Seed Culture

BAC plasmids of interest were transformed into the relevant electrocompetent *E.coli* strains using a BioRad Genepulser Xcell electroporator (Hercules, CA) at 2.5KV and then recovered in SOC media for 1hour. The SOC recoveries were then plated onto LB Agar plates containing chloramphenicol as a selection antibiotic. The plates were grown overnight in a stationary incubator at 37°C. The following day, colonies from the transformations were seeded into the 96DWPs containing 400uL of fermentation growth media (proprietary to Manus Bio) with either galactose or glucose, depending on the enzyme being tested. The wells were seeded in triplicate with a single transformation

colony used to inoculate each well. Seed cultures were incubated at 37°C in a shaking incubator at 250RPM for 24hrs.

Production Culture

Production cultures were inoculated from the 24 hour seed cultures by adding 40µL of the seed culture to 360µL of fresh fermentation growth media in a 96DWP. Production cultures were incubated at 37°C in a shaking incubator at 250RPM for 48hrs.

Fermentation Extraction

Liquid-liquid extractions were done with the purpose of extracting out reaction products and intermediates from the cell cultures into a matrix that could be analyzed by the applicable analytical instrumentation. The extraction was performed by adding 400uL of 100% acetonitrile, pulse vortexing for 10 min and centrifuging for 10min. Supernatants from the spun down plate were diluted an additional 20-50X depending on how much glucose or galactose was fed. 75% acetonitrile was used for LC-QQQ dilutions and pure water was used for UPLC-RID dilutions. Final dilutions were run on the LC-QQQ for tagatose quantification and on the UPLC-RID for glucose and galactose quantification.

Table 1. Definitions of Genomic Parts and Loci.

Part/loci	Abbreviation Used	Part Type	Description
pCC1BAC	1BAC	Plasmid Backbone	Single copy plasmid backbone
MP3	3	Promoter	A Manus Bio promoter, the weakest promoter used in this project
MP6	6	Promoter	A Manus Bio promoter, stronger than MP3 but weaker than FAB46
FAB46	46	Promoter	A very strong promoter, the strongest promoter used in this project
BCD18	18	Ribosome Binding Site	A "bicistronic design" ribosome binding sequence, the weakest one used in this project.
BCD14	14	Ribosome Binding Site	A "bicistronic design" ribosome binding sequence, stronger than BCD18 but weaker than BCD14.
BCD7	7	Ribosome Binding Site	A "bicistronic design" ribosome binding sequence, the strongest one used in this project.
Intergenic region between atpI and rsmG	igA	*E.coli* Genomic Locus	A chromosomal locus used in this project to integrate genes of interest into.
Intergenic region between lafU and dinB	lafU-dinB	*E.coli* Genomic Locus	A chromosomal locus used in this project to integrate genes of interest into.

"Bicistronic design" Ribosome Binding Sequences were previously characterized by the Mutalik et al group and unlike standard ribosome binding sites, have no dependence on the downstream genetic elements they control. The FAB46 promoter was also characterized by the Mutalik et al group using a modular promoter library screen. (Mutalik et al., 2013). MP, "Manus Promoters" are proprietary to Manus Bio. Relative strengths of promoters in order of increasing strength: MP3>MP6>FAB46. Relative strengths of BCDs in order of increasing strength: BCD18>BCD14>BCD7.

Promoters and BCDs are often referred to by their abbreviations, separated by hyphens. For example, 1BAC-46-7-*B.co*LAI = pCC1BAC-FAB46-BCD7-*B.co*LAI = *B.co*LAI cloned into a pCC1BAC plasmid under control of the FAB46 promoter and the BCD7 ribosome binding site. Plasmids complimented into strains are noted by a "+" sign. For example, s171 + 1BAC-46-7-*B.co*LAI = the pCC1BAC-FAB46-BCD7-*B.co*LAI plasmid complemented into the s171 strain.

Integrations into the genome are noted by the strain name followed by a "Δ" symbol to indicate which strain and at which locus was the destination of the gene of interest. A double colon then indicates the following genetic material was integrated. For example: s171ΔigA::46-7-*B.co*LAI = 46-7-*B.co*LAI was integrated into the s171 strain at the igA chromosomal locus. Deletions are simply noted by a "Δ" symbol with no double colon following it. For example: s100ΔgalKM = galKM was deleted in the s100 strain.

Analytical Instrumentation Methods

There were 2 analytical instruments used in this project. These were quantitative ultra-performance liquid chromatography coupled with triple quadrupole mass spectrometry (LC-QQQ, Agilent 6460 Triple quad Mass Spectrometer) and ultra-performance liquid chromatography coupled with refractive index detection (UPLC-RID, Agilent 1260 Infinity II Refractive Index Detector) Each one of these instruments was used in different ways to assess the various products of interest throughout this project.

Analytics for Epimerase Discovery Work

- Reactions to be Observed:
    - UDP-Galactose -> UDP-Glucose
    - Galactose-1-phosphate -> Glucose-1-phosphate
- Method Details:
    - Buffer A: 100mM Ammonium Acetate pH 9.5
    - Buffer B: 100% Acetonitrile
    - Instrument: LC-QQQ
    - Column: ThermoFisher Scientific Accucore-150mm Amide HILIC (16726-152130)
    - Run Time = 81 min

Analytics for Phosphatase and LAI Discovery Work

- Reactions to be Observed:
    - Desired Reaction for phosphatases: Galactose-1-P -> Galactose
    - Desired Reaction for LAI: D-Galactose -> D-Tagatose
- Method Details:
    - Buffer A: 100mM Ammonium Bicarbonate
    - Buffer B: 100% Acetonitrile
    - Instrument: LC-QQQ
    - Column: Milipore SeQuant ZIC-pHILIC 150x4.6 mm (1.50461.0001)
    - Run Time = 31 min

Analytics for Phosphatase and LAI *In Vivo* work

After the initial discovery work was completed the chosen LAI and G1PP were cloned into BAC plasmids and tested using an *in vivo* fermentation assay as described in the previous chapter section. Analytics for these used a modified, shorter method, derived from the one used to screen the previous *in vitro* phosphatase reactions. The analytical methods for these screens are as follows:

- Compounds Observed (Instrument used to analyze the compound):
    - D-Tagatose (LC-QQQ)
    - D-Galactose (UPLC-RID)
    - D-Glucose (UPLC-RID)
- Method Details for the LC-QQQ Tagatose Detection:
    - Buffer A: 100mM Ammonium Bicarbonate
    - Buffer B: 100% Acetonitrile
    - Instrument: LC-QQQ
    - Column: Milipore SeQuant ZIC-pHILIC 150x4.6 mm (1.50461.0001)
    - Run Time = 18 min
- Method Details for the UPLC-RID Glucose and Galactose Detection:
    - Buffer A: 100% Water
    - Buffer B: N/A
    - Instrument: UPLC-RID
    - Column: Rezex RCM-Monosaccharide Ca+, 100x7.8mm. Phenomenex (00D-0130-K0), controlled at 80°C
    - Run Time = 15 min

Chapter III.

Results

The results of the project are heavily weighted towards data from route 2 (the phosphatase route) as it proved to be the more viable route. However, chapter 3 begins with the preliminary data from the *in vitro* epimerase enzyme screening assays for route 1 and describes the reasons why the route was ultimately dropped to pursuit route 2 instead.

Route 1 Experimental Results

The epimerases were tested using an *in vitro* enzyme screening assay, described in the previous chapter, where galactose-1-phospahte was fed to a cell free extract of the epimerases and screened for conversion to glucose-1-phosphate. None of the epimerases displayed any conversion from galactose-1-phosphate to glucose-1-phosphate (**see Figure 13a**). This reaction was screened in the reverse direction that it was intended to function *in vivo* due to limitations in analytical method used. The method had some peak tailing associated with it so it would have been easier to see the earlier eluting glucose-1-phosphate appear after conversion from galactose-1-phosphate (**see Figure 13**). All of the epimerases tested were active on the control UDP-galactose substrate, with the exception of the *Lactococcus lactis IL1403* epimerase which is shown as a blue line in **Figure 13b**.

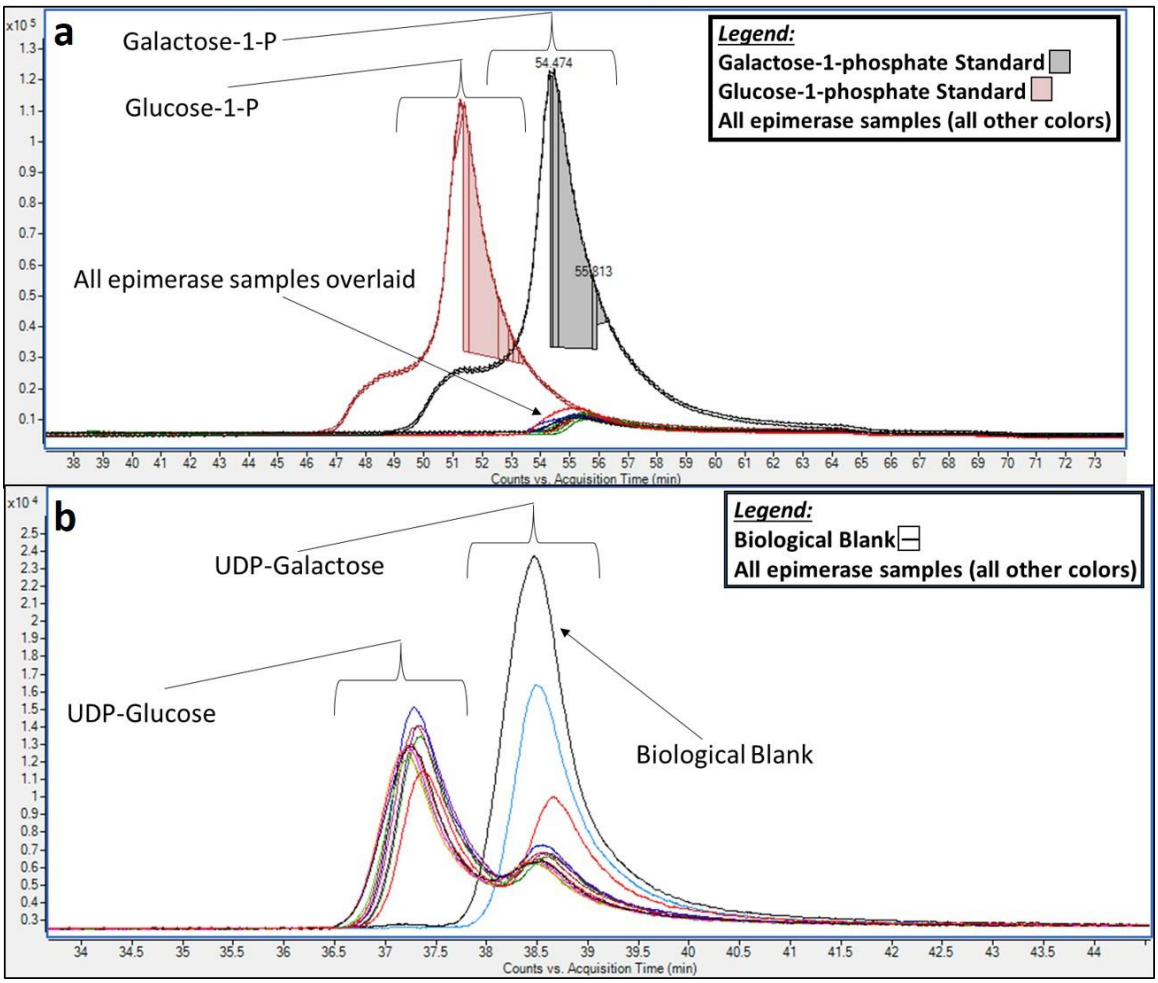


Figure 13. Total Ion Count Chromatograms of the Epimerase *In Vitro* Enzyme Assays.

*Total ion counts for all 16 of the epimerase reactions that were collected from LC-QQQ single ion monitoring scans of the reactants and products are depicted here. (**a**)The galactose-1-phosphate reactant standard is shown in grey. The glucose-1-phosphate desired product standard is shown in burgundy. All the epimerase samples in addition to the biological blank sample are overlaid together to demonstrate their collective lack of activity on galactose-1-P. The previously observed retention time windows for galactose-1-phosphate and glucose-1-phosphate are outlined with labeled black brackets. (**b**) The control reactions for the epimerases. The biological blank control is shown as a black line. The previously observed retention time windows for the UDP-galactose reactant and the expected UDP Glucose product are outlined with labeled black brackets. All other lines are results of the individual epimerase reactions on UDP-galactose.*

Figure 13 displays results that confirm the epimerases were expressed and active in the cell free extracts, this is seen in the conversion of UDP-galactose to UDP-glucose. However, they were not active on the desired galactose-1-phosphate substrate. Although the data is not shown here this same result was seen when the reaction was supplemented with 10mM UMP as well. After this result was collected for the epimerases, route 1 was removed as a viable route to tagatose for this project and all efforts were refocused onto route 2.

Route 2 *In Vitro* Discovery Experimental Results

Using route 2 to make tagatose *in vivo* from glucose was the primary focus of this research project. This experimental data is broken down into the *in vitro* discovery results followed by the *in vivo* experimentation results

Initial Enzyme Discovery Data

The cloned LAIs and G1PPs for the project were first vetted for their ability to do the desired reactions using an *in vitro* enzyme screening assay. The LAI was fed galactose and screened for tagatose production, and the G1PPs were fed galactose-1-phosphate and screened for galactose production. Only 2 LAIs were selected for the project. One of them was removed from consideration after the recombinant protein harvest step as it did not express well in *E.coli*. Because of this, only the *Bacillus coagulans* NL01 LAI (*B.co*LAI) was tested for its ability to convert galactose to tagatose, **see Figure 14**.

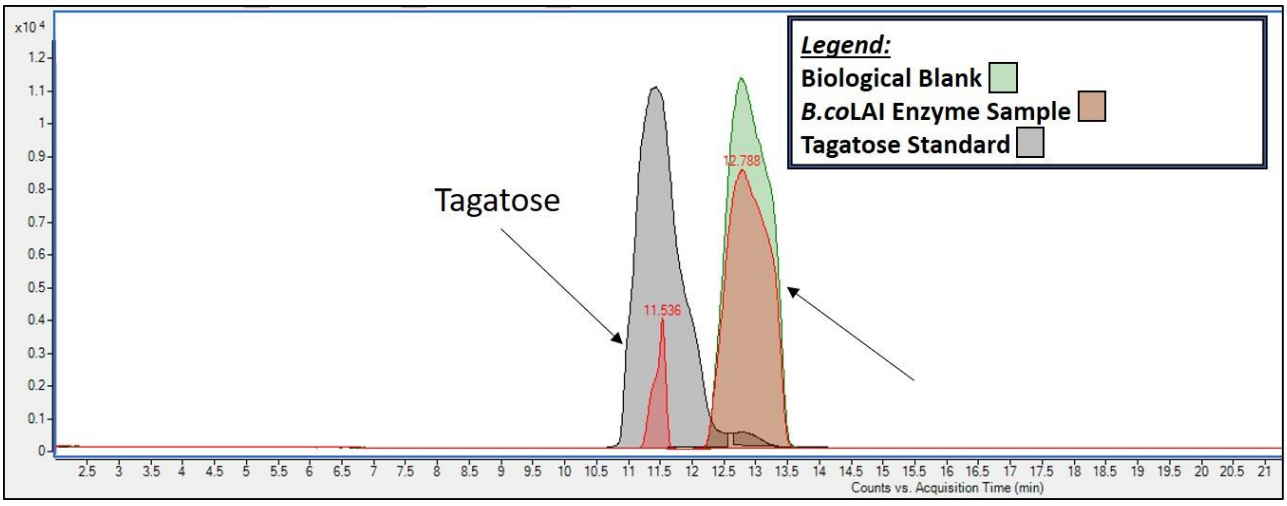


Figure 14. LC-QQQ Single Ion Monitoring Scans from the *In Vitro B.co*LAI Screen

Reactants (galactose) and products (tagatose) are pictured here for the Bacillus coagulans NL01 LAI in vitro enzyme screen. 3 samples total are shown: The biological blank sample in green, A 12.5mM tagatose standard in grey, and the Bacillus coagulans NL01 in vitro enzyme screening assay sample in red.

As seen in **Figure 14**, the biological blank sample which is displayed in green, shows only the fed galactose and displays no conversion to tagatose. The tagatose product is only seen when galactose is fed to the *Bacillus coagulans* NL01 enzyme. In this condition, shown in red, the fed galactose has an observed decrease and a new peak is formed that overlays perfectly with the grey tagatose standard, confirming that the *B.co*LAI enzyme can convert galactose to tagatose. After this characterization of the LAI was completed, the phosphatases were then tested *in vitro*.

Out of the 9 phosphatases tested on galactose-1-phosphate, only one of them, *Dictyostelium discoideum* (*D.dis*G1PP), showed higher conversion to galactose than the biological blank. This result is highlighted below.

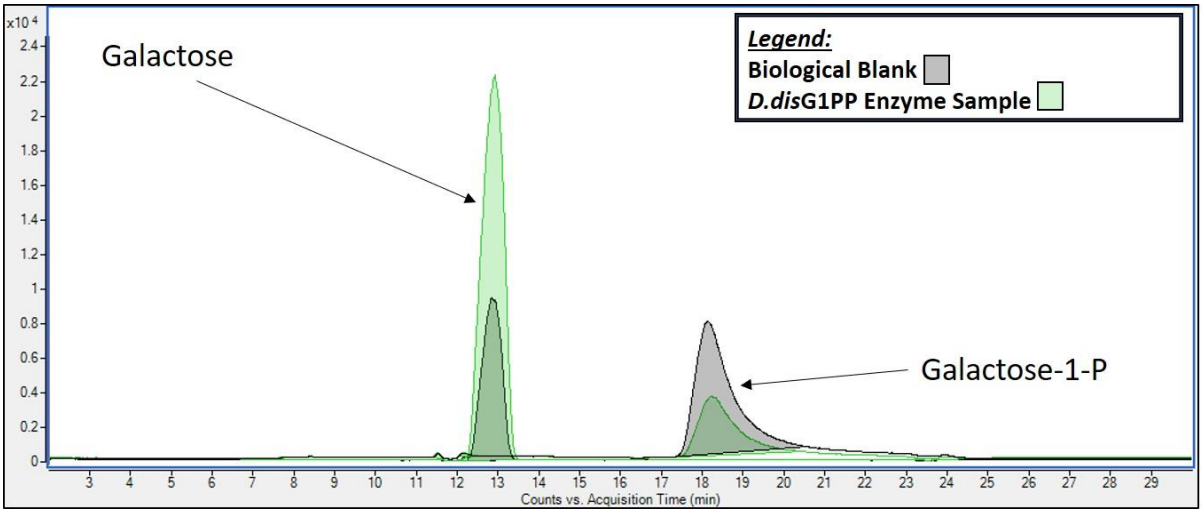


Figure 15. LC-QQQ Single Ion Monitoring Scans from the *In Vitro* Phosphatase Screen.

Reactants (galactose-1-phosphate) and products (galactose) are pictured here for the Dictyostelium discoideum phosphatase in vitro enzyme screen. 2 samples total are shown: The biological blank sample in grey, and the Dictyostelium discoideum phosphatase in vitro enzyme screening assay sample in green.

As seen in **Figure 15**, the biological blank sample, which is shown in grey, does show some background activity of converting galactose-1-phosphate to galactose. This is a cell free extract containing background expressed *E.coli* proteins for growth and metabolism so it is not unexpected to have seen some background activity on this substrate. However, the *Dictyostelium discoideum* phosphatase sample, shown in green, shows a much lower level of galactose-1-phosphate and a much higher level of galactose than the biological blank sample. This confirms the ability of the enzyme to dephosphorylate galactose-1-phosphate much more readily than the background *E.coli* enzymes, proving it to be an excellent candidate for the route 2 pathway.

Follow Up *In Vitro D.dis*G1PP Characterization Data

An additional substrate specificity assay was conducted after the initial selection of the *Dictyostelium discoideum* galactose-1-phosphate phosphatase (*D.dis*G1PP).The goal of this substrate specificity assay was to investigate how specific the *D.dis*G1PP was on galactose-1-phosphate and if it may interfere with other intermediates in the proposed route 2 pathway. Three of these pathway intermediates were tested with the *D.dis*G1PP, galactose-1-phosphate, glucose-1-phosphate, and glucose-6-phosphate. The results of this specificity assay are outlined in detail in **Figure 16**.

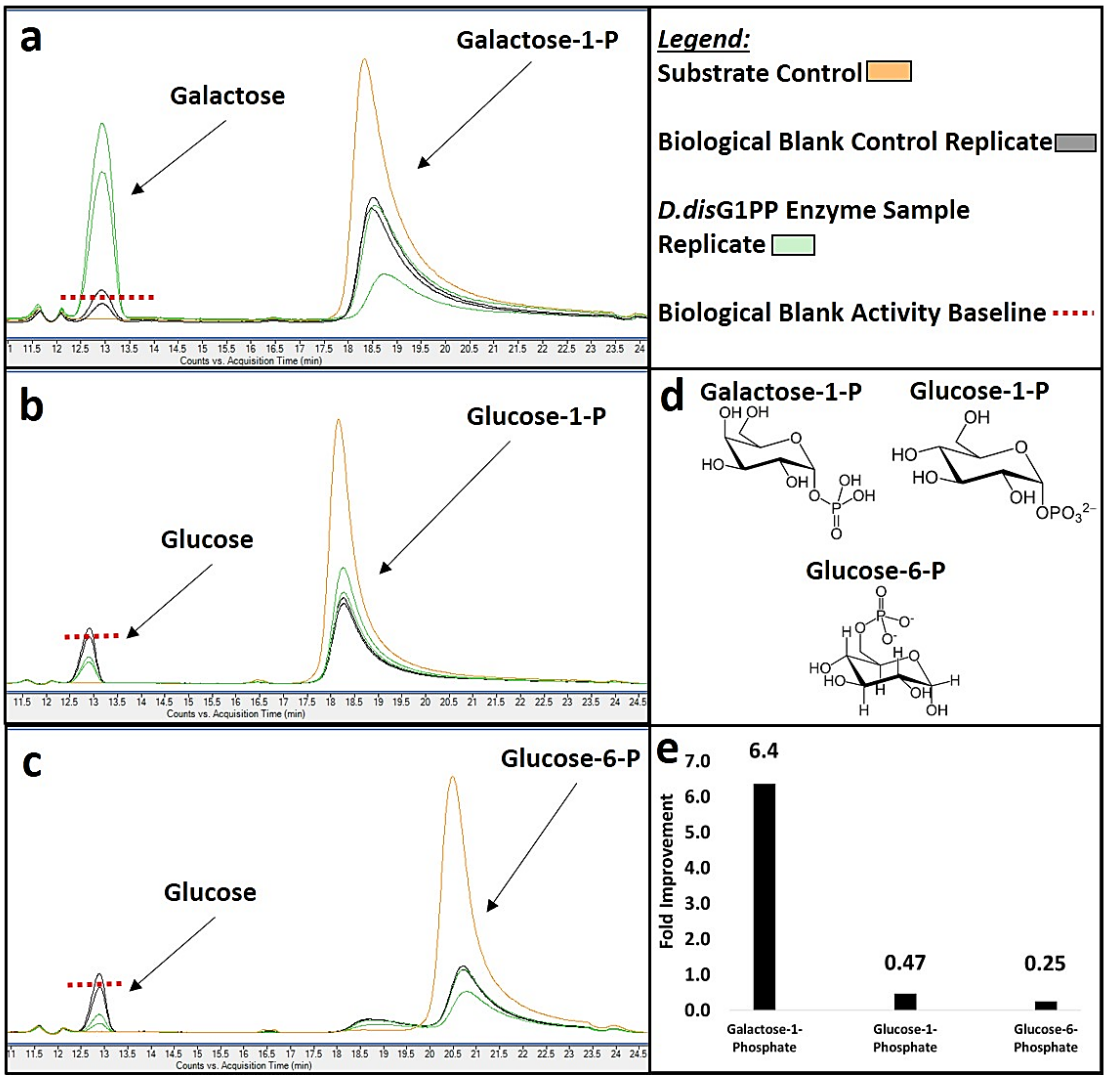


Figure 16. *D.dis*G1PP *In Vitro* Substrate Specificity Assays.

*Three total ion count chromatograms of the results from LC-QQQ single ion monitoring scans of the D.disG1PP specificity assays. 5 samples total are shown in each substrate feed chromatogram: A substrate standard control (substrate incubated alongside the other reactions without the presence of a cell free extract to control for potential temperature degradation) shown in light orange, 2 replicates of the biological blank control shown in black, and 2 replicates of the D.disG1PP enzyme sample shown in green. The biological blank activity baseline is annotated for visual convenience with a dotted red line. (**a**) Galactose-1-phosphate substrate feed results, (**b**) Glucose-1-phosphate substrate feed results, (**c**) Glucose-6-phosphate substrate feed results. (**d**) The chemical structures of galactose-1-phosphate, glucose-1-phosphate, and glucose-6-phosphate. (**e**)The fold improvements over the biological blank of D.disG1PP activity on the 3 different substrates.*

The *D.dis*G1PP was found to be highly specific with its phosphatase activity and preferred to act on galactose-1-phosphate much more than the other 2 substrates. This is shown clearly in **Figure 16a** as the galactose peak is far larger than the glucose peak from either **Figure 16b or c**. Additionally, when compared to the biological blank activities on these substrates, galactose-1-phosphate is the only substrate in which the *D.dis*G1PP exhibits significant activity over the biological blank, further indicating the enzyme's lack of activity on both glucose-1-phosphate and glucose-6-phosphate. This is seen in **Figure 16a-c** when comparing the level of the green lined peaks with the dotted red line indicating the biological blank activity baseline. The only substrate in which the green peaks are higher than the biological blank activity is in the galactose-1-phosphate substrate feed chromatogram (**see Figure 16a**). For both the glucose-1-phosphate and the glucose-6-phosphate feeds, the *D.dis*G1PP activity is below the biological blank baseline activity. To further illustrate this finding, **Figure 16e** displays the calculated fold improvements of the *D.dis*G1PP activity over the biological blank sample. Fold improvement is a calculated ratio of the LC-QQQ galactose or glucose response of 2 different samples. In this case, the *D.dis*G1PP enzyme sample response was divided by the biological blank sample response, achieving a ratio to easily compare the 2 conditions. It is shown here that the *D.dis*G1PP exhibits very high activity on galactose-1-phosphate (6.4 fold over the blank) and is indeed less active than the baseline blank activity on both glucose-1-phosphate and glucose-6-phosphate (0.47 and 0.25 fold over the blank respectively).

After the *B.co*LAI and the *D.dis*G1PP enzymes were selected for route 2 the enzymes were characterized *in vivo* to see if it was possible to express them in *E.coli* and

feed glucose to make tagatose. The fully developed proposed pathway after the *in vitro* enzyme discovery assays is outlined below.

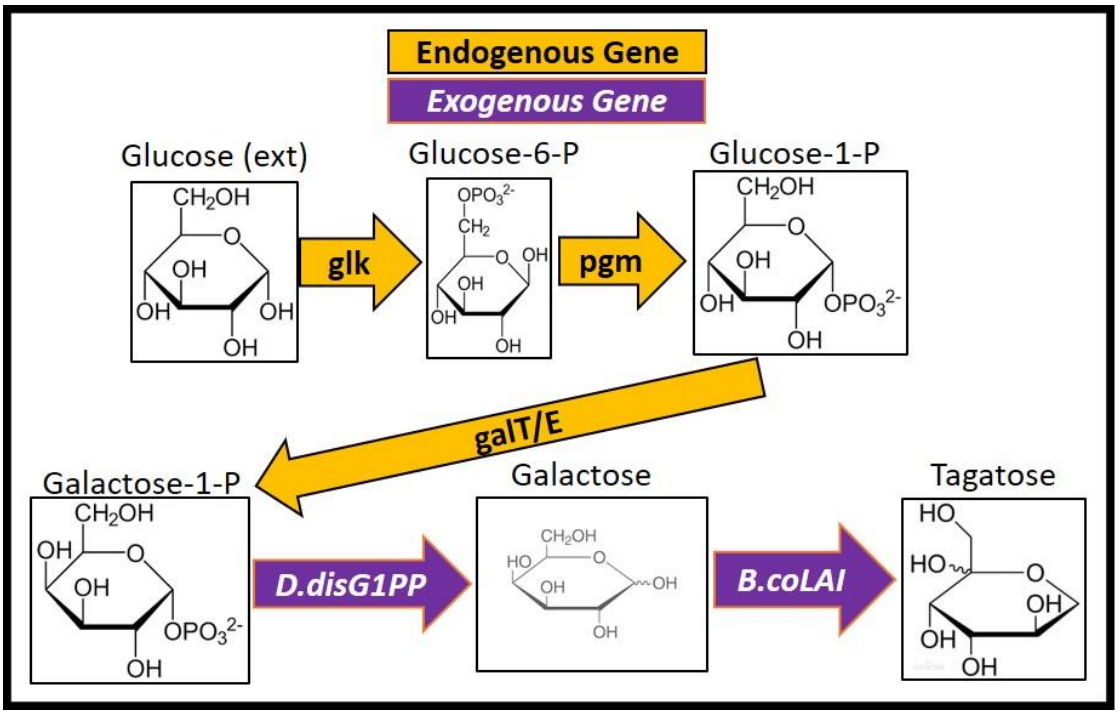


Figure 17. The Fully Developed Route 2 Pathway After Enzyme Discovery.

As seen in the key, endogenous E.coli genes are shown in yellow, while the selected exogenous enzymes are shown in purple. This pathway includes all of the necessary enzymes in the pathway but omits some supporting reactions and biproducts for simplicity. "ext"= external.

Route 2 *In Vivo* Experimental Results

In vivo experimentation was done in multiple steps, first by characterizing the *B.co*LAI on galactose, integrating it into the *E.coli* genome, and then characterizing the *D.dis*G1PP in the presence of the integrated LAI when feeding glucose. Genomic parts are often referred to from here on in. **See Table 1** for a list of definitions to these

genomic parts and loci. Additionally, see the Materials and Methods Chapter section, "Naming and Notation Conventions for *In Vivo* Data" for an extensive explanation of the naming conventions used in this project.

Chassis Development for *In Vivo* Screening

Before the recently selected *B.co*LAI and *D.dis*G1PP could be tested using *in vivo* fermentation assays, modifications to the chassis had to be made that would allow for galactose production *in vivo* while eliminating the cell's ability to metabolize the accumulated galactose. Allowing galactose to accumulate in the cell would allow the *B.co*LAI to have access to its intended galactose substrate and to isomerize it to tagatose. Under normal circumstances, *E.coli* will to convert galactose to glucose-1-phosphate using the Leloir pathway for cellular energy, (**see Figure 30**). In this project however, these same Leloir pathway genes are harnessed to run in the reverse direction as they normally would and convert fed glucose into galactose. This unusual phenotype is driven by the conversion of galactose-1-phosphate to galactose by the selective phosphatase, *D.dis*G1PP. This reaction removes the galactose-1-phosphate product and therefore drives the pathway in reverse toward galactose-1-phosphate. Additionally, this phenotype is driven by a few targeted modifications made to the chassis, these modifications are described in the following **Table 2** and the results of these individual modifications are outlined in **Figure 18** and **Figure 19**.

Table 2. A List of Strain Modifications Made.

Mod #	Mod Made to Strain	Endogenous Gene Name	Endogenous Gene Function	Mod Rationale
1	Knockout of galKM ΔgalKM	galK – galactokinase galM – galactose-1-epimerase	galK – phosphorylates galactose to prime it for catabolism galM - isomerizes α- and β-galactose	Keep galactose dephosphorylated so that it cannot be metabolized galM was knocked out due to its proximity to galK, it is a nonessential gene for the process and is mainly involved in lactose catabolism (Bouffard et al., 1994)
2	Knockout of pgi Δpgi	pgi - glucose-6-phosphate isomerase	Isomerize glucose-6-phosphate to fructose-6-phosphate to carry out glycolysis	Reduces competition for glucose-6-phosphate which can be converted to glucose-1-phosphate using pgm (see Mod 3)
3	Knock In of 46-7-pgm ::46-7-pgm	Pgm - phosphoglucomutase	Interconverts glucose-6-phosphate and glucose-1-phosphate	Necessary to make the desired glucose-1-phosphate intermediate

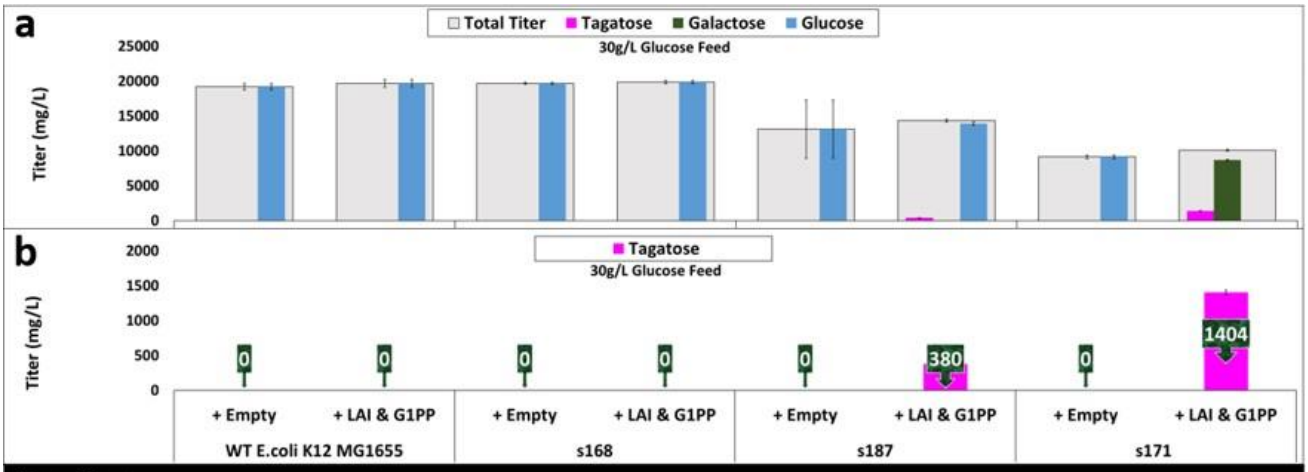
Modifications (Mod) were made to the chassis in the order of how they are numbered, 3 being the final modification made.

These modifications were made and verified sequentially. The strain lineage for this project is as follows:

1. s168 = Wild Type *E.coli* K12 MG1655ΔgalKM
2. s171 = s168Δpgi::46-7-pgm
3. s184 = s171ΔigA::46-7-*B.co*LAI
4. s186 = s184ΔlafU-dinB::46-7-*D.dis*G1PP

Strain modification's effect in relation to glucose, galactose, and tagatose production was assessed by complementing a 1BAC-46-7-B.coLAI-7-D.disG1PP

plasmid into the entire strain lineage. This experiment is outlined below in **Figures 18 and 19** and displays how the different modifications made to the strain affect the production potential. Strains are referred to by their strain numbers.

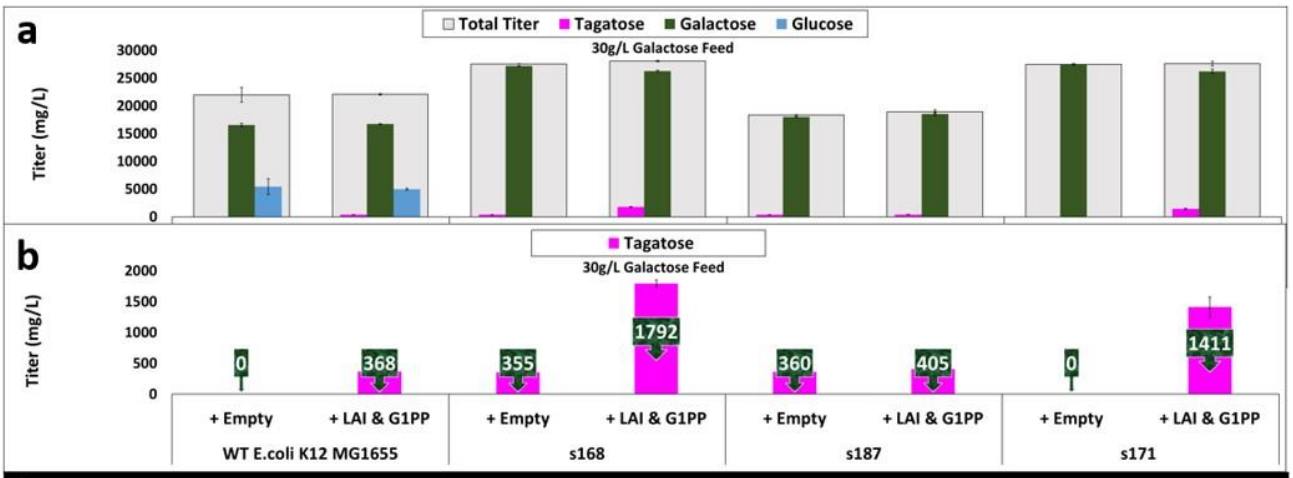


C Modification	WT *E.coli*	s168	s187	s171
ΔgalKM	✗	✓	✓	✓
Δpgi	✗	✗	✓	✓
::46-7-pgm	✗	✗	✗	✓
::*B.co*LAI	✗	✗	✗	✗
::*D.disG1PP*	✗	✗	✗	✗

Figure 18. Strain Modification Phenotypes on a 30g/L Glucose Feed.

*"+Empty" refers to the complementation of an empty 1BAC into the strain, "+LAI & G1PP" refers to the complementation of the formerly mentioned 46-7-B.coLAI-7-D.disG1PP plasmid into the strain.**(a)** A histogram displaying all measured metabolites and their added totals. **(b)** A histogram displaying only tagatose and its exact titers. **(c)** A table displaying the different modifications present in each strain. A green check mark signifies the modification is present in the strain and a red "X" signifies the modification is not present in the strain.*

When using glucose as a substrate, tagatose production cannot be seen until the deletion of pgi is present in the strain lineage, this implicates the pgi knockout's role in the loss of catabolite repression of the gal genes. Under normal conditions, in the presence of glucose, none of the gal genes would be expressed. However, it seems that once pgi is knocked out, gal gene repression is alleviated and they become expressed, directing carbon flux into the Leloir pathway. The production of tagatose is then dramatically increased (~3.7 fold) with the overexpression of pgm, see **Figure 18b**. There is also an accumulation of galactose in the product profile after the overexpression of pgm is incorporated into the chassis, indicating pgm is converting glucose-6-phosphate into glucose-1-phosphate as intended. This production of glucose-1-phosphate allows galT to convert it to galactose-1-phosphate, which is then dephosphorylated by the *D.dis*G1PP, making galactose *in vivo*.



C Modification	WT *E.coli*	s168	s187	s171
ΔgalKM	✗	✓	✓	✓
Δpgi	✗	✗	✓	✓
::46-7-pgm	✗	✗	✗	✓
::*B.co*LAI	✗	✗	✗	✗
::*D.disG1PP*	✗	✗	✗	✗

Figure 19. Strain Modification Phenotypes on a 30g/L Galactose Feed.

"+ Empty" refers to the complementation of an empty 1BAC into the strain, "+ LAI & G1PP" refers to the complementation of the formerly mentioned 46-7-B.coLAI-7-D.disG1PP plasmid into the strain. ***(a)*** *A histogram displaying all measured metabolites and their added totals.* ***(b)*** *A histogram displaying only tagatose and its exact titers.* ***(c)*** *A table displaying the different modifications present or not in each strain. A green check mark signifies the modification is present in the strain and a red "X" signifies the modification is not present in the strain.*

Figure 19 displays the effects of the modifications made to the chassis when feeding 30g/L galactose. A ΔgalKM *E.coli* strain cannot metabolize galactose and would normally not be able to grow at all in these types of conditions. However, the media here is supplemented with yeast extract that can be utilized for cell growth. Additionally, small amounts of tagatose are made in some of the "+ empty" conditions. This is likely

due to background activity of the endogenous *E.coli* LAI, araA, converting the available galactose to tagatose.

Here, the effect of the ΔgalKM modification becomes apparent when looking at the difference between wildtype *E.coli* and s168(ΔgalKM), there is a large increase in tagatose production (~5 fold). There is also an accumulation of glucose in the wildtype *E.coli* but this phenotype does not appear once the ΔgalKM modification is introduced into the chassis. This is likely due to the Leloir pathway working the way it was intended in the wildtype *E.coli* strain and converting the fed galactose into glucose-1-phosphate. The glucose-1-phosphate appears to then be dephosphorylated into free glucose, perhaps due to native phosphatase activities.

In Vivo Enzyme Characterization

*B.co*LAI was first characterized by an *in vivo* fermentation assay. This was done on plasmid to confirm that the LAI would in fact convert galactose to tagatose *in vivo*. After this initial confirmation, the *B.co*LAI was integrated into the genome and assessed in its ability to convert 5g/L galactose to tagatose both on plasmid and when chromosomally integrated. These results are outlined below.

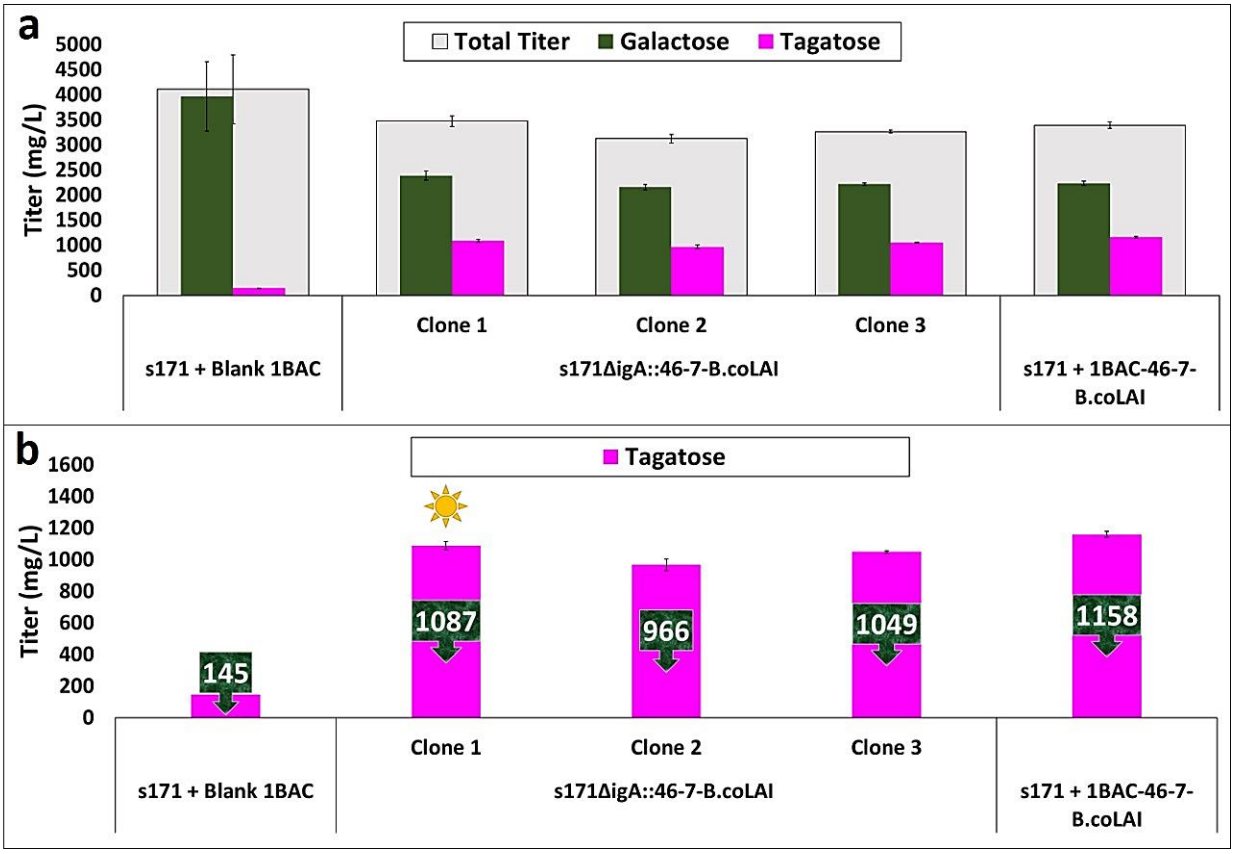


Figure 20. Characterization of the *B.co*LAI on a 5g/L Galactose Feed.

*A depiction of 3 clones of 46-7-B.coLAI integrated into the s171 genome. The assay is also controlled with the complementation of 1BAC-46-7-B.coLAI and an empty 1BAC into s171(**a**) A histogram displaying all measured metabolites and their added totals. (**b**) A histogram displaying only tagatose and its exact titers. The golden sun signifies which clone was deemed the best and moved forward with.*

As seen in **Figure 20**, the blank 1BAC condition again shows small amounts of galactose to tagatose conversion, likely due to the native *E.coli* LAI, araA. Clone 1 of the 46-7-*B.co*LAI integration displays the most ability to convert the available galactose into tagatose (~1.08g/L), yielding a % conversion of 31.3%. The plasmid complementation of

46-7-*B.co*LAI was able to produce slightly higher amounts of tagatose (~1.15g/L), yielding a % conversion of 34.2%.

The *B.co*LAI integrated into the genome was cataloged as s184 (ΔgalKM, Δpgi::46-7-pgm, ::46-7-B.coLAI). This strain allowed for the addition of the *D.dis*G1PP into the system to test for the production tagatose directly from glucose. This was first done on plasmid and then the phosphatase was integrated into the genome to begin exploring chassis modifications. **Figure 21** depicts the first data set collected with the *D.dis*G1PP on a 30g/L glucose feed and outlines its effect on the production of tagatose under these conditions.

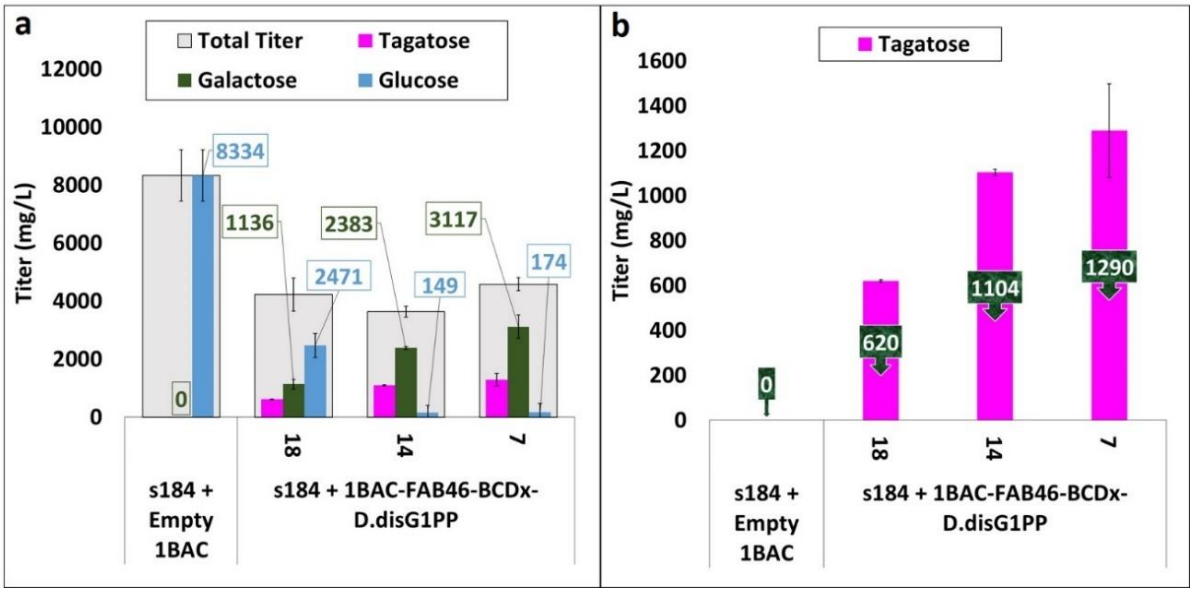


Figure 21. *D.dis*G1PP Tagatose Production on Plasmid from a 30g/L Glucose Feed.

*A depiction of 3 expression levels of FAB46-D.disG1PP with varying BCD strengths complemented into a LAI integrated strain, s184 (ΔgalKM, Δpgi::46-7-pgm, ::46-7-B.coLAI). The assay is controlled with a complementation of an empty 1BAC into s184(**a**) A histogram displaying all measured metabolites and their added totals. (**b**) A histogram displaying only tagatose and its exact titers.*

As seen in **Figure 21**, when both the LAI and *D.dis*G1PP are present, the system is able to make tagatose from a 30g/L glucose feed. This is different from the above results in **Figure 20**, which only includes the LAI and therefore, the system needs to be fed galactose to make tagatose. This result contains the initial, full proof of concept for route 2 to make tagatose *in vivo* directly from glucose.

Additionally, tagatose, galactose, and glucose titers are all affected by the differing levels of *D.dis*G1PP. Production of tagatose and galactose increases with the increasing expression of *D.dis*G1PP. The highest tagatose production titer being made from 1BAC-46-7-*D.dis*G1PP is 1290mg/L, (~4.3% yield, based on 30g/L glucose fed) and this corresponds with a galactose titer of 3117mg/L (~10.4% yield, based on 30g/L glucose fed). Additionally, glucose drops far lower than the empty 1BAC control once the *D.dis*G1PP is introduced. Glucose appears to be essentially depleted once the *D.dis*G1PP is expressed at the 46-14 level, indicating the enzymes effectiveness at making galactose from glucose *in vivo*. It is important to note that total titers are low here in terms of the 30g/L glucose that was fed. This could be explained by the cell metabolizing the additional glucose that is not reported as product. The cell will always use some of the fed glucose for growth, so achieving 100% yield is theoretically impossible.

After the *D.dis*G1PP was characterized *in vivo*, and the initial proof of concept was achieved, some additional follow up studies were run to further investigate the *D.dis*G1PP and its effectiveness due to specificity. **Figure 22** depicts 2 additional tests done with this phosphatase. The first test is an exploration of 5 homologs of the *D.dis*G1PP that came from a protein BLAST search of the *D.dis*G1PP enzyme. When

looking at the top 100 BLAST search returned sequences, the top 5 enzymes were 72-79% identical to the *D.dis*G1PP. After that, the % identity dropped off to ~50%, indicating that these top 5 enzymes were much more closely related to the *D.dis*G1PP than any other enzyme. These 5 enzymes were cloned into BACs and tested alongside the *D.dis*G1PP for tagatose production from glucose. **See Appendix 3, Figure 31** for a full homolog tree depicting the results of this BLAST search.

An additional comparison with 2 overexpressed endogenous *E.coli* phosphatases was also performed. The purpose of these endogenous phosphatases was to assess how important the specificity of the phosphatase to galactose-1-phosphate was, or if it was simply a matter of overexpressing any phosphatase to get the system to make tagatose. The 2 endogenous phosphatases tested were agp and PhoA. Agp is documented to be a glucose-1-phosphatase. PhoA is documented to be a highly versatile *E.coli* phosphatase, performing 47 different phosphatase reactions *in vivo*, (Karp et al., 2014).

When looking at **Figure 22** below, all of the homologs of the *D.dis*G1PP make galactose and tagatose, with the exception of #4. Additionally, homolog #6 appears to make more galactose than the *D.dis*G1PP, however it is still making less tagatose. Homologs #2 and #6 also have large amounts of glucose in their profile, which is highly contrasting with that of the *D.dis*G1PP. These higher glucose titers potentially imply dephosphorylation of important intermediates such as glucose-6-phosphate and glucose-1-phosphate. Perhaps these enzymes are more active than the *D.dis*G1PP but less specific in terms of which phosphorylated sugars they will act on. Acting on these other intermediates pulls away from the carbon flux that can be moved towards tagatose. This may explain why these two homologs show lower tagatose titers.

Lastly, the complementations of both 46-7-agp and PhoA do not make any galactose or tagatose. When looking at **Figure 22**c, this is illustrated in further detail as there is no tagatose peak observed at all for either PhoA or agp. This result provides further evidence to the importance of the phosphatase's specificity to galactose-1-phosphate for the success of route 2 in this project.

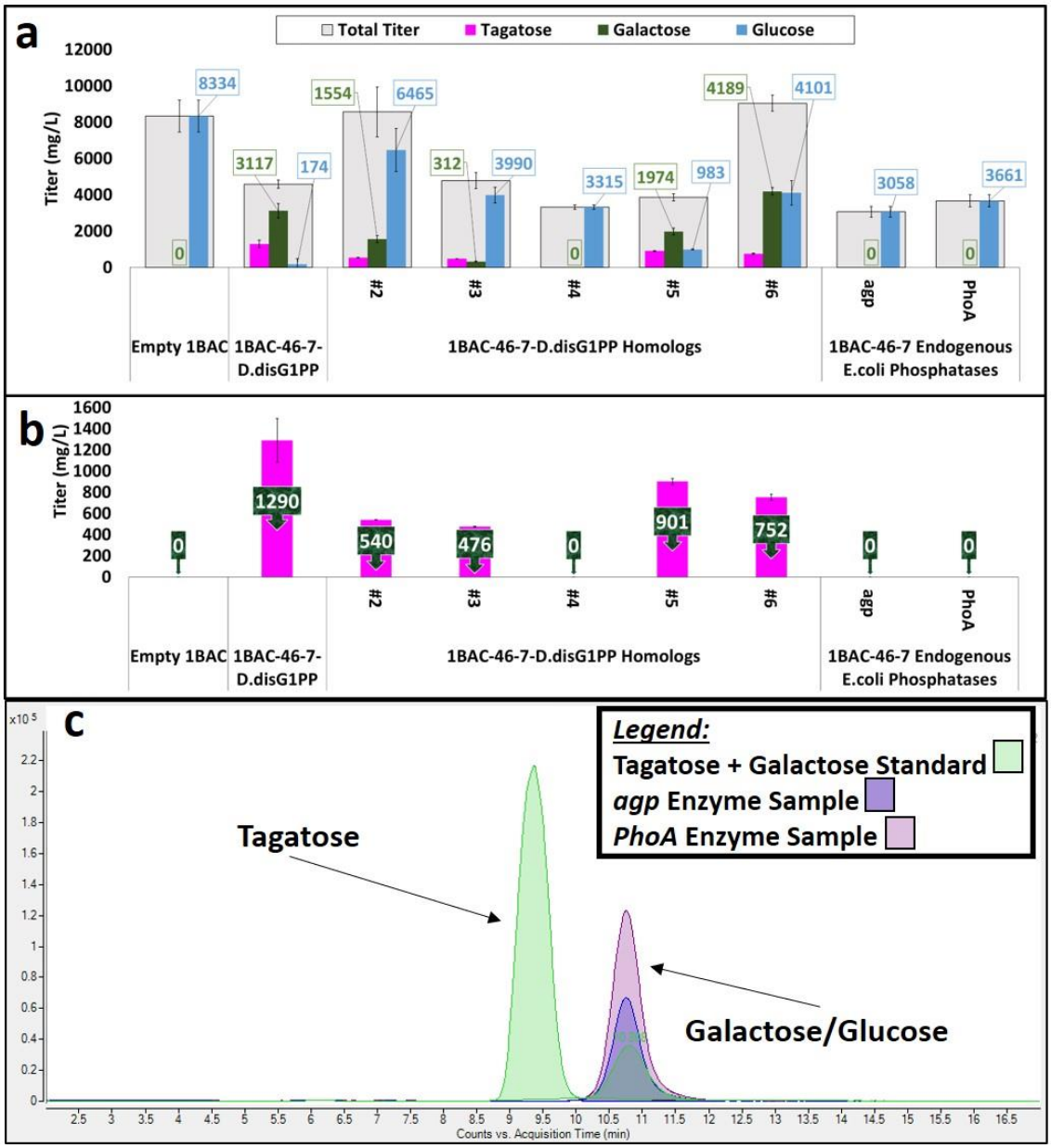


Figure 22. Other Phosphatases on a 30g/L Glucose Feed in s184.

*A depiction of 5 homologs of the D.disG1PP as well as 2 endogenous E.coli phosphatases complemented into a LAI integrated strain, s184 (ΔgalKM, Δpgi::46-7-pgm, ::46-7-B.coLAI). The assay is controlled with a complementation of an empty 1BAC into s184.(**a**) A histogram displaying all measured metabolites and their added totals. (**b**) A histogram displaying only tagatose and its exact titers. (**c**) LC-QQQ Single Ion Monitoring scans for the following samples: In green, a 100mg/L Tagatose/Galactose Standard. In purple: A biological sample of 1BAC-46-7-agp. In light purple: A biological sample of 1BAC-46-7-PhoA.*

After the information on the *D.dis*G1PP homologs and the endogenous *E.coli* phosphatases were collected, it was determined that the *D.dis*G1PP at a 46-7 expression strength was the best option to move forward with on the basis of tagatose production. This was then integrated into the s184 genome to create a fully integrated tagatose producing strain from glucose. The results of this integration are outlined below.

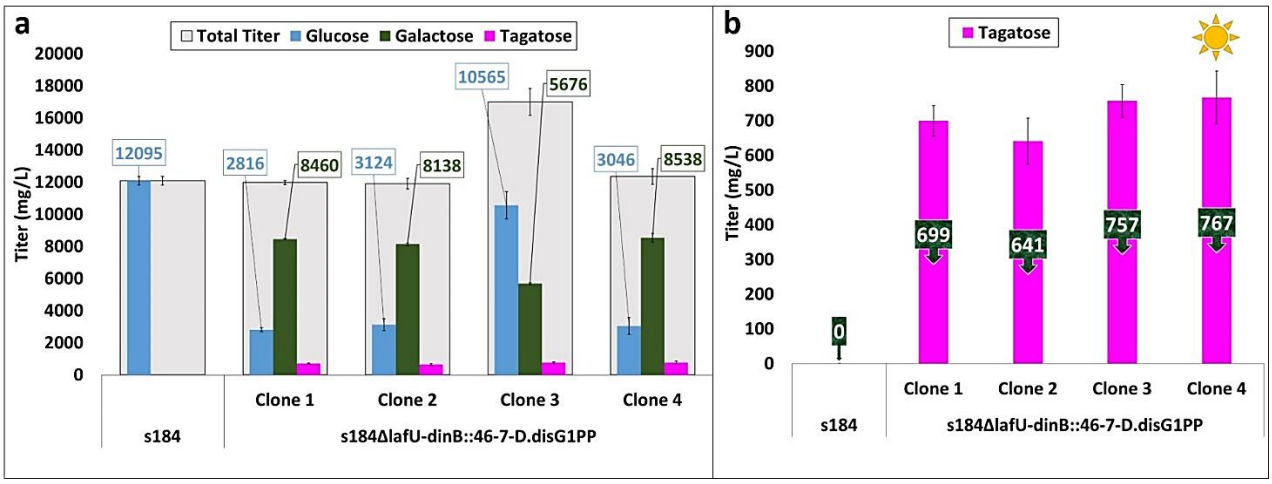


Figure 23. Integration of the *D.dis*G1PP on a 30g/L Glucose Feed.

*A depiction of 4 clones of 46-7-D.disG1PP integrated into the s184 genome. The assay is controlled with s184 (ΔgalKM, Δpgi::46-7-pgm, ::46-7-B.coLAI).(**a**) A histogram displaying all measured metabolites and their added totals. (**b**) A histogram displaying only tagatose and its exact titers. The golden sun signifies which clone was deemed the best and moved forward with.*

Initially, it appears that the glucose and galactose titers are much higher now that the enzyme has been integrated into the genome. When the results from **Figure 23** are compared with those from the previous **Figure 22,** these titers are increasing 50-100% after integration. This could be due to differential expression of these genes when on

plasmid versus when integrated into the genome. From this data in **Figure 23**, clone 4 was chosen as the lead candidate for a fully integrated tagatose strain. This clone made the most tagatose and the most galactose, indicating a functional *D.dis*G1PP as well as a functional *B.co*LAI. The integrated clone 4 was later named s186 (ΔgalKM, Δpgi::46-7-pgm, ::46-7-*B.co*LAI, ::46-7-*D.dis*G1PP) and is the strain used for all *in vivo* chassis assessments.

In Vivo Chassis Limitation Assessments

Chassis optimization targets for galactose-1-phosphate flux were assessed after s186 was made. The first assessment was to complement varying expression levels of 3 different gal genes, galE, galT, and galU, to see if this would help bring more flux into the Leloir pathway, (**see Appendix 2, Figure 30**). It is important to note that all the gal genes are endogenous to *E.coli* and that they seem to be expressed at some basal level in the fully integrated tagatose producing strains. This was discussed in the previous chassis development section, and likely has to do with catabolite repression loss as a result of the pgi knockout in the strain. The results of the complementations of these gal genes are outlined in **Figure 24**.

When assessing chassis limitations, 2 additional metrics were considered as a basis for differentiating results, "% pulled into pathway" and "relative % change." The % pulled into pathway metric was calculated by adding the % yields of galactose and tagatose together, where % yield was calculated on the basis of a 30g/L glucose feed. Both galactose and tagatose would have been made by converting galactose-1-phosphate into galactose and tagatose, hence they are both "pulled into the pathway."

The second metric, “relative % change,” was used to compare modifications made to the strain with the control condition. It is calculated in the following way: ((Final Titer – Initial Titer)/(Initial Titer))*100, where initial titer is the titer from the control condition, in this case, s186 + Empty 1BAC. Using this calculation, a positive value for relative % change indicates and increase in relation to the control titer, and a negative value indicates a decrease in relation to the control titer.

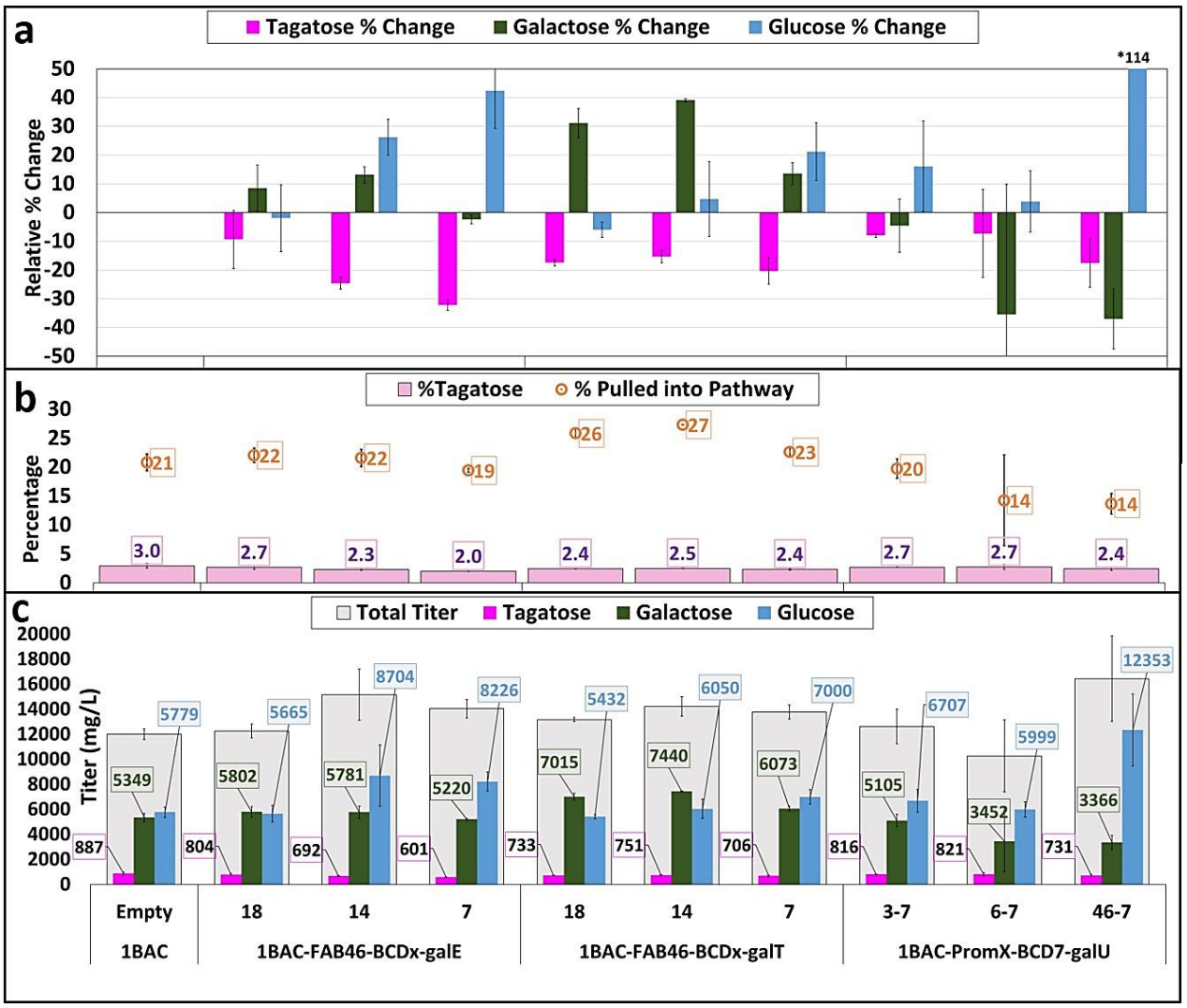


Figure 24. Complementation of Some Gal Genes into s186 on a 30g/L Glucose Feed.

(a) *A histogram depicting the percent changes of all the measured metabolites in the pathway relative to an s186 + Empty 1BAC control.****(b)*** *A graphical representation of the percentage of tagatose being made and the carbon flux successfully pulled into the pathway relative to the 30g/L of carbon fed as glucose.* ***(c)*** *A histogram displaying all measured metabolites and their added totals.*

Looking at **Figure 24**, the baseline percentage of carbon pulled into the pathway is 21%, this is in the control condition, s186 + empty 1BAC. Additionally, all of the gal

gene complementations here are deleterious for tagatose production. Complementations of 46-18 and 46-14 galT show both a 5% and 6% boost in % pulled into pathway and a 31% and 39% relative increases in galactose production respectively, this complementation however, is deleterious at the 46-7 expression level. The galE complementation has a similar relative % increase in galactose at the 46-18 and 46-14 level (8% and 13% respectively). These both result in 1% boosts in % pulled into the pathway. Finally, galU seems to be entirely deleterious to this process in terms of its relative % galactose changes, which are all negative, it also has large relative % increases in glucose titers leading to large decreases in the % pulled into the pathway. It additionally has some high error associated with it, which implicates it as a stressful modification. Overall, this complementation data suggests that additional expression of galT and also galE could be advantageous for the overall pathway, however, both complementations were still deleterious to tagatose production.

In addition to the investigation of a potential limitation in some of the gal genes, limitations of the enzymes in the pathway responsible for the novel chemistry were also assessed by additional complementation. **Figure 25** shows the results of these experiments in which complementation of the *B.co*LAI and *D.dis*G1PP enzymes were assessed as well as the complementation of some additional operons containing galE and galT together.

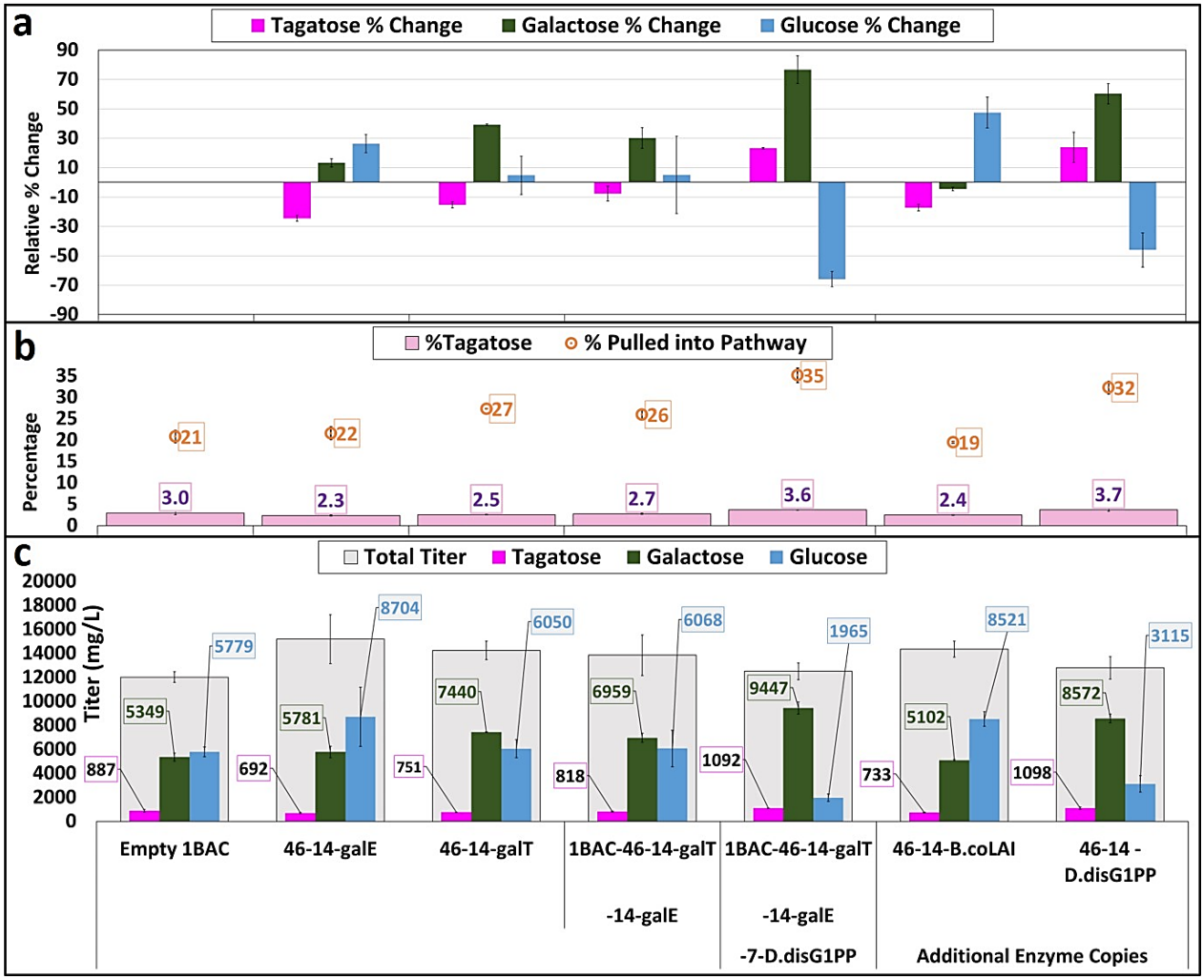


Figure 25. Additional Complementations into s186 on a 30g/L Glucose Feed.

(a) *A histogram depicting the percent changes of all the measured metabolites in the pathway relative to an s186 + Empty 1BAC control.****(b)*** *A graphical representation of the percentage of tagatose being made and the carbon flux successfully pulled into the pathway relative to the 30g/L of carbon fed as glucose.* ***(c)*** *A histogram displaying all measured metabolites and their added totals.*

Looking at **Figure 25**, the effect of additional complementation of 46-14 galE and 46-14-galT is again advantageous in terms of both pulling more carbon into the pathway (1%, and 6% respectively) and their relative percent change on galactose (13% and 39%

respectively). However, when a plasmid containing both 14-galE and 14-galT in an operon is complemented, their combined effect is less beneficial than just the complementation of 46-14-galT alone (5% more pulled into the pathway, 30% relative galactose increase). All 3 of these complementations are deleterious to the relative % change in tagatose as well.

Building off this, when *D.dis*G1PP is added into the operon, expressed at 46-7, a large shift is observed in favor of carbon being pulled into the pathway and in tagatose production (77% relative change in galactose, 23% relative change in tagatose, and 14% more carbon pulled into the pathway (35% total)). This result demonstrates a 31.5% yield of galactose from glucose. A similar effect is also seen with the additional complementation of 46-14-*D.dis*G1PP only (60% relative change in galactose, 24% relative change in tagatose, and 11% more carbon pulled into the pathway). This complementation makes slightly more tagatose than the 3 gene operon suggesting that the system is somewhat limited by its ability to make galactose, and subsequently tagatose, by *D.dis*G1PP expression.

Lastly, the additional complementation of the *B.co*LAI has a largely negative effect on the pathway (47% relative change in glucose, -17% relative change in galactose, and 2% less carbon pulled into the pathway). It is unclear why this complementation alone would have such a negative impact on the pathway. It would seem overall that the system is accumulating galactose, and additional *B.co*LAI would help to convert it to tagatose. Perhaps the overexpression of the enzyme alone had an impact on overall cell health.

Transport Assessment

Further experimentation where transporters were assessed on their ability to assist with tagatose production was conducted. The goal of these transporters was to assist in the conversion of galactose to tagatose by transporting the converted tagatose out of the cell. This in turn would drive the equilibrium of this isomerization reaction towards the tagatose product as it would be exported out of the cell and be partitioned away from the *B.co*LAI by the cell membrane. This concept has been described and assessed by a few groups, including the most recent Bober & Nair study which was discussed in the previous introduction chapter.

Two transporters were tested in this project, galP and mglBAC, both of which are endogenous to *E.coli* and are reported to be galactose importers. GalP is also reported to import glucose (Karp et al., 2014). While these are both documented galactose importers, the mglBAC transporter was implicated to be able to export tagatose from the *E.coli* membrane in the J.-H. Kim et al study, previously described in the introduction chapter. GalP was also included for this reason as perhaps it would be able to export tagatose as well.

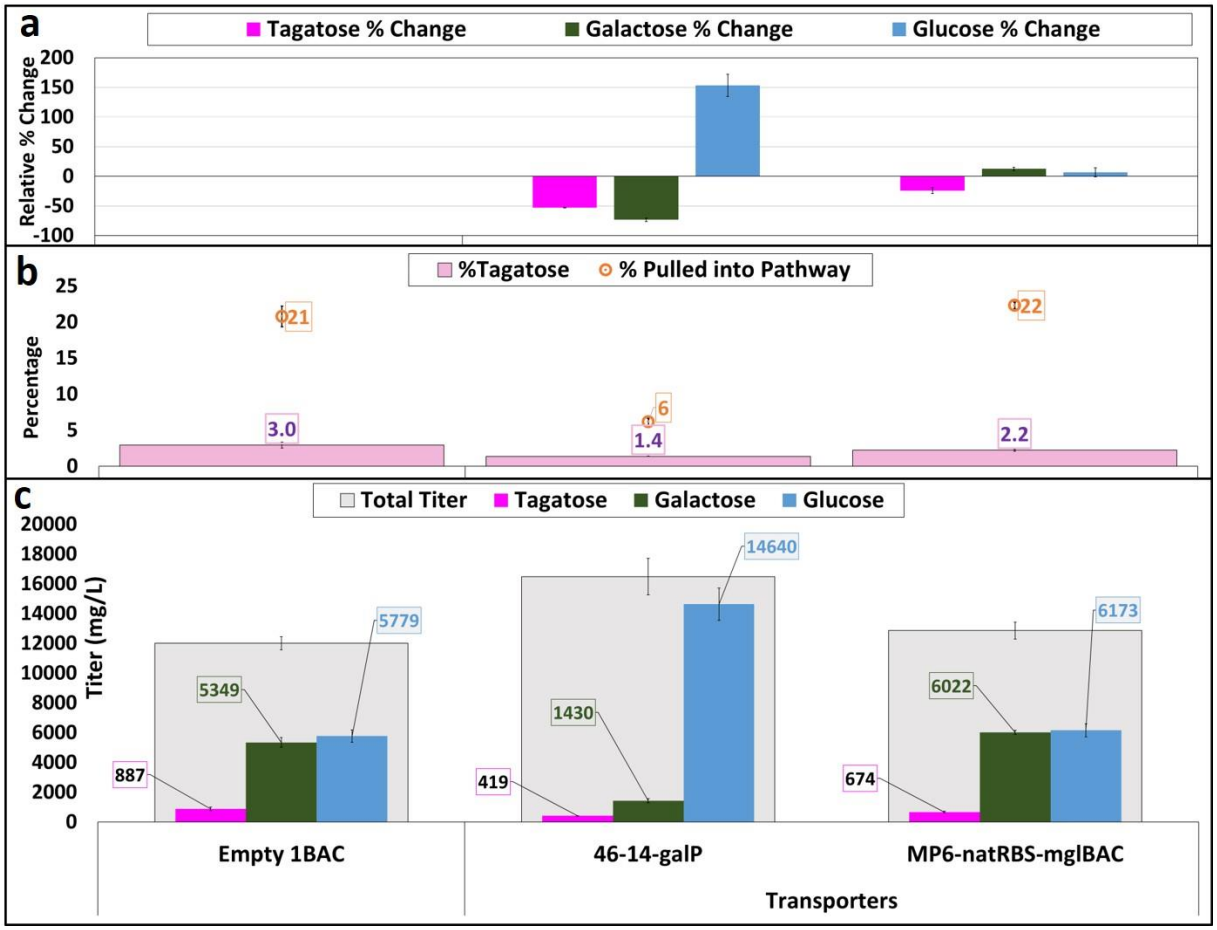


Figure 26. Transporter Complementations into s186 on a 30g/L Glucose Feed.

"natRBS" = native RBS as the mglBAC transporter was cloned with its native RBS instead of a BCD. ***(a)*** *A histogram depicting the percent changes of all the measured metabolites in the pathway relative to an s186 + Empty 1BAC control.* ***(b)*** *A graphical representation of the percentage of tagatose being made and the carbon flux successfully pulled into the pathway relative to the 30g/L of carbon fed as glucose.* ***(c)*** *A histogram displaying all measured metabolites and their added totals.*

Figure 26 shows that neither of the transporters were advantageous for tagatose production, in fact, the galP transporter was extremely deleterious, leading to a 15% reduction in carbon pulled into the pathway and a 53% relative reduction in tagatose titers. However, mglBAC did exhibit a slight increase in carbon pulled into the pathway

(1% increase) and also showed a 13% relative increase in galactose. It is hard to say exactly what may be happening with these transporters, mglBAC may be exporting galactose out of the cell instead of tagatose. This would lead to higher galactose titer but less tagatose as a result. Additionally, galP could be importing more glucose into the cell at a faster rate, which could lead to more catabolite repression of the gal genes, resulting in a very deleterious effect on the pathway

Full Pathway Limitation Follow Ups

After the collection of all the previously described data, it became apparent that the production of tagatose using the current pathway was most likely limited by the activity of the *B.co*LAI. Looking back on **Figures 24-26**, the s186 strain (ΔgalKM, Δpgi::46-7-pgm, ::46-7-*B.co*LAI, ::46-7-*D.dis*G1PP) consistently produces ~5g/L of galactose, however it is only making ~800mg/L of tagatose. This implies that something besides the expression of the *B.co*LAI is limiting the conversion of galactose, as it was seen in the previous data in **Figure 25** that the additional complementation of *B.co*LAI did not produce any more tagatose. Previous reports on the *B.co*LAI indicate that the LAI is most active at 60°C with an optimal concentration of its manganese (Mn2+) cofactor at 0.5mM (Mei et al., 2016). However, upon further investigation, the growth media used to collect the data so far had Mn2+ at a concentration of ~0.05mM.

To address the potential pathway limitation of the LAI, an experiment was designed to test the effect of additional supplementation of Mn2+, of running the reaction at 60°, and additionally, of allowing the reaction to be carried out for longer. This longer condition was added due to the possibility that the isomerization of galactose to tagatose

is incomplete at earlier timepoints. **Figure 27 and Figure 28** show the results of this experiment.

In **Figure 27**, it is seen that in both the 37°C and 60°C conditions, additional Mn2+ supplementation raises tagatose titers. It is also shown that the tagatose titers are at their highest in the first timepoint for both temperature conditions, indicating that either equilibrium is being reached between galactose and tagatose at this timepoint, or there is some other unknown limiting factor. The highest tagatose titers seen from the 30g/L glucose feed are in the 24hr at 60°C condition with additional supplementation of Mn2+ (1622mg/L corresponding to a 5.4% yield). Additionally, galactose titers are seen at their highest in the 48hr, 37°C timepoint (~6g/L). These galactose titers dramatically decrease throughout the remainder of the time course with no additional rise in tagatose titers.

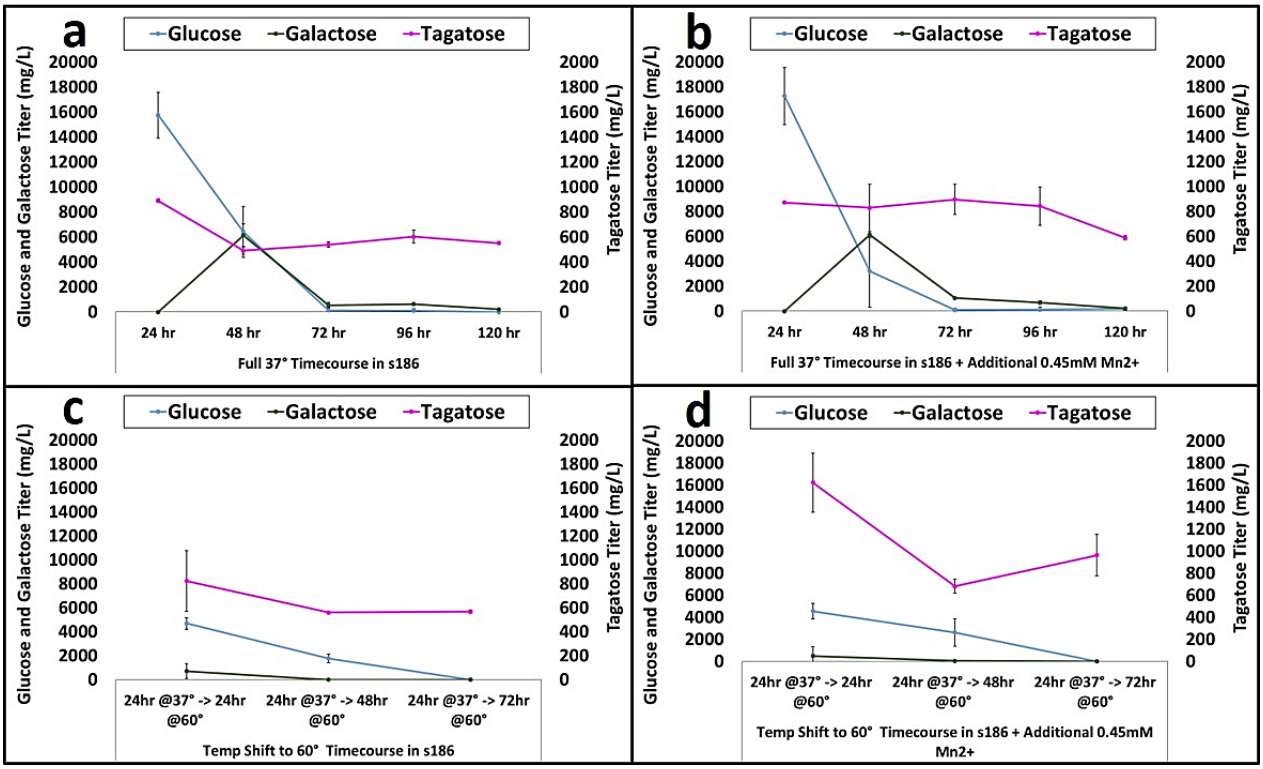


Figure 27. *B.co*LAI Limitation Follow Up Results on a 30g/L Glucose Feed.

*All plots display galactose and glucose on the left Y axis and tagatose on a separate right Y axis. **(a)** Titer results from a time course run entirely at 37°. **(b)** Titer results from a time course run entirely at 37° with the media supplemented with 0.45mM additional manganese. **(c)** Titer results from a time course run at 37° for the first 24hr and then shifted to 60° and sampled at the timepoints indicated. **(d)** Titer results from a time course run at 37° for the first 24hr and then shifted to 60° and sampled at the timepoints indicated, all media here was supplemented with additional 0.45mM manganese.*

In **Figure 28**, from the 30g/L galactose feed, it is again seen that in both the 37°C and 60°C conditions additional Mn2+ supplementation raises tagatose titers. A more expected time course result is seen for tagatose titers here as the titer generally increases throughout the course of the timepoints, which contrasts with the previous glucose results. When feeding galactose, the higher 60°C temp has a largely beneficial effect on

tagatose production. The highest tagatose titers seen from the 30g/L galactose feed are in the 72hr at 60°C condition with additional supplementation of Mn2+ (15487mg/L, corresponding to a 51.6% yield). **Figure 28a and b** show an odd decrease in tagatose titer at the 120hr timepoint, it is not clear why this occurs at this timepoint, perhaps product degradation.

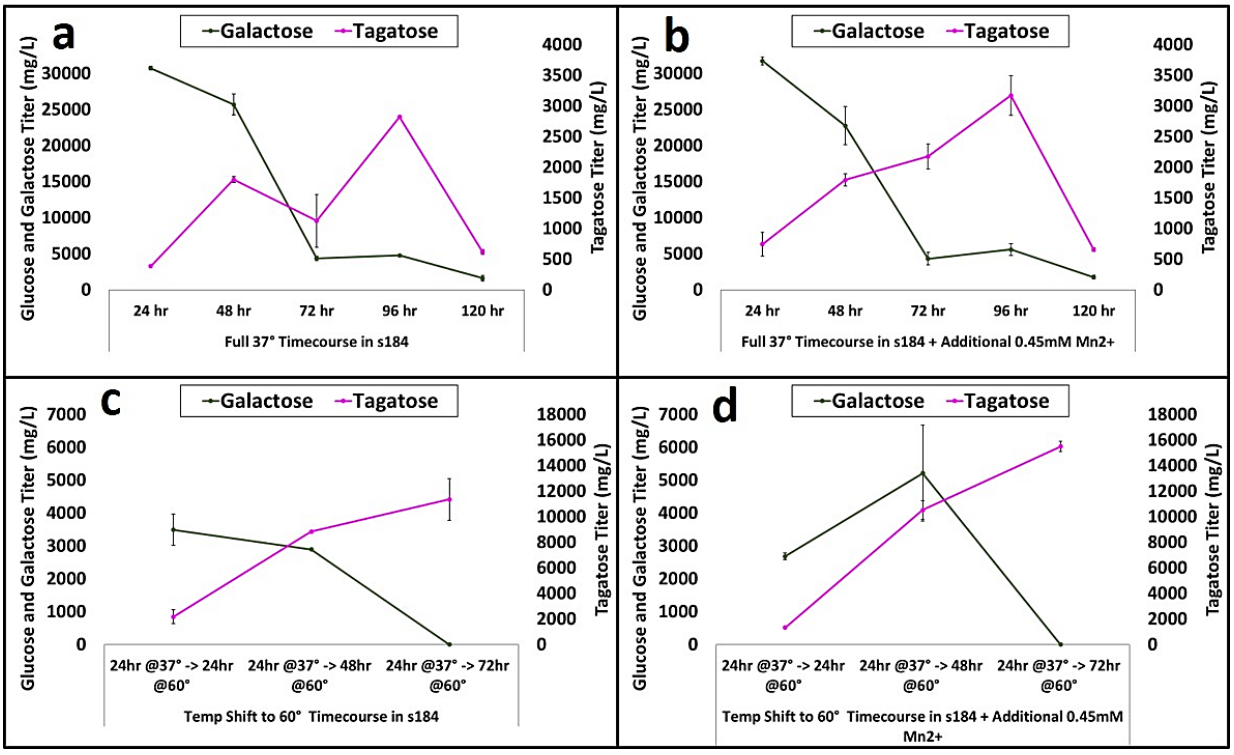


Figure 28. *B.co*LAI Limitation Follow up Results on a 30g/L Galactose Feed.

*All plots display galactose and glucose on the left Y axis and tagatose on a separate right Y axis. **(a)** Titer results from a time course run entirely at 37°. **(b)** Titer results from a time course run entirely at 37° with the media supplemented with 0.45mM additional manganese. **(c)** Titer results from a time course run at 37° for the first 24hr and then shifted to 60° and sampled at the timepoints indicated. **(d)** Titer results from a time course run at 37° for the first 24hr and then shifted to 60° and sampled at the timepoints indicated, all media here was supplemented with additional 0.45mM manganese.*

Figures 27 and 28 both show overall beneficial effects of additional supplementation of Mn2+ and elevated temperatures for tagatose production. However, they also both highlight potential issues with the system, as they show decreases in galactose concentration without increases in tagatose production. These results indicate potential galactose catabolism. Theoretically, in s186, once the *D.dis*G1PP dephosphorylates galactose-1-phosphate into galactose it should not be metabolized as galK has been knocked out of the strain. This principle applies to galactose feeding in s184 as well. However, in both the galactose and the glucose feeding time courses the level of galactose is going down over time without a comparable increase in tagatose.

Chapter IV.

Discussion

In this study, a novel pathway for the production of D-tagatose from glucose was designed, implemented, and assessed using a fully *in vivo* microbial host approach. To make this tagatose production possible, a novel bioproduction scheme of galactose from glucose was also implemented in this project. This bioproduction of galactose *in vivo* was made possible by the discovery of a novel chemistry performed by the *Dictyostelium discoideum* phosphatase on galactose-1-phosphate. In addition to this novel chemistry, the study demonstrates the largely specific nature of the *D.dis*G1PP on galactose-1-phosphate in relation to other common central carbon intermediates. The specificity of the *D.dis*G1PP is critically important for this process to work and it does not function with non-specific phosphatases.

This project lays fundamental groundwork for the future development of rare, natural, low calorie, and low glycemic sugars, such as tagatose, by using bioproduction tactics. The current industrial tactics for making these rare sugars incur large costs, are difficult to implement, and are less sustainable. Here, a potential alternative to those tactics is demonstrated using a cheap glucose feed source and an entirely *in vivo* microbial bioproduction process. Alternative sweeteners are of growing interest to worldwide consumers because they are healthier alternatives to commonly used artificial sweeteners such as sucralose, and aspartame.

Follow-Up Studies to Improve Tagatose Production

As mentioned in the previous results section, this process is currently limited by the terminal isomerization reaction of galactose to tagatose by the *B.co*LAI. Attempts to solve this limitation were taken in the previous results section such as allowing the reaction to run for a longer duration, supplementing with additional manganese cofactor, and also running at elevated temperatures. However, these did not seem to entirely solve the issue. Other approaches to solving the problem of *B.co*LAI limitation could be to select a different LAI from the literature. Additionally, further looking into manganese supplementation and optimizing it for the enzyme's *in vivo* performance cold be helpful as the 0.5mM optimal concentration was established in an *in vitro* setting (Mei et al., 2016).

Apart from assessing the *B.co*LAI limitation, the system's overall limitations could also be explored. An additional hypothesis as to why this final step may not be occurring is that the galactose being made is metabolized by the cell. Deletion of the galK gene should eliminate the cell's ability to phosphorylate and subsequently metabolize galactose. However, this genetic system is not yet fully understood and the combination of both a galK and pgi knockout with the accumulation of galactose could lead to unusual changes in transcriptional regulation of other genes such as kinases. These kinases could become overexpressed in this context and phosphorylate galactose, making it metabolizable and jeopardizing this system's ability to produce tagatose. To investigate this issue, a panel of kinases known to phosphorylate sugars could be knocked out of the s186 strain and tested for its increase in tagatose and galactose production. There are

hundreds of endogenous *E.coli* kinases, however, a few candidates for these knock outs could include YegV, Yeil, YgcE, gntK, idnK, glpK and alsK.

A final alternative hypothesis to unusual galactose catabolism routes is that the galactose made is being exported from the cell. This may happen with this unusual strain genotype as the cell may become stressed with high concentrations of unmetabolizable intracellular sugars. A systematic knockout of different known sugar transporters may help to alleviate this issue. Candidates for these knock outs include non-selective porins such as ompC, ompF, and PhoE as well as genes known to facilitate sugar transport such as galP, lacY, setA, setB, and setC.

Lastly, there are many endogenous *E.coli* phosphatases reported to dephosphorylate glucose-1-phosphate into glucose. All activities of this nature will have a direct negative impact on the carbon flux that can be pulled into the pathway. Viable candidates for targeted knock out to prevent this from happening include yihX, agp, ybiV, yidA, yigL, and PhoA.

Other Follow Up Studies Proposed to Characterize the System Further

An assessment of the cellular partitioning of galactose and tagatose would also be a logical next step. It is important to characterize where these metabolites are in relation to the cell membrane as it has been shown in previous studies that the cell membrane has the ability to partition galactose from the tagatose product and therefore shift the equilibrium of the reaction toward the desired product. Establishing the cellular localization of these metabolites could inform how to move forward and potentially leverage this partitioning for the improvement of tagatose production.

Additionally, as it was first tested and characterized here, the further characterization of the *D.dis*G1PP would be advantageous to understanding the system's full potential. Mainly, it is still unknown how much this phosphatase will act on other important central carbon intermediates. Although it was characterized to prefer galactose-1-phosphate over glucose-1-phosphate and glucose-6-phosphate, a full study in which it is tested with other important central carbon intermediates would be beneficial. Other metabolites proposed for testing would be fructose 1,6 bisphosphate, glyceraldehyde 3-phosphate, DHAP, 6-phosphogluconolactone, 6-phosphosgluconate, ribulose-5-phosphate, xylulose-5-phosphate, ribose-5-phosphate, sedoheptulose-7-phosphate, erythrose-4-phosphate and fructose-6-phosphate as these are all key intermediates for cell growth.

Theoretical Yield of Tagatose Using the Proposed Pathway

The theoretical yield of tagatose from glucose is 100% considering that it is a monosaccharide isomerization and requires no additional inputs of mass. However, this yield is impossible *in vivo* and will be directly impacted by any glucose that is consumed for cellular growth. When considering all the conversions and associated supporting reactions, the following table demonstrates the fully elucidated chemical conversion from glucose to tagatose *in vivo.*

Table 3. All Reactions Associated with Route 2.

Step #	Reactants	Products	$\Delta_r G$
1	ATP + Glucose	Glucose-6-Phosphate + ADP	-20.4
2	Glucose-6-Phosphate	Glucose-1-Phosphate	7.4
3	Glucose-1-Phosphate + UTP	UDP-Glucose + PPi	-1.2
4	UDP-Glucose	UDP-Galactose	2.9
5	Glucose-1-Phosphate + UDP-Galactose	Galactose-1-Phosphate + UDP-Glucose	0.9
6	ADP + Galactose-1-Phosphate	Galactose + ATP	16.3
7	Galactose	Tagatose	-7.2
Total	Glucose + UTP + Glucose-1-Phosphate	PPi + UDP-Glucose + Tagatose	-1.3

All ΔrG values were obtained from https://equilibrator.weizmann.ac.il/, (Flamholz et al., 2012)

As there is a negative value for the ΔrG of the overall reaction from glucose to tagatose, this process is slightly thermodynamically favorable. To make it more favorable however, the following reaction mechanism of active transport could be considered: Glucose + ATP + H2O ⇌ Tagatose + ADP + Pi (ΔrG = -46.9) (Flamholz et al., 2012). Active transport of tagatose from the cell could not only enhance the equilibrium in favor of tagatose, but could also drive the ΔrG further negative and make the process even more thermodynamically favorable.

Appendix 1.

Information about Enzymes Tested for Discovery

Table 4. Enzymes Cloned for Cell Free Extract Assays

#	Gene Name	Gene Class	Gene Length	MW (KD)
1	Arabidopsis thaliana (Ep)	Epimerase	1059	39.23
2	Bacillus subtilis (Ep)	Epimerase	1017	36.91
3	Erwinia amylovora (Ep)	Epimerase	1017	36.76
4	Pisum sativum (Ep)	Epimerase	1056	39.04
5	X. axonopodis (Ep)	Epimerase	1173	43.1
6	Heliobacillus mobilis (Ep)	Epimerase	972	35.24
7	rhizobium leguminosarum (Ep)	Epimerase	984	35.93
8	bradyrhizobium japonicum (Ep)	Epimerase	1017	36.1
9	lactobacillus casei (Ep)	Epimerase	999	36.41
10	clostridium acetobutylicum (Ep)	Epimerase	993	36.55
11	Thermus thermophilus (Ep)	Epimerase	936	33.83
12	Lactococcus lactis IL1403 (Ep)	Epimerase	945	35.18
13	Oceanobacillus iheyensis (Ep)	Epimerase	1005	36.7
14	Arabidopsis thaliana (G1PP)	Phosphatase	816	29.12
15	Dictyostelium discoideum (G1PP)	Phosphatase	822	29.97

16	Caenorhabditis elegans (G1PP)	Phosphatase	861	31.07
17	Bacillus subtilis (G1PP)	Phosphatase	822	29.39
18	Archaeoglobus fulgidus (G1PP)	Phosphatase	675	24.55
19	Aquifex aeolicus (G1PP)	Phosphatase	645	23.97
20	Thermoplasma acidophilum (G1PP)	Phosphatase	678	25.35
21	S. cerevisiae S288c (G1PP)	Phosphatase	741	27.1
22	S. cerevisiae S288c PYP1 (G1PP)	Phosphatase	729	27.55
23	L. sakei (strain 23K) (LAI)	L-Arabinose Isomerase	1428	53.76
24	Bacillus coagulans NL01 (LAI)	L-Arabinose Isomerase	1425	53.53
25	S. cerevisiae S288c (Ep)	Epimerase	2142	78.28
26	B. longum JCM 1217 (Ep)	Epimerase	1065	37.3
egalE (27)	E.coli galE (Ep)	Epimerase	1017	37.3

There were 26 enzymes total cloned for the purpose of testing their activities. Abbreviations next to the gene names indicate their enzyme's intended function. Abbreviations are as follows: Ep(Galactose-1-Phosphate Epimerase), G1PP(Galactose-1-Phosphate Phosphatase, LAI(L-Arabinose Isomerase).

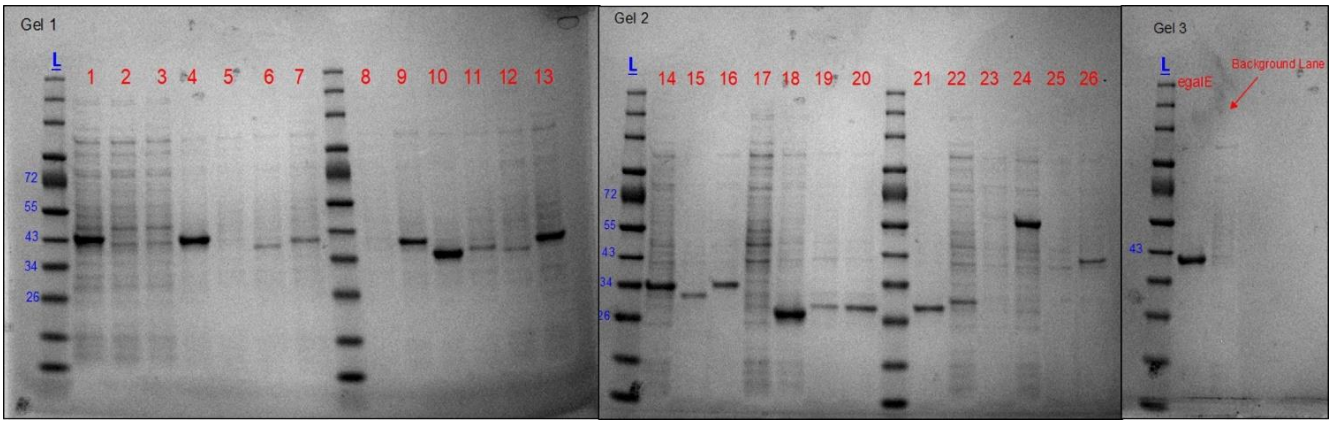


Figure 29. SDS-PAGE Gels Used to Assess Protein Expression.

Refer to Table 4 for the corresponding names and molecular weights to the numbered enzymes in this gel picture. Enzymes 2, 3, 5, 8, 17, 23, and 25 showed little to no apparent expression in this gel and were therefore dropped from the project. "L" Stands for protein ladder and the relevant sized bands are marked on the left with their molecular weights. The "background lane" contains a cell free extract produced from a blank pET28a(+) vector that had no enzyme cloned into it, hence this lane contains the "background" native E.coli proteins expressed only and gives a sense for how much of this background is expected to be in all the cell free extracts.

Appendix 2.

The Leloir Pathway

The Leloir pathway is endogenous to *E.coli* and is involved in the metabolism of galactose. This is accomplished by multiple transfers of UDP groups to ultimately convert the galactose into glucose-1-phosphate which can then be used for generation of cellular energy in the glycolysis pathway. The Leloir pathway involves 3 main genes, galactokinase (galK), galactose-1-phosphate uridylyltransferase (galT) and UDP-glucose 4-epimerase (galE). Additionally, galactose-1-epimerase (galM) is involved as well in converting the free galactose beta anomers into the biologically usable α anomer form. A depiction of the full Leloir pathway is pictured below in **Figure 30.**

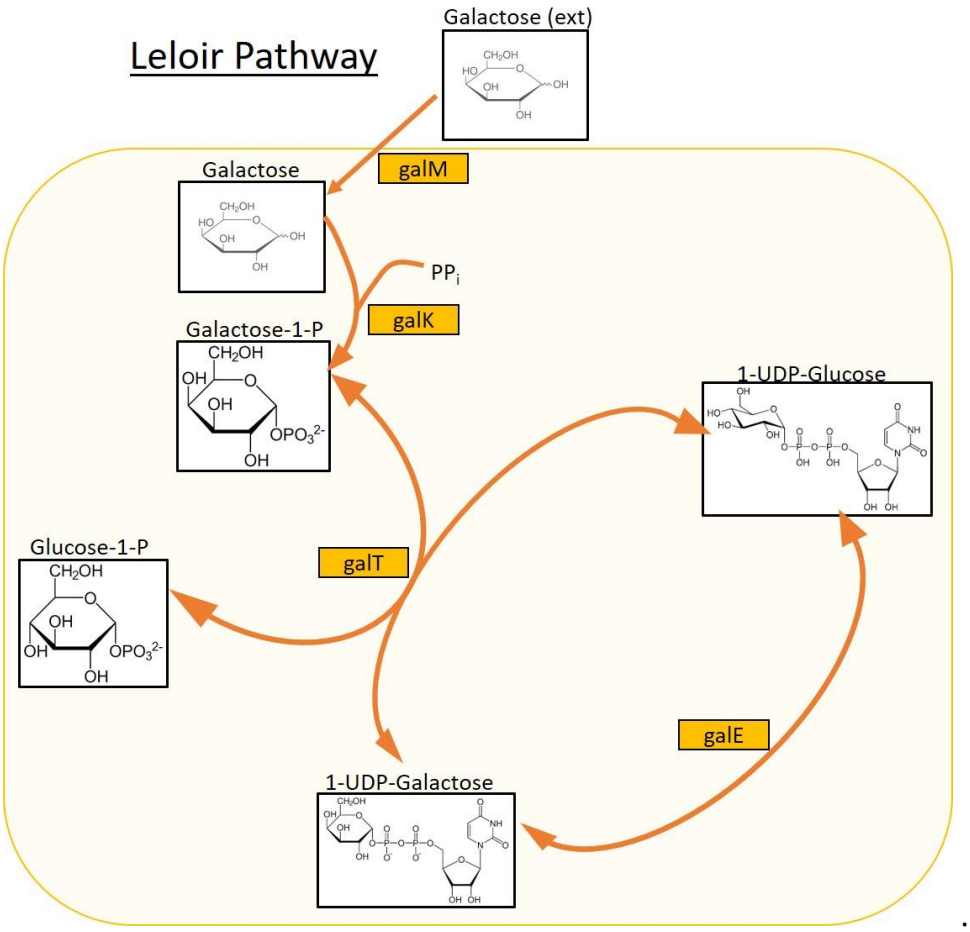


Figure 30. The Leloir Pathway

All gal genes are catabolite repressed in the presence of glucose. (Karp et al., 2014)

Appendix 3

*D.dis*G1PP Homolog Tree

The homolog tree below depicts the top 100 BLAST results from a protein blast run against the *D.dis*G1PP sequence. 5 clear homologs were chosen from this blast search as they had a natural partition from the others with greater than 70% sequence identity to the *D.dis*G1PP. The homologs also form a well-defined clade from the rest of the results, indicating that they may have all evolved from the *D.dis*G1PP enzyme and are its closest characterized relatives.

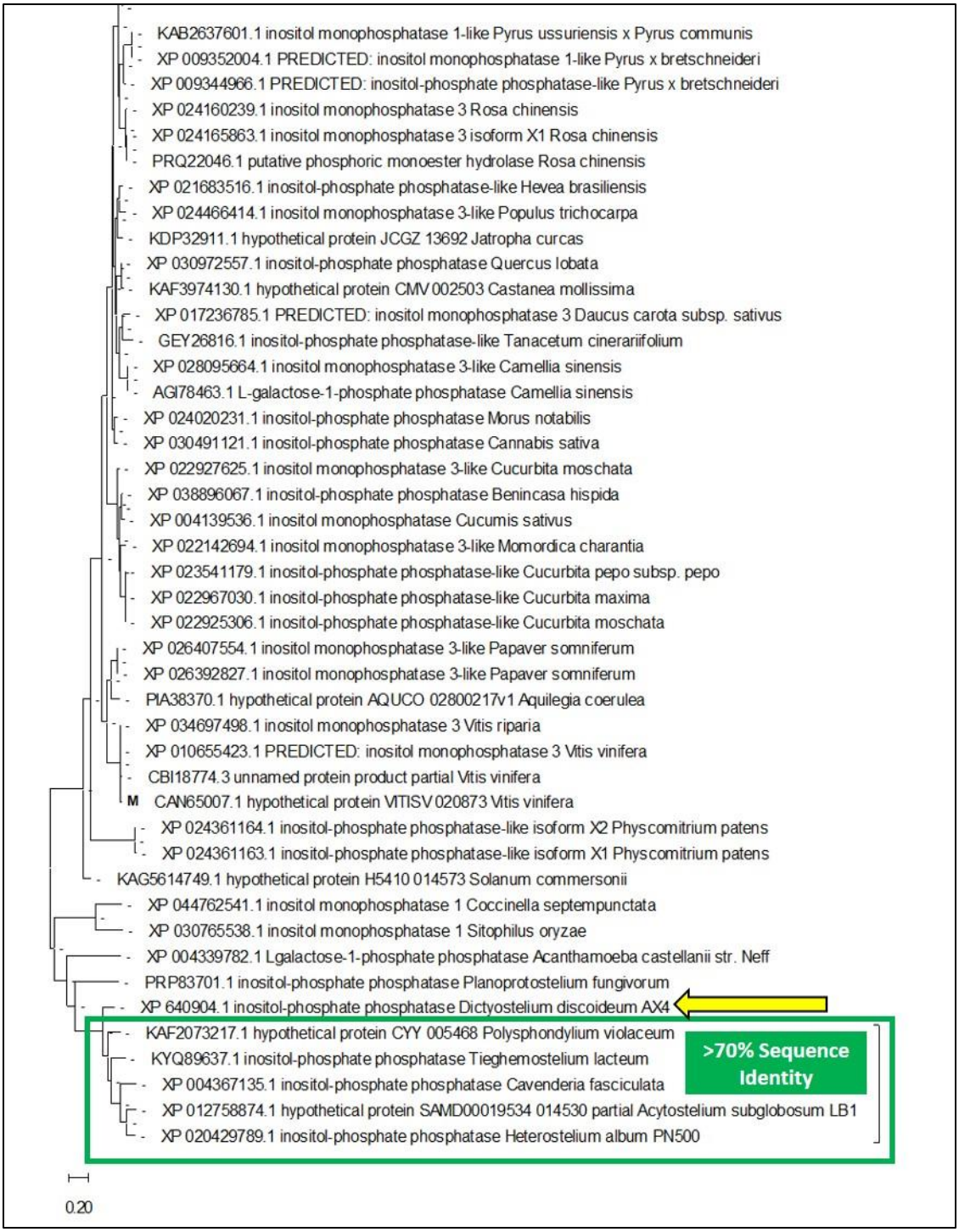


Figure 31. A Homolog Tree from the Top 100 BLAST Results for *D.dis*G1PP

D.disG1PP homologs 2-6 are highlighted with a green box and the D.disG1PP enzyme is pointed to by a yellow arrow.

Appendix 4.

Sources and Sequences for Genes and Plasmids

Below are **Tables 5 and 6** containing all sources for protein sequences and codon optimized DNA sequences for all genes used in this project.

Table 5. Sources and Sequences for Initial Enzymes Tested.

Enzyme #	Source	Sequence Name	Optimized DNA Sequence
1	https://www.uniprot.org/uniprot/Q42605.fasta	Arabidopsis thaliana (Ep)	ATGGCGGGCAGCAGCGTGGAACAGAACATTCTGGTGACCGGCGGCGCGGGCTTTATTGGCACCCATACCGTGGTGCAGCTGCTGAA AGATGGCTTTAAAGTGAGCATTATTGATAACTTTGATAACAGCGTGATTGAAGCGGTGGATCGCGTGCGCGAACTGGTGGGCCCGG ATCTGAGCAAAAAACTGGATTTTAACCTGGGCGATCTGCGCAACAAAGGCGATATTGAAAAACTGTTTAGCAAACAGCGCTTTGATG CGGTGATTCATTTTGCGGGCCTGAAAGCGGTGGGCGAAAGCGTGGAAAACCCGCGCCGCTATTTTGATAACAACCTGGTGGGCACC ATTAACCTGTATGAAACCATGGCGAAATATAACTGCAAAATGATGGTGTTTAGCAGCAGCGCGACCGTATATGGTCAGCCGGAAAAA ATTCCGTGCATGGAAGATTTTGAACTGAAAGCGATGAACCCGTATGGCCGCACCAAACTGTTTCTGGAAGAAATTGCGCGCGATATTC AGAAAGCGGAACCGGAATGGCGCATTATTCTGCTGCGCTATTTTAACCCGGTGGGCGCGCATGAAAGCGGCAGCATTGGCGAAGAT CCGAAAGGCATTCCGAACAACCTGATGCCGTATATTCAGCAAGTGGCGGTGGGCCGCCTGCCGGAACTGAACGTGTACGGCCATGAT TATCCGACCGAAGATGGCAGCGCGGTGCGCGATTATATTCATGTGATGGATCTGGCGGATGGCCATATTGCGGCGCTGCGCAAACTG TTTGCGGATCCGAAAATTGGCTGCACCGCGTATAACCTGGGCACCGGCCAAGGCACGAGCGTGCTGGAAATGGTGGCGGCGTTTGA AAAAGCGAGCGGCAAAAAAATTCCGATTAAACTGTGCCCGCGCCGCAGCGGCGATGCGACCGCGGTGTATGCGAGCACCGAAAAAG CGGAAAAAGAACTGGGCTGGAAAGCGAAATATGGCGTGGATGAAATGTGCCGCGATCAGTGGAAATGGGCGAACAACAACCCGTG GGGCTATCAGAACAAACTG
2	https://www.uniprot.org/uniprot/P55180.fasta	Bacillus subtilis (Ep)	ATGGCGATTCTGGTGACCGGCGGCGCGGGCTATATTGGCAGCCATACCTGCGTGGAACTGCTGAACAGCGGCTATGAAATTGTGGT GCTGGATAACCTGAGCAACAGCAGCGCGGAAGCGCTGAACCGCGTGAAAGAAATTACCGGCAAAGATCTGACCTTTTATGAAGCGG ATCTGCTGGATCGCGAAGCGGTGGATAGCGTGTTTGCGGAAAACGAAATTGAAGCGATTCATTTTGCGGGCCTGAAAGCGGTGGGC GAAAGCGTGGCGATTCCGCTGAAATATTATCATAACAACCTGACCGGCACCTTTATTCTGTGCGAAGCGATGGAAAAATATGGCGTG AAAAAAATTGTGTTTAGCAGCAGCGCGACCGTGTATGGCGTGCCGGAAACGAGCCCGATTACCGAAGATTTTCCGCTGGGCGCGACC AACCCGTATGGTCAGACCAAACTGATGCTGGAACAGATTCTGCGCGATCTGCATACCGCGGATAACGAATGGAGCGTGGCGCTGCTG CGCTATTTTAACCCGTTTGGCGCGCATCCGAGCGGCCGCATTGGCGAAGATCCGAACGGCATTCCGAACAACCTGATGCCGTATGTG GCGCAAGTGGCGGTGGGCAAACTGGAACAGCTGAGCGTGTTTGGCAACGATTATCCGACCAAAGATGGCACCGGCGTGCGCGATTA TATTCATGTGGTGGATCTGGCGGAAGGCCATGTGAAAGCGCTGGAAAAAGTGCTGAACAGCACCGGCGCGGATGCGTATAACCTGG GCACCGGCACCGGCTATAGCGTGCTGGAAATGGTGAAAGCGTTTGAAAAAGTGAGCGGCAAAGAAGTGCCGTATCGCTTTGCTGAT CGACGGCCGGGCGATATTGCGACCTGTTTTGCGGATCCGGCGAAAGCTAAACGCGAACTGGGCTGGGAAGCGAAACGCGGCCTGG AAGAAATGTGCGCGGATAGCTGGCGCTGGCAGAGCAGCAACGTGAACGGCTATAAAAGCGCGGAA
3	https://www.uniprot.org/uniprot/P35673.fasta	Erwinia amylovora (Ep)	ATGGCGAGCATTCTGGTGACCGGCGGCGCGGGCTATATTGGCAGCCATACCGTGCTGAGCCTGCTGCAGCGCGGCGATGATGTGGT GATTCTGGATAACCTGAGCAACGCGAGCCGCGAAAGCATTAACCGCGTGGAAAAACTGACCGGCAAAACCGCGACCTTTTTTGAAGG CGATCTGCTGGATCGCAGCTGCCTGCGCAGCGTGTTTAGCGCGCATCGCATTAGCGCGGTGATTCATTTTGCGGGCCTGAAAGCGGT GGGCGCGAGCACCCGCAAACCGCTGGAATATTATCAGAACAACGTGACCGGCACCCTGGTGCTGCTGGAAGAAATGCGCAGCGCGG GCGTGAATCAGTTTATTTTTAGCAGCAGCGCGACCGTGTATGGCGCGGATGCGCCGGTGCCGTATGTGGAAACCACCCCGATTGGCG GCACCACGAGCCCGTATGGCACGAGCAAACTGATGGTGGAACAGATTCTGCGCGATTATGCGAAAGCGAACCCGGAATTTAAAACC ATTGCGCTGCGCTATTTTAACCCGGTGGGCGCGCATGAAAGCGGTCAGATGGGCGAAGATCCGAACGGCATTCCGAACAACCTGCTG CCGTATATTGCGCAAGTGGCGATTGGCCGCCTGGAAAAACTGGGCATTTTTGGCGATGATTATCCGACCGAAGATGGCACCGGCGTG CGCGATTATATTCATGTGATGGATCTGGCGGAAGGCCATCTGAAAGCGCTGGATCATCTGAGCGCGATTGAAGGCTATAAAGCGTAT AACCTGGGCGCGGGCAAAGGCTATAGCGTGCTGGAAATGGTGAAAGCGTTTGAAAAAGCGAGCGGCGGCACCGTGGCGTATCAGA TTAGCCCGCGCCGCGATGGCGATCTGGCGGCGTTTTGGGCGGATGCGACCCTGGCGGATAAAGAACTGAACTGGCGCGTGAGCCGC GGCATTGATGAAATGATGCGCGATACCTGGAACTGGCAGAGTCAGAACCCGCAAGGCGATAGC

4	https://www.uniprot.org/uniprot/B0M3E8.fasta	Pisum sativum (Ep)	ATGGCGGTGGCGAGCAGTCAGAAAATTCTGGTGACCGGCGGCGCGGGGCTTTATTGGCACCCATACCGTGGTGCAGCTGCTGAACAA CGGCTTTAACGTGAGCATTATTGATAACTTTGATAACAGCGTGATGGAAGCGGTGGAACGCGTGCGCGAAGTGGTGGGCAGCAACC TGAGTCAGAACCTGGAATTTACCCTGGGCGATCTGCGCAACAAAGATGATCTGGAAAAACTGTTTAGCAAAAGCAAATTTGATGCGG TGATTCATTTTGCGGGCCTGAAAGCGGTGGGCGAAAGCGTGGAAAACCCGCGCCGCTATTTTGATAACAACCTGGTGGGCACCATTA ACCTGTATGAAGTGATGGCGAAACATAACTGCAAAAAAATGGTGTTTAGCAGCAGCGCGACCGTATATGGTCAGCCGGAAAAAATTC CGTGCGTGGAAGATTTTAAACTGCAAGCGATGAACCCGTATGGCCGCACCAAACTGTTTCTGGAAGAAATTGCGCGCGATATTCAGA AAGCGGAACCGGAATGGCGCATTGTGCTGCTGCGCTATTTTAACCCGGTGGGCGCGCATGAAAGCGGCAAACTGGGCGAAGATCCG CGCGGCATTCCGAACAACCTGATGCCGTATATTCAGCAAGTGGCGGTGGGCCGCCTGCCGGAACTGAACGTGTACGGCCATGATTAT CCGACCCGCGATGGCAGCGCGATTCGCGATTATATTCATGTGATGGATCTGGCGGATGGCCATATTGCGGCGCTGCGCAAACTGTTT ACGAGCGAAAACATTGGCTGCACCGCGTATAACCTGGGCACCGGCCGCGGCAGCAGCGTGCTGGAAATGGTGGCGGCGTTTGAAAA AGCGAGCGGCAAAAAAATTGCGCTGAAACTGTGCCCGCGCCGCCCGGGCGATGCGACCGAAGTGTATGCGAGCACCGCGAAAGCG GAAAAAGAACTGGGCTGGAAAGCGAAATATGGCGTGGAAGAAATGTGCCGCGATCAGTGGAACTGGGCGAAAAACAACCCGTGGG GCTATAGCGGCAAACCG
5	https://www.uniprot.org/uniprot/A0A2S6YC55.fasta	X. axonopodis (Ep)	ATGGCGCGCATTCTGGTGACCGGCGGCACCGGCTATATTGGCAGCCATACCTGCGTGGAACTGGCGCGCCGCGGCCATGAAGTGTG CATTGTGGATAACCTGAGCAACAGCAGCGAACGCGTGCTGGATCATCTGCAGCATCTGATGGGCTTTCGCCCGGAATTTCATCGCAT GGATGTGCGCGCGCCGGAACTGGCGGATCTGCTGATTAGCAAACGCATTGATGCGGTGCTGCATTTTGCGGCGCTGAAAGCGGTGG GCGAAAGCGTACGCGAACCGCTGCTGTATTTTAACAACAACGTGACCGGCACCCTGGCTCTGCTGCGCGCGATGCGCACCGCGAAAG TGTGCAACCTGGTGTTTAGCAGCAGCGCGACCGTGTATGGCGATGCGAACCGCAGCCCGATTGATGAAAGCGCGCCGCTGAAAGCG ATTAACCCGTATGGCCGCACCAAACTGATGATGGAAGAAATGATTGGCGATCTGAGCGCGGCGTGGCCGGATTTTAACGCGGCCCTG CTGCGCTATTTTAACCCGGTGGGCGCGCATCCGAGCGGCTATCTGGGCGAAGATCCGCAAGGCATGCCGAACAACCTGATGCCGTAT ATTGCGCAAGTGGCGGTGGGCCGTCGCGCTGCGCTTCAAGTGTTTGGCGATGATTATCCGACCGATGATGGCACCGGCGTGCGCGAT TATCTGCATGTGATGGATCTGGCGCGCGCGCATGTGGATGCGATTGATTATCTGCAGCGCGAACGCAAAGGCCTGGTGGTGAACCTG GGGACCGGTCGCGGATACAGCGTCCGCGAGGTGGCGGCTGCATTTGCGCGCGCGAGCGGCCGCCGCATTCCGCTGAACATTGCGCC GCGCCGCGATGGCGATGTGGCGGTGTATTTTGCGAGCACCGCGCTGGCGAACCGCCTGCTGGGCTGGACCGCGGAATATGATCTGG ATCGCATGTGCCGCGATACCTGGCGCTGGCAGAGCATGCATCCCGATGGTTATGCCGGTAGCCGTTTACCGAGACAAAGGTCAGCCG CCGTGGTGCCTGAACGCGATTCTGCTCGCGAACTGCCGGCGCCTGCCCGTCGTCTGTCGCCGGGCCCCGATACCCGCCGACGCAGCG AACCGGTGCGCACCCTGGTGGTGGCGGATGGCTTTAGCCGCGCG
6	https://www.uniprot.org/uniprot/Q8GE14.fasta	Heliobacillus mobilis (Ep)	ATGGCGCGCTTTCTGGTGACCGGCGGCGCGGGCTATATTGGCAGCCATACCGTGCTGGCGCTGCAAGCGGCGGGCCATGATGTGGT GATTCTGGATAACCTGACCACCGGCCATGCGTATCTGGTGCCGGATGGCGTGCCGTTTTATCAAGGCGATGTGGGCGATAGCAGCCT GCTGGAAAAAATTTTTCATGATTATCCGATTGATGGCGTGCTGCATTTTGCGGCGAAAAGCCTGGTGGGCGAAAGCATGCAAGATCC GGGCAAATATTTTCTGGCGAACACGAGTCAGAGCGCGGTGCTGCTGAGCGCGATGGCGAAACATAAAATTCCGTATTTTGTGCTGAG CAGCACCGCGGCGGTGTATGGCGAACCGCAAGATGTGCCGATTCCGGAAAGTCATCCGATTCGCCCGACCAACCCGTATGGCCTGAG CAAACATCTGATTGAACAGATGCTGCCGTGGTTTGAAAAAGTGCATCGCATGAAATGGATGGCGCTGCGCTATTTTAACGTGGCGGG CGCGGATCCGCAAGGCCGCAGCGGCGAAAGCCATGATCCGGAAACCCATCTGATTCCGAACATTCTGAAAGTGGCGCAAGGCCATC GCGATTTTGTGAGCGTGTTTGGCAACGATTATCCGACCGAAGATGGTACCTGCATTCGCGATTATATTCATGTGAGCGATCTGGCGGA TGCGCATGTGCTGGCGATTGAAGCGCTGTTTAGCGGCAAACCGAGCGGCGTGTGCAACCTGGGCAACGGCCAAGGCTTTAGCGTGC TGAGCGTGATTGAAAGCGCGCGCCGCGTGACCGGACATCCGATTCCGGTGCGCATTGAACCGCGCCGTCCGGGAGATCCCGCCGTG CTGGTGGCGAGCAACGCGAAAGCGGTGGCGGAACTGGGCTGGAAACCGAAATTTACCGATCTGAGCACCATTGTGGAAACCGCGTG GCTGTGGCAGCATCGC
7	https://www.uniprot.org/uniprot/Q59745.fasta	rhizobium leguminosarum (Ep)	ATGGCGGGCGAAACCGTGCTGGTGGTGGGCGGCGCGGGCTATATTGGCAGCCATACCTGCTTAGATCTGGCGAACAAAGGCTATCG CCCGGTGGTGTTTGATAACTTTAGCAACGGCCATCGCGAATTTGTGCGCTGGGGCCCGGCGGAAGAAGGCGATATTCGCGATCGCGC GCGCCTGGATGAAGTGCTGGCGAAACATAAACCGGCGGCGATTCTGCATTTTGCGGCGCTGATTGAAGTGGGCGAAAGCGTGAAAG ATCCGGTGAGCTTTTATGAAAACAACGTGATTGGCACCCTGACCCTGCTGAGCGCGGCGCAAGCGGCGGGCATTAACGCGTTTGTGT TTAGCAGTACCTGCGCGACCTATGGCCTGCCGCAGAGCGTGCCGCTGGATGAAACCCATCGCCAAGTGCCGATTAACCCGTATGGCC GCACCAAATATATTGTGGAACAAGCGCTGGCGGATTATGATCAGTATGGCAGCCTGCGCAGCGTGGTGCTGCGCTATTTTAACGCGG CGGGCGCGGATTTTGAAGGCCGCATTGGCGAATGGCATCAGCCGGAAACCCATGCGATTCCGCTGGCGATTGATGCGGCGCTGGGC CGCCGCCAAGGCTTTAAAGTGTTTGGCAGCGATTATGAAACCCGCGATGGTACCTGCGTGCGCGATTATATTCATGTGCTCGATCTGG CGGATGCGCATGTGCGCGCGGTGGAATATCTGCTGAAAGGCGGCGATAGCGTGGCGCTGAACCTGGGCACCGGCACCGGCACCACC GTGAAAGAACTGCTGGGCGCGATTGAAGAAGTGAGCAACCGCCCGTTTCCGGTGGAATATATTGGCCGCCGCGAAGGCGATAGCCA TACCCTGGTGGCGAACAACGATAAAGCGCGCGATGTGCTGGGCTGGGTGCCGCAGTATGATCTGAGCGAAATTATTCGCAGCGCGT GGGATTGGCATGCGAAAAGCAATCAGCAT
8	https://www.uniprot.org/uniprot/Q9L5C0.fasta	bradyrhizobium japonicum (Ep)	ATGGCGACCGTGCTGGTGACCGGCGGCGCGGGCTATATTGGCAGCCATATGGTGCATGCGCTGGTGGATGCGGGCGAAAGCGTAG TGGTGATTGATAACCTGAGCACCGGCTTTAGCGCGTTTCTGCCGGAAGGCGTGCCGCTGTTTATTGGCGATGCGGGCGATGAAAACC TGGTGGAAGGCGTGATTGCGCAGCATGGCATTGATAGCATTATTCATTTTGCGGGCAGCGTGGTGGTGCCGGATAGCATGCGCGATC CGCTGGGCTATTATCGCAACAACACCATGACCACCGGCAGCCTGCTGAACGCGGCGGTGAAAGGCGGCGTGAGCCGCTTTATTTTTA GCAGCACCGCGGCGGTGTATGGCAACCCGGATCGCGTGCCGGTGCCGGAAATTGCGCCGACCCGCCCGCTGAGCCCGTATGGCAGC AGCAAACTGATGACCGAAATTATGCTGCATGATGTGGCGAGCGCGCATGGCATGAGCTATGTGGTGCTGCGCTATTTTAACGTGGCG GGCGCGGATCCGAAAGGCCGCGTGGGCCTGGCGACCACCGGCGCGACCCATCTGCTGAAAATTGCGGTGGAAGCGGCGACCGGTC AGCGCGCGAAAATTGATGTGTTTGGCACCGATTATCCGACCCAAGATGGCAGCTGCATTCGCGATTTTATTCATGTGAGCGATCTGGT GGAAGCGCATCGCGCGGCGCTGAGCTATCTGCGCGCGGGCGGTCAGAGCGTGACCCTGAACTGCGGCTATGGCCGCGGCTATAGCG TGTTAGAAACCATTGAAGCGGTGCGCCGCGTGAGCATGCGCAACTTTGCGGTGGCGTATGCGGCGCGCCGCCCGGGCGATATTATG ACCATGGTGGCGGATACCACCCGCATTCGCAGCCTGCTGGATTGGACCCCGCGCTTTGATGATCTCGAAACCATTGCGAGCCATGCG CTGGCGTGGGAAGAAAAACTGTTTCGCGAACGCGCGGGCAGCGCGCGCCAAGCGGAAAGCGCG

9	https://www.uniprot.org/uniprot/O84903.fasta	lactobacillus casei (Ep)	ATGGCGACCATTGCGGTGCTGGGCGGCGCGGGCTATATTGGCAGCCATACCGTGAAACAGCTGCTGGCGGCGGGCGAAGATGTGG TGGTGCTGGATAACCTGATTACCGGCCATCGCAAAGCGGTGGATCCGCGCGCGCGCTTTTATCAAGGCGATATTCGCGATTATCATTT TCTGAGCCAAGTGTTTAGCCAAGAAAAAATTGATGGCATTGTGCATTTTGCGGCGTTTAGCATTGTGCCGGAAAGCATGAAAGATCC GCTGAAATATTTTGATAACAACACCGGCGGCATGATTACCCTGCTGGAAGCGATGAATCAGTTTGGCATTAAAAAAATTGTGTTTAGC AGCACCGCGGCGACCTATGGCGAACCGAAACAAGTGCCGATTAAAGAAACCGATCCGCAAGTGCCGACCAACCCGTATGGCGAAAG CAAACTGGCGATGGAAAAAATTATGCATTGGGCGGATGTGGCGTATGGCCTGAAATTTGTGGCGCTGCGCTATTTTAACGTGGCGGG CGCGATGCCGGATGGCAGCATTGGCGAAGATCATCATCCGGAAACCCATATTGTGCCGATTATTCTGCAAGTGGCGGCGGGCACCCG CACCGGCCTGCAGATTTATGGCGATGATTATCCGACCAAAGATGGCACCAACGTGCGCGATTATGTGCATGTGGTGGATCTGGCGGA TGCGCATATTCTGGCGCTGAAATATCTGGATGCGGGCAACAAAAGCAGCGCGTTTAACATTGGCAGCGCGCATGGCTTTAGCAACCT GGAAATTCTGAACGCGGCGCGCAAAGTAACCGGCCAAGAAATTCCGGCGACCATGGGCCCGCGCCGTGCGGGCGATCCGAGCACCC TGATTGCGAGCAGCGAAAAAGCGCGCGATATTCTGGGCTGGAAACCGAACTATGATGATATTGATAAAATTATTGAAACCGCGTGGA ACTGGCATGAAAACCATCCGGAAGGCTTTGGCGATCGCAAC
10	https://www.uniprot.org/uniprot/Q97EZ5.fasta	clostridium acetobutylicum (Ep)	ATGGCGAGCATTCTGGTGTGCGGCGGCGCGGGCTATATTGGCAGCCATACCGTGCATGAACTGATTGAAAGCGGCAAAGAAGTGGT GATTGTGGATAACCTGCAGAGCGGCCATCTGCAAGCGGTGAACAAAAAAGCGAAATTTTATAAAGGCGATATTCGCGATAGCGAATT TCTGGATAAAGTGTTTAGCGAAAACAAAATTCAGAGCATTATTCATTTTGCGGCGAACAGCCTGGTGGGCGAAAGCATGGAAAAACC GCTGCTGTATTTTAACAACAACGTGTATGGCATGCAGATTCTGCTGGAAAGCATGGTGAAACATAACATTAAAAACATTGTGTTTAGC AGCACCGCGGCGGTGTATGGCGAACCGAAAAAAATTCCGATTCTGGAAAGCGATGATACCAACCCGACCAACACCTATGGCGAAAC CAAACTGACCATGGAAAAAATGATGAAATGGAGCAGCAAAGCGTATGGCATTAACTATGTGGCGCTGCGCTATTTTAACGTGGCGG GCGCGCTGGATGATGGCAGCATTGGCGAAGATCATAACCCGGAAACCCATCTGATTCCGCTGATTCTGCAAGTGCCGCTGAAAAAAC GCCCGTTTATTACCGTGTTTGGCACCGATTATCCGACCCCGGATGGTACCTGCATTCGCGATTATATTCATGTGATTGATCTGGCGGCG GCGCATATTAAAGCGGTGGAATATCTGGAAAACGGCGGCAAAAGCAGCATTTTTAACCTGGGCAACGGCGTGGGCTTTAGCGTGAA AGAAATGATTGATGCGGCGCGCGAAGTGACCAAACAGAACATTAAAGTGGTGCTGGGCGAACGCCGCAGCGGCGATCCGGCGCAG CTGGTGGCGAGCAGTCAGAAAGCGAAAGAAGTGCTGGGCTGGAGCCCGAAATTTACCAACGTGAAAGATATTATTAAAGATGCGTG GAAATGGCATGAAACCCATCCGAACGGCTTTGATGCG
11	https://www.uniprot.org/uniprot/A0A3P4ASQ4.fasta	Thermus thermophilus (Ep)	ATGGCGCGCGTGCTGGTGACCGGAGGCGCGGGCTTTATTGGCAGCCATATTGTGGAAGATCTGCTGGCGCGCGGCCTGGAAGTGGC GGTGCTGGATAACCTGGCGACCGGCAAACGCGAAAACGTGCCGAAAGGCGTGCCGTTTTTTCAAGTGGATCTGCGCGATAAAGAAG GCGTGGAACGCGCGTTTCGCGAATTTCGCCCGACCCATGTGAGCCATCAAGCGGCGCAAGCGAGCGTGAAAGTGAGCGTGGAAGAT CCGGTGCTGGATTTTGAAGTCAACTTACTGGGAGGTCTGAACCTGCTGGAGGCTTGCCGTCAGTATGGCGTGGAAAAACTGGTGTTT GCGAGCACCGGCGGCGCGATTTATGGCGAAGTGCCGGAAGGCGAACGCGCGGAAGAAACCTGGCCCCCGCGCCCGAAAAGCCCGT ATGCGGCGAGCAAAGCGGCGTTTGAACATTATCTGAGCGTGTATGGTCAGAGCTATGGCCTGAAATGGGTGAGCCTGCGCTATGCG AACGTGTATGGCCCGCGCCAAGATCCGCATGGCGAAGCGGGCGTGGTGGCGATTTTTGCGGAACGCGTGCTGAAAGGCGAACCGGT GACCCTGTATGCGCGCAAAACCCCGGGCGATGAAGGCTGCGTGCGCGATTATATTTATGTGAAAGATGTGGCGGAAGCACATGCTCT GGCGCTGTTTAGCCTGGAAGGCATTTATAACGTGGGCACCGGCGAAGGCCATACCACGCGCGAAGTCTTAGAAGCGGTCGCGGAGG CCGCTGGGAAAGCTCCGCAAGTGCAGCCGGCGCCACCGCGCCCCGGCGATCTGGAACGCAGCGTGCTGAGCCCGCTGAAACTGATG GCGCATGGCTGGCGCCCGAAAGTGGGCTTTCAAGAAGGCATTCGCCTGACCGTGGATCATTTTCGCGGCGCG
12	https://www.uniprot.org/uniprot/Q9CIZ5.fasta	Lactococcus lactis IL1403 (Ep)	ATGGCGATTACCAAAAGCAACATTCTGATTACCGGCGGCGCGGGCTTTATTGGCAGCAGCCTGGCGAACGAACTGCTGCCGCAGAAC AAAATTACCGTGATTGATAACCTGAGCATGGGCGATTTTAACAACCTGCATGAAACGAGCAACCTGACCAAAATTCTGGGCGATGTG ACCGATAAAAACCTGCTGGTGAAAGTGCTGGAAGAAAACGATTTTGATTATATTTTTCATCTGGCGGCGATTGCGAGCGTGGCGGAT AGCGTGGCGCGCCCGTATGAAACCCATCAAGTGAACTTTGATAGCACCATGACCCTGCTGGAAATTCTGCGTCAGAACAAAAAAAGC CTGAAACGCTTTGTGTTTAGCAGCAGCGCGGCGGTGTATGGCGATGAACCGACCCTGCCGAAACAAGAAGAAGGCACCATTCGCCC GCTGACCCCGTATGCGATTGATAAATTTGCGAGCGAAAAAATGACCATGATTTATAACAACCTGTATGATGTGCCGACGAGCGCGAC CCGCTTTTTTAACGTGTATGGCCCGAATCAGAACCCGAACAGCCCGTATAGCGGCTTTATTAGCATTCTGGTGGATCGCCTGCGCGAA AACACCGAACTGACCATTTTTGGCGATGGCGAACAGAGCCGCGATTTTGTGTATATTGAAGATGTGATTCAAGCGCTGCTGCTGATTG CGACGAGCGAACAGAGCTTTGGCGAAGTGTATAACGTGGGCACCGGCGTGAAAAACAGCATTAACGATCTGACCAAATTTGCGCAG AAATTTACCAACAAAGAACTGAGCATTAAATTTGATGATGTGCGCCAAGGCGATATTAAAGATAGCGTGAGCGATATTAGCAAACTG AAAGATATTGGCTATAGCCCGAAATTTGATCTGAGCAACGGCATGAAAAAATATCTGAACTATGAATTTAAA
13	https://www.uniprot.org/uniprot/A0A2P1WNG8.fasta	Oceanobacillus iheyensis (Ep)	ATGGCGAGCATTCTGGTGCTGGGAGGCGCGGGCTATATTGGCAGCCATGCGGTGTATCAGCTGATTGAAAGCGGCAAAGATGTGGT GGTGGTGGATAACCTGCTGACCGGCCATAAAGCGGCGGTGCATCCGAAAGCGACCTTTTATAAAGGCGATATTCGCGAACGCGATTT TCTGCATGAAGTGTTTAGCAACGAAAGCATTGAAAGCGTGATTCATTTTGCGGCGAGCAGCCTGGTGGGCGAAAGCATGGATAAACC GCTGCAGTATTTTGATAACAACGTGTATGGCACGCAGATTCTGCTGGAAGAAATGGTGGAACATGGAGTGAAAGAAATTGTGTTTAG CAGCACCGCGGCGACCTACGGCGAGCCGGAGGCGGTGCCGATTACCGAAGCGATGCCGACCAACCCGGCGAGCACGTATGGCGAA ACCAAGCTGACCATGGAAAAAATGATGCGCTGGGCGGCGCAAGCGCATGGCATTCGCTATGTGAGCCTGCGCTATTTTAACGTGGCG GGCGCGCGCGCGAGCGGCGAAATTGGCGAAGATCATCGCCCGGAAACCCATCTGGTGCCGATTATTCTGCAAGCGGCGCTGGGTCA GCGCGATCATATTACCATTTTTGGCGAAGATTATGATACCGCGGATGGTACCTGCATTCGCGATTATGTGCATGTGGAAGATCTGATT GAAGCGCATCTGCTGGCGCTGGAATATCTGAAAAACGGCGGCAGCAGCGATATTTTTAACCTGGGCAGCAGCCAAGGCTTTTCTGTG AAAGAAATGATTGAAGCGGCGCGCACCGTGACCGGCAAAAAAATTCCGACGCAGAGCGGCCCGCGCCGCGCGGGCGATCCGGCGG TGCTGATTGCGAGCCCGGAAAAAGCGAAACGCATTCTGGGCTGGAACCCGAAACGCACCGCGATTCATACCATTATGCAAGATGCGT GGAACTGGCATACGAGCCATCCGAACGGCTATGAAGAAGGCGATAGCAAA

14	https://www.uniprot.org/uniprot/Q9M8S8.fasta	Arabidopsis thaliana (G1PP)	ATGGCGGATAACGATAGCCTGGATCAGTTTCTGGCGGCGGCGATTGATGCGGCGAAAAAAGCGGGTCAGATTATTCGCAAAGGCTT TTATGAAACCAAACATGTGGAACATAAAGGCCAAGTGGATTTAGTGACCGAAACCGATAAAGGCTGCGAAGAACTGGTGTTTAACCA TCTGAAACAGCTGTTTCCGAACCATAAATTTATTGGCGAAGAAACCACCGCGGCGTTTGGCGTGACCGAACTGACCGATGAACCGAC CTGGATTGTGGATCCGCTGGATGGCACCACCAACTTTGTGCATGGCTTTCCGTTTGTGTGCGTGAGCATTGGCCTGACCATTGGCAAA GTGCCGGTGGTGGGCGTGGTGTATAACCCGATTATGGAAGAACTGTTTACCGGCGTGCAAGGCAAAGGCGCGTTTCTGAACGGCAA ACGCATTAAAGTGAGCGCGCAGAGCGAACTGCTGACCGCGCTGCTCGTGACCGAAGCGGGCACCAAACGCGATAAAGCGACCCTGG ATGATACCACCAACCGCATTAACAGCCTGCTGACCAAAGTGCGCAGCCTGCGCATGAGCGGCAGCTGCGCGCTGGATCTGTGCGGCG TGGCGTGCGGCCGCGTGGATATTTTTTATGAACTGGGCTTTGGCGGCCCGTGGGATATTGCGGCGGGCATTGTGATTGTGAAAGAAG CGGGCGGCCTGATTTTTGATCCGAGCGGCAAAGATCTGGATATTACGAGTCAGCGCATTGCGGCGAGCAACGCGAGCCTGAAAGAA CTGTTTGCGGAAGCGCTGCGCCTGACCGGCGCG
15	https://www.uniprot.org/uniprot/Q54U72.fasta	Dictyostelium discoideum (G1PP)	ATGGCGGAAAACATTACCCTGGATCAGTATCTGCAGAGCGCGGTGGATGTGGTGAAAGAAATTGGCCCGATGATTCTGAAAAACTAT AACAGCCGCAGCAAACAGATTGAATATAAAGGCGCGATTGATCTGGTGACCGATACCGATAAAGCGGTGGAAGAACATATTATTAA AACCCTGACCACCAAATATCCGCATACCAAAATTCTGGGCGAAGAAAGCACCAAAGATGGCATTTATAACTGGGGCAACGAACCGAC CTGGGTGATTGATCCGATTGATGGCACCACCAACTTTGTGCATCGCTTTCCGCTGTTTTGCGTGAGCATTGCGCTGAGCATTAACAAA GAAATTGTGGTGGCGTGCCTGTATGCGCCGGTGCTGGATGAACTGTTTACCGCGACCAAAGGCGGCGGCGCGTTTCTGAACGGCGA AAGCATTAGCGTGAGCAGCGTGGAACATCTGAGTCAGAGCATTATTAGCACCAACGTGGGCTATGATCGCAGCGATAAAGGCATTG AATTTATGCTGACCAACTTTAAAAACATTCTGAAAGATAACGTGCAAGCGCTGCGCTTTAGCGGCACCGCGGCGTGGGAAATGGCGA GCGTGAGCTGCGGCCGCGTGGATAGCTTTTATGAATGGGGCATTCATCCGTGGGATATTGCTGCGGCGAGCCTGCTGATTACCGAAG CGGGCGGCGTGGTGGTGGATCCAAGCGGCGGCAAATGCGATATGGAAAGCCGCAAAGTGCTGTGCGGCAACCCGAACATTGTGAA CAAACTGAGCAAACTGCTGATTGAAAAACCGACGAGCAAC
16	https://www.uniprot.org/uniprot/Q19420.fasta	Caenorhabditis elegans (G1PP)	ATGGCGGTGTTTCAGCCGATTCATGAAGAAGAACAAGTGTTTGTGGATTATGCGATTGAACTGGTGAAAAAAGCGGGCACCCTGGT GCGCACCGCGTTTGATAGCCCGGAAAGCAAAGTGGATACCAAAAGCAGCAACACCGATCTGGTGACCGAAACCGATCAAGCGGTGG AAAAACTGCTGATTGAAGGCCTGAGCGAACGCTTTAAAGGCCATCGCTTTATTGGCGAAGAAAGCGTGGCGGGCGGCGCGAAAATT GAATGGACCGATGCGCCGACCTGGATTATTGATCCGATTGATGGCACCACCAACTTTGTGCATCGCATTCCGATGATTGCGATTTGCG TGGGCCTGGCGATTAAAAAACAGATTCGCGCGGGCATTGTGTATAACCCGATTACCAACGAACTGTATCTGGCGCAGCTGGGCAAAG GCGCGTTTAAAAACGGCTTTCCGATTCGCGCGAGCAAAAATCAGCTGCTGAGCAAAGGCGTGCTGTGTCAGAGCCTGGGCCTGCATA ACCGCGTGCAGTTTGGCGATCGCTGGCTGGATATTGCGCAGAGCAACATGCGCAACCAAGTGATGGCGGGCGTGCGCGGCCATCGC AGCTTTGGCAGCGCGGCGATTAACATGGTGATGGTGGCGCAAGGCAGCTGCGATGGCTATGTGGAATATGGCATTCATGCGTGGGA TGTGGCGGCGCCGAGCATTATTGTGACCGAAGCGGGCGGCGTGGTGACCGATCCGACCGGCAGCCCGTTTGATGTGATGAGCCGCA AAGTGCTGTGCGCGGGCACCGCGGAACTGGGCCGCGATCTGAGCGCGTGCCTGACCCATGTGGATTTTGAACCGGAAGCG
17	https://www.uniprot.org/uniprot/P94526.fasta	Bacillus subtilis (G1PP)	ATGGCGCGCATTATGGCGAGCCATGATACCCCGGTGAGCCCGGCGGGAATTCTGATTGATCTGGATGGCACCGTGTTTCGCGGCAAC GAACTGATTGAAGGCGCGCGCGAAGCGATTAAAACCCTGCGCCGCATGGGCAAAAAAATTGTGTTTCTGAGCAACCGCGGCAACATT AGCCGCGCGATGTGCCGCAAAAAACTGCTGGGCGCGGGCATTGAAACCGATGTGAACGATATTGTGCTGAGCAGCAGCGTGACCGC GGCGTTTCTGAAAAAACATTATCGCTTTAGCAAAGTGTGGGTGCTGGGCGAACAAGGCCTGGTGGATGAACTGCGCCTGGCGGGCG TGCAGAACGCGAGCGAACCGAAAGAAGCGGATTGGCTGGTGATTAGCCTGCATGAAACCCTGACCTATGATGATCTGAACCAAGCG TTTCAAGCGGCGGCGGGCGGCGCGCGCATTATTGCGACCAACAAAGATCGCAGCTTTCCGAACGAAGATGGCAACGCGATTGATGT GGCGGGCATGATTGGCGCGATTGAAACGAGCGCGCAAGCGAAAACCGAACTGGTGGTGGGCAAACCGAGCTGGCTGATGGCGGAA GCGGCGTGCACCGCGATGGGCCTGAGCGCGCATGAATGCATGATTATTGGCGATAGCATTGAAAGCGATATTGCGATGGGCAAACT GTATGGCATGAAAAGCGCGCTGGTGCTGACCGGCAGCGCGAAACAAGGCGAACAGCGCCTGTATACCCCGGATTATGTGCTGGATA GCATTAAAGATGTGACCAAACTGGCGGAAGAAGGTATTCTGATT
18	https://www.uniprot.org/uniprot/O29805.fasta	Archaeoglobus fulgidus (G1PP)	ATGGCGTTTAAACCGAAAGCGATTGCGGTGGATATTGATGGCACCCTGACCGATCGCAAACGCGCGCTGAACTGCCGCGCGGTAGA AGCGCTGCGCAAAGTGAAAATTCCGGTGATTCTGGCGACCGGCAACATTAGCTGCTTTGCGCGCGCGGCGGCGAAACTGATTGGCG TGAGCGATGTGGTGATTTGCGAAAACGGCGGCGTGGTGCGCTTTGAATATGATGGCGAAGATATTGTGCTGGGCGATAAAGAAAAA TGCGTGGAAGCGGTGCGCGTGCTGGAAAAACATTATGAAGTGGAACTGCTGGATTTTGAATATCGCAAAAGCGAAGTGTGCATGCG CCGCAGCTTTGATATTAACGAAGCGCGCAAACTGATTGAAGGCATGGGCGTGAAACTGGTGGATAGCGGCTTTGCGTATCATATTAT GGATGCGGATGTGAGCAAAGGCAAAGCGCTGAAATTTGTGGCGGAACGCCTGGGCATTAGCAGCGCGGAATTTGCGGTGATTGGC GATAGCGAAAACGATATTGATATGTTTCGCGTGGCGGGCTTTGGCATTGCGGTGGCGAACGCGGATGAACGCCTGAAAGAATATGC GGATCTGGTGACCCCGAGCCCGGATGGCGAAGGCGTGGTTGAAGCGCTGCAGTTTCTGGGCCTGCTGCGC

19	https://www.uniprot.org/uniprot/O67359.fasta	Aquifex aeolicus (G1PP)	ATGGCGCGCGTGATTCTGTTTGATCTGGATGGCACCCTGATTGATAGCGCGAAAGATATTGCGTTAGCGCTGGAAAAAACCCTGAAA GAACTGGGCCTGGAAGAATATTATCCGGATAACGTGACCAAATATATTGGCGGCGGCGTGCGCGCGCTGCTGGAAAAAGTGCTGAA AGATAAATTTCGCGAAGAATATGTGGAAGTGTTTCGCAAACATTATCTGGAAAACCCGGTGGTGTATACCAAACCGTATCCGGAAATT CCGTATACCCTGGAAGCGCTGAAAAGCAAAGGCTTTAAACTGGCGGTGGTGAGCAACAAACTGGAAGAACTGAGCAAAAAAATTCT GGATATTCTGAACCTGAGCGGCTATTTTGATTTAATTGTGGGCGGCGATACCTTTGGCGAAAAAAAACCGAGCCCGACCCCGGTGCT GAAAACCCTGGAAATTCTGGGCGAAGAACCGGAAAAAGCGCTCATTGTGGGCGATACCGATGCGGATATTGAAGCGGGCAAACGC GCGGGCACCAAAACCGCGCTGGCGCTGTGGGGCTATGTGAAACTGAACAGTCAGATTCCGGATTTTACCCTGAGCCGCCCGAGCGAT CTGGTGAAACTGATGGATAACCATATTGTGGAATTT
20	https://www.uniprot.org/uniprot/Q9HLQ2.fasta	Thermoplasma acidophilum (G1PP)	ATGGCGATTCGCCTGGCGGCGATTGATGTGGATGGCACCCTGACCGATCGCGATCGCCTGATTAGCACCAAAGCGATTGAAAGCATT CGCAGCGCGGAAAAAAAAGGCCTGACCGTGAGCCTGCTGAGCGGCAACGTGATTCCGGTGGTGTATGCGCTGAAAATTTTTCTGGG CATTAACGGCCCGGTGTTTGGCGAAAACGGCGGCATTATGTTTGATAACGATGGCAGCATTAAAAAATTTTTTAGCAACGAAGGCAC CAACAAATTTCTGGAAGAAATGAGCAAACGCACGAGCATGCGCAGCATTCTGACCAACCGCTGGCGCGAAGCGAGCACCGGCTTTG ATATTGATCCGGAAGATGTGGATTATGTGCGCAAAGAAGCGGAAAGCCGCGGCTTTGTGATTTTTTATAGCGGCTATAGCTGGCATC TGATGAACCGCGGCGAAGATAAAGCGTTTGCGGTGAACAAACTGAAAGAAATGTATAGCCTGGAATATGATGAAATTCTGGTGATT GGCGATAGCAACAACGATATGCCGATGTTTCAGCTGCCGGTGCGCAAAGCGTGCCCGGCGAACGCGACCGATAACATTAAAGCGGT GAGCGATTTTGTGAGCGATTATAGCTATGGCGAAGAAATTGGTCAGATTTTTAAACATTTTGAACTGATG
21	https://www.uniprot.org/uniprot/P38774.fasta	S. cerevisiae S288c (G1PP)	ATGGCGGAATTTAGCGCGGATCTGTGCCTGTTTGATCTGGATGGCACCATTGTGAGCACCACCGTGGCGGCGGAAAAAGCGTGGAC CAAACTGTGCTATGAATATGGCGTGGATCCGAGCGAACTGTTTAAACATAGCCATGGCGCGCGCACCCAAGAAGTGCTGCGCCGCTT TTTTCCGAAACTGGATGATACCGATAACAAAGGCGTGCTGGCGCTGGAAAAAGATATTGCGCATAGCTATCTGGATACCGTGAGCCT GATTCCGGGCGCGGAAAACCTGCTGCTGAGCCTGGATGTGGATACCGAAACGCAGAAAAAACTGCCGGAACGCAAATGGGCGATTG TGACGAGCGGCAGCCCGTATCTGGCGTTTAGCTGGTTTGAAACCATTCTGAAAAACGTGGGCAAACCGAAAGTGTTTATTACCGGCT TTGATGTGAAAAACGGCAAACCGGATCCGGAAGGCTATAGCCGCGCGCGCGATCTGCTGCGCCAAGATCTGCAGCTGACCGGCAAA CAAGATCTGAAATATGTGGTGTTTGAAGATGCGCCGGTGGGCATTAAAGCGGGCAAAGCGATGGGCGCGATTACCGTGGGCATTAC GAGCAGCTATGATAAAAGCGTGCTGTTTGATGCGGGCGCGGATTATGTGGTGTGCGATCTGACCCAAGTGAGCGTGGTGAAAAACA ACGAAAACGGCATTGTGATTCAAGTGAACAACCCGCTGACCCGCGCG
22	https://www.uniprot.org/uniprot/P53981.fasta	S. cerevisiae S288c PYP1 (G1PP)	ATGGCGGTGAAAGCGGTGATTTTTACCGATTTTGATGGCACCGTGACCCTGGAAGATAGCAACGATTATCTGACCGATACCCTGGGC TTTGGCAAAGAAAAACGCCTGAAAGTGTTTGAAGGCGTGCTGGATGATACCAAAAGCTTTCGCCAAGGCTTTATGGAAATGCTGGAG AGTATTCATACCCCGTTTCCGGAATGCATTAAAATTCTGGAAAAAAAAAATTCGCCTGGATCCGGGCTTTAAAGATACCTTTGAATGGG CGCAAGAAAACGATGTGCCGGTGATTGTGGTGAGCAGCGGCATGAAACCGATTATTAAAGTGCTGCTGACCCGCCTGGTGGGCCAA GAAAGCATTCATAAAATTGATATTGTGAGCAACGAAGTGGAAATTGATGCGCATGATCAGTGGAAAATTATTTATAAAGATGAAAGC CCGTTTGGCCATGATAAAAGCCGCAGCATTGATGCGTATAAAAAAAAATTTGAAAGCACCCTGAAAGCGGGCGAACAGCGCCCGGT GTATTTTTATTGCGGCGATGGCGTGAGCGATCTGAGCGCGGCGAAAGAATGCGATCTGCTGTTTGCGAAACGCGGCAAAGATCTGGT GACCTATTGCAAAAAACAGAACGTGCCGTTTCATGAATTTGATACCTTTAAAGATATTCTGGCGAGCATGAAACAAGTGCTGGCGGG CGAAAAAACCGTGGCGGAACTGATGGAAAAC
23	https://www.uniprot.org/uniprot/Q38UH2.fasta	L. sakei (strain 23K) (LAI)	ATGGCGCTGAACACCGAAAACTATGAATTTTGGTTTGTGACCGGCAGTCAGAGCCTGTATGGCGAAGAAACCCTGCGCAGCGTGGAA AAAGATGCGAAAGAAATTGTGGAAAAACTGAACGCGAGCCATCAGCTGCCGTATCCGATTGTGTTTAAACTGGTGGCGACCACCGCG GATAACATTACCAAAGTGATGAAAGAAGCGAACTATAACGATCATGTGGCGGGCGTGATTACCTGGATGCATACCTTTAGCCCGGCG AAAAACTGGATTCGCGGCACCAAACTGCTGCAGAAACCGCTGCTGCATCTGGCGACGCAGTTTCTGAACAAAATTCCGTATGATACCA TTGATTTTGATTATATGAACCTGAATCAGAGCGCGCATGGCGATCGCGAATATGCGTTTATTAACGCGCGCCTGCGCAAAAACAACAA AATTATTAGCGGCTATTGGGGCGATGAAGATGTGCAGAAAGCGATGGCGAAATGGATGGATGTGGCGGTGGCGTATAACGAAAGC TTTAAAATTAAAGTGGTGACCTTTGCGGATAAAATGCGCAACGTGGCGGTGACCGATGGCGATAAAGTGGAAGCGCAGATTAAATTT GGCTGGACCGTGGATTATTGGGGCGTGGGCGATCTGGTGGCGGAAGTGAACGCGGTGAGCGAAGCGGATATTGATGCGAAATATG CGGATCTGCAGAAAGAATATGATTTTGTGGAAGGTCAGAACACCCCGGAAAAATTTGAACATAACGTGAAATATCAGATTCGCGAAT ATTTTGGCCTGAAAAAATTTATGGATGATCGCGGCTATACCGCGTTTACCACCAACTTTGAAGATCTGGTGGGCCTGGAACAGCTGCC GGGCCTGGCGGCGCAGCTGCTGATGGCGGAAGGCTATGGCTTTGCGGGCGAAGGCGATTGGAAAACCGCGGCGCTGGATCGCCTG CTGAAAATTATGGCGCATAACGAAAAAACCGTGTTTATGGAAGATTATACCCTGGATCTGCGCCAAGGCCATGAAGCGATTCTGGGC AGCCATATGCTGGAAGTGGATCCGAGCATTGCGAGCGATAAACCGCGCGTGGAAGTGCATCCGCTGGATATTGGCGATAAAGATGA TCCGGCGCGCCTGGTGTTTACCGGCATGCAAGGCGATGCGGTGGATGTGACCATGGCGGATTATGGCGATGAATTTAAACTGATGA GCTATGATGTGCGCGGCAACAAACCGGAAGCGGATACCCCGCATCTGCCGGTGGCGAAACAACTGTGGACCCCGAAACAAGGCCTG CGCGAAGGCGCGGTAGGCTGGCTGACCGTGGGCGGCGGCCATCATACCGTGCTGAGCTTTGCGGTGGATAGCGAACAGCTGCAAG ATCTGAGCCATCTGTTTGATCTGACCTATGTGAACATTAAA

24	https://doi.org/10.1186/s12896-016-0286-5	Bacillus coagulans NL01 (LAI)	ATGGCGCTGAAAATTAAAGATTATGAATTTTGGTTTATTGTGGGCAGTCAGCATCTGTATGGCAAAGAAACCCTGGATGCGGTGAAA GAAGATGCGGAAAAAATTATTAACGAAATTAACGCTAGCGGCAAACTTCCGTATCCGATTGTGTTTAAAACCGTGGCGACCACCGCG GATAGCATTACCCAAGTGATGAAAGAAGTGAACTATAACGATAACGTGGCGGGCGTGATTACCTGGATGCATACCTTTAGCCCGGCG AAAAACTGGATTCGCGGCACCAAACTGCTGCAGAAACCGCTGCTGCATCTGGCGACGCAGTTTCTGGATCATATTCCGTGGGAAACC ATTGATATGGATTATATGAACCTGCATCAGAGCGCGCATGGCGATCGTGAATATGGCTTTATTAACGCGCGCCTGAACAAAAACAAC AAAATTGTGGTGGGCCATTGGAAAGATGAAAAAGTGCAGACGCAGATTGGTCAGTGGATGGATGTGGCGGTGGCGTTTAACGAAA GCTTTAACATTAAAGTGGCGCGCTTTGGCGATAACATGCGCAACGTGGCGGTGACCGATGGCGACAAAATTGAAGCGCAGATTCAGT TTGGCTGGACCGTGGATTATTATGGCATAGGCGACCTGGTAGCGGAAATGAAGGAAGTGACCGATCAAGAAATTGAAGCGGTGTGC GCGGAATGCCAAGAAAAATATGAACTGGTGGTGGGCAACAACGATCCGGCGTATTTTGAAGATCATGTGAAAGAACAGATTCGCAT TGAAATTGCGCTGCGCAAATTTCTGGATCGCGGCGGCTATACCGCGTTTACCACCAACTTTGAAGATCTGTGGGGCATGAAACAGCT GCCGGGCATGGCGGTGCAGCGCCTGAACGCGGAAGGCTATGGCTTTGCGGGCGAAGGCGATTGGAAAACCGCGGCGCTGGATCGC CTGGTGAAAATTATGGCGCATAACGAAAAAACCGGCTTTATGGAAGATTATACCTATAACCTGGTGAAAGGCCATGAAGAAATTCTG GGCAGCCATATGCTGGAAGTGGATCCGACCCTGGCTAGTGGCAAAATTCGCGTGGAAGTGCATCCGCTGGGTATTGGCGATCGCGA GGATCCGGCGCGCCTGGTGTTTGATGGCACAGACGGCGACGCGGTTAACCTGACCGTGAGCGATTTTGGCGATCAGTTTAAACTGGT GATGTATGAAGTGGATGGCAAAAAACCGGCGGAAGCGGCGCCGAAACTGCCGGTGGCGCGTCAGCTGTGGACCCCGAAACCGGGC TTTTATGAAGGCGTGCAGAAATGGATTGAAAACGGCGGCGGCCATCATACCGTGCTGAGCTTTGCGGTGACCGCGGAACAGATTGA AGATTTTGCGAAAATGGTGGGCCTGAAAACCGTGAAAATT
25	https://www.uniprot.org/uniprot/P04397.fasta	S. cerevisiae S288c (Ep)	ATGGCGACCGCGCAGCTGCAGAGCGAAAGCACGAGCAAGATTGTGCTGGTGACCGGCGGCGCGGGCTATATTGGCAGCCATACCGT GGTGGAACTGATTGAAAACGGCTATGATTGCGTGGTGGCGGATAACCTGAGCAACAGCACCTATGATAGCGTGGCGCGCTTAGAAG TGCTGACCAAACATCATATTCCGTTTTATGAAGTGGATCTGTGCGATCGCAAAGGCCTGGAAAAAGTGTTTAAAGAATATAAAATTGA TAGCGTGATTCATTTTGCGGGCCTGAAAGCGGTGGGCGAAAGCACGCAGATTCCGCTGCGCTATTATCATAACAACATTCTGGGCAC CGTGGTGCTGCTGGAACTGATGCAGCAGTATAACGTGAGCAAATTTGTGTTTAGCAGCAGCGCGACCGTGTATGGCGATGCGACCCG CTTTCCGAATATGATCCCGATACCGGAAGAATGCCCGCTGGGCCCGACCAACCCGTATGGCCATACCAAATATGCGATTGAAAACATT CTGAACGATCTGTATATAACAGCGATAAAAAAAGCTGGAAATTTGCGATTCTGCGCTATTTTAATCCGATAGGCGCGCATCCGAGCGGCC TGATTGGCGAAGATCCGCTGGGCATTCCGAACAACCTGCTGCCGTATATGGCGCAAGTGGCGGTGGGCCGCCGCGAAAAACTGTAT ATTTTTGGCGATGATTATGATAGCCGCGATGGCACCCCGATTCGCGATTATATTCATGTGGTGGATCTGGCGAAAGGCCATATTGCGG CGCTGCAGTATCTGGAAGCGTATAACGAAAACGAAGGCCTGTGCCGCGAATGGAACCTGGGCAGCGGCAAAGGCAGCACCGTGTTT GAAGTGTATCATGCGTTTTGCAAAGCGAGCGGCATTGATCTGCCGTATAAAGTGACCGGCCGCCGCGCGGGCGATGTGCTGAACCTG ACCGCGAAACCGGATCGCGCGAAACGCGAACTGAAATGGCAGACCGAACTGCAAGTGGAAGATAGCTGCAAAGATCTGTGGAAAT GGACCACCGAAAACCCGTTTGGCTATCAGCTGCGCGGCGTGGAAGCGCGCTTTAGCGCGGAAGATATGCGCTATGATGCGCGCTTTG TGACCATTGGCGCGGGCACCCGCTTTCAAGCGACCTTTGCGAACCTGGGCGCGAGCATTGTGGATCTGAAAGTGAACGGTCAGAGC GTGGTGCTGGGCTATGAAAACGAAGAAGGCTATCTGAACCCGGATAGCGCGTATATTGGCGCGACCATTGGCCGCTATGCGAACCG CATTAGCAAAGGCAAATTTAGCCTGTGCAACAAAGATTATCAGCTGACCGTGAACAACGGCGTGAACGCGAACCATAGCAGCATTGG CAGCTTTCATCGCAAACGCTTTCTGGGCCCGATTATTCAGAACCCGAGCAAAGATGTGTTTACCGCGGAATATATGCTGATTGATAAC GAAAAAGATACCGAATTTCCGGGCGATCTGCTGGTGACCATTCAGTATACCGTGAACGTGGCGCAGAAAAGCCTGGAAATGGTGTAT AAAGGCAAACTGACCGCGGGCGAAGCGACCCCGATTAACCTGACCAACCATAGTTATTTTAACCTGAACAAACCGTATGGCGATACC ATTGAAGGCACCGAAATTATGGTGCGCAGCAAAAAAAGCGTGGATGTGGATAAAAACATGATTCCGACGGGAAACATTGTGGATCG CGAAATTGCGACCTTTAACAGCACCAAACCGACCGTGCTGGGCCCGAAAAACCCGCAGTTTGATTGCTGCTTTGTGGTGGATGAAAA CGCGAAACCGAGTCAGATTAACACCCTGAACAACGAACTGACCCTGATTGTGAAAGCGTTTCATCCGGATAGCAACATTACCCTCGAA GTGCTGAGCACCGAACCGACCTATCAGTTTTATACCGGCGATTTTCTGAGCGCGGGCTATGAAGCGCGCCAAGGCTTTGCGATTGAA CCGGGCCGCTATATTGATGCGATTAACCAAGAAAACTGGAAAGATTGCGTGACCCTGAAAAACGGCGAAACCTATGGCAGCAAAATT GTGTATCGCTTTAGC
26	https://www.uniprot.org/uniprot/E8MF10.fasta	B. longum JCM 1217 (Ep)	ATGGCGACCACCGTGCTGGTGACCGGCGGCGCGGGCTTTATTGCGACCCATACCGATATTGAACTGCTGAACAAAGGCTATGATGTG ATTAGCGTGGATAACTATGGCAACAGCAGCCCGGTGGCGCTGGAACGCGTGGAACAGATTACCGGCAAACCGGTGAAACGCTATGA TGGCGATGTGCGCGATGAAGCGCTGATGGAACGCGTGTTTGCGGAAAACAACATTGATTGGGTGATTCATTTTGCGGGCCTGAAAG CTGTGGGCGAAAGCGTGGCGAAACCGATTGAATATTATGATAACAACCTGTATAGCACCCTGGTGCTGCTGAAAGTGATGAAAAAAC ATAACGTGAAAAAAATTATTTTTAGCAGCAGCGCGACCGTGTATGGCACCCCGAAAGAACTGCCGATTACCGAAGAAACCCCGACCG GCGGCACCACCAACCCGTATGGCACGAGCAAACTGTTTCAAGAACAGATTCTGCGCGATGTGCATGTGGCGGATCCGAGCTGGACCA TTGTGCTGCTGCGCTATTTTAACCCGGTGGGCGCGCATGAAAGCGGCCTGCTGGGCGAAGATCCGAAAGGCATTCCGGCGAACCTGA CCCCGTATGTGGCGAAAGTGGCCGTGGGCGAACTGAAAGAAGTGCAAGTGTATGGCGATGATTATGATACCCCGGATGGCACCGGC GTGCGCGATTATATTCATGTGGTGGATCTGGCGAAAGGCCATGTGGCGGTGATTGATCATATTGATAAAGAAGGCGTGTTTGTGTAT AACCTGGGCACCGGCCATGGCTATAGCGTGCTGGAAGTGATTAAAGCGTATGAAAAAGCGGCGGGCCATCCGATTCCGTATGCGATT AAACCGCGCCGCCCGGGCGATATTGCGGCGTGCTATGCGGATGCGAGCAAAGCGGAAAAAGAACTGGGCTGGAAAGCGGAACTGA CCATTGATGATATGGCGGCGAGCAGCCTGAACTGGCAGACCAAAAACCCGAACGGCTTTCGCGATGCGGAA
27	https://rest.uniprot.org/uniprotkb/P09147.fasta	E.coli galE (Ep)	ATGAGAGTTCTGGTTACCGGTGGTAGCGGTTACATTGGAAGTCATACCTGTGTGCAATTACTGCAAAACGGTCATGATGTCATCATTC TTGATAACCTCTGTAACAGTAAGCGCAGCGTACTGCCTGTTATCGAGCGTTTAGGCGGCAAACATCCAACGTTTGTTGAAGGCGATAT TCGTAACGAAGCGTTGATGACCGAGATCCTGCACGATCACGCTATCGACACCGTGATCCACTTCGCCGGGCTGAAAGCCGTGGGCGA ATCGGTACAAAAACCGCTGGAATATTACGACAACAATGTCAACGGCACTCTGCGCCTGATTAGCGCCATGCGCGCCGCTAACGTCAA AAACTTTATTTTTAGCTCCTCCGCCACCGTTTATGGCGATCAGCCCAAAATTCCATACGTTGAAAGCTTCCCGACCGGCACACCGCAAA GCCCTTACGGCAAAAGCAAGCTGATGGTGGAACAGATCCTCACCGATCTGCAAAAAGCCCAGCCGGACTGGAGCATTGCCCTGCTGC GCTACTTCAACCCGGTTGGCGCGCATCCGTCGGGCGATATGGGCGAAGATCCGCAAGGCATTCCGAATAACCTGATGCCATACATCG CCCAGGTTGCTGTAGGCCGTCGCGACTCGCTGGCGATTTTTGGTAACGATTATCCGACCGAAGATGGTACTGGCGTACGCGATTACAT CCACGTAATGGATCTGGCGGACGGTCACGTCGTGGCGATGGAAAAACTGGCGAACAAGCCAGGCGTACACATCTACAACCTCGGCG CTGGCGTAGGCAACAGCGTGCTGGACGTGGTTAATGCCTTCAGCAAAGCCTGCGGCAAACCGGTTAATTATCATTTTGCACCGCGTC GCGAGGGCGACCTTCCGGCCTACTGGGCGGACGCCAGCAAAGCCGACCGTGAACTGAACTGGCGCGTAACGCGCACACTCGATGAA ATGGCGCAGGACACCTGGCACTGGCAGTCACGCCATCCACAGGGATATCCCGATTAA

Table 6. Sources and Sequences for Additional Phosphatases Tested.

Homolog #	Source	Sequence Name	Optimized DNA Sequence
2	https://www.ncbi.nlm.nih.gov/protein/XP_012758874.1?report=genbank&log$=prottop&blast_rank=2&RID=NX9S2ZWC013	Acytostelium subglobosum LB1 (G1PP)	ATGGCGCAGACCAACATTACCCTGCAGCCGTATAACATTTATAAATATCGCATGACCGATATTAGCT ATACGCAGTATCTGAACAGCGCGGTGAACGTGGCGAAAAGCGTGGGCCCGCTGATTCTGGAAAAC TATAACAACCGCGATAAAAAAGTGGAATATAAAGGCGCGGTGGATCTGGTGACCGAAACCGATAA AAACGTGGAAGAAATTATTATTAAAACCCTGACCACCGAATATCCGCATACCAAAATTCTGGGCGA AGAAAGCACCAAAGATGGCGTGTATAACTGGGGCAACGAACCGACCTGGATTATTGATCCGATTG ATGGCACCACCAACTTTGTGCATCGCTTTCCGCTGTTTTGCGTGAGCATTGGCCTGAGCATTAACAA AGAAATTGTGGTGGGCTGCATTTATAGCCCGGTGCTGAACGAAATTTTTACCGCGACCAAAGGCG GCGGCAGCTATCTGAACGGCAACAAACTGCAAGTGACGAGCGTGGAAACCCTGCAGCAAGCGGT GGTGGCGACCAACGTGGGCTATGATCGCAGCGAATATGGCGTGGAATTTATGATGAGCAACCTGA AAGCGATTATGATGCAGAACGTGCAGAGCATTCGCTTTAGCGGCACCGCGGCGTGGGAAATGGCG AGCGTGGCGTGCGGCCGCCTGGATTGCTTTTATGAATGGGGCATTCATCCGTGGGATATTGCTGCG GCGAGCCTGATGATTACAGAAGCGGGCGGCGTGGTGGTGGATCCAAGCGGGGGCCCCGTGCAAA TGGAAGGCCGCCGCGTGCTGTGCGGCAACGCGACCATTTGCCAAGTGGTGAGCAAACTGCTGGGC GTGCGCAAACCGTAA
3	https://www.ncbi.nlm.nih.gov/protein/KYQ89637.1?report=genbank&log$=prottop&blast_rank=3&RID=NX9S2ZWC013	Tieghemostelium lacteum (G1PP)	ATGGCGATTAACAACATTAGCTATCAAGAATTTCTGCAAGCGGCGGTGGATGTGACCCAAGAAGT GAGCCCGCTGATTCTGAAAAACTATCATAGCCGCGAACGCAAAATTGAATATAAAGGCGCGGTGG ATCTGGTGACCGATACCGATAAAGCGGTGGAAGAACATATTATTAAAACCCTGACCACCAAATATC CGTTTACCAAAATTCTGGGCGAAGAAAGCACCAAAGATGGCATTTATGGCTGGACCGATGAACCG ACCTGGATTATTGATCCGATTGATGGCACCACCAACTTTGTGCATCGCTTTCCGCTGTTTTGCGTGA GCATTGCGCTGGCGTATAAAAAAGAAGTGGTGGTGGGCGTGATTTATTGCCCGACCCTGAACGAA ACCTTTACCGCGACCCTGGGCGGCGGCGCGTATCTGAACGGCGAAAAAATTAGCGTGAGCCAAGT GGATAACCTGCAGCATGCGGTGATTGCGACCAACGTGGGCTATGATCGCAGCAAAAGCGGCATTG ATTTTATGCTGGGCAACCTGAGCAACATTCTGAATCAGAACGTGCAGAGCATTCGCTTTAGCGGCA CCGCGGCGTGGGAAATGGCGAGCGTGAGCTGCGGCCGCATTGATTGCTTTTATGAATGGGGCATT CATCCGTGGGATATTGCGGCGGCGAGCATTCTGATTACCGAAGCGGGCGGCGTGGTGGTGGATCC GACCAACACCAAAGTGAACATGGAAAGCCGCAAAGTGCTGTGCGGCAACCCGAAAATTGTGGAAC TGGTGAGCAAACAGCTGCTGGAAAAACCGGGCCATCCGAAATAA
4	https://www.ncbi.nlm.nih.gov/protein/XP_004367135.1?report=genbank&log$=prottop&blast_rank=4&RID=NX9S2ZWC013	Cavenderia fasciculata (G1PP)	ATGGCGAGCACCCCGACCTTTGAACAGTATCTGGCGAGCGCGGTGGCGGTGGCGAAAGAAGTGA GCCCGATGGTGCTGGAAAACTATAACAACCGCGATAAAAAAGTGGAATATAAAGGCGCGGTGGAT CTGGTGACCGAAACCGATAAAGCGGTGGAAGCGCATATTATTAAACGCCTGACCGAACAGTATCC GGAAACCAAAATTATGGGCGAAGAAAGCACCAAAGATGGCGTGTATGGCTGGACCGATGAAGCG ACCTGGATTATTGATCCGATTGATGGCACCACCAACTTTGTGCATCGCTTTCCGCTGTTTTGCATTAG CATTGCGCTGAGCATTAACAAACAGATTGTGGTGGGCTGCATTTATAGCCCGGTGCTGGATGAACT GTTTACCGCGACCAAAGGCGGCGGCGCGTTTCTGAACGGTCAGCGCATGAGCGTGACGAGCGTGG ATAGCCTGCAGCAAGCGGTGGTGGCGACCAACGTGGGCTATGATCGCACCCCGTATGGCATTGAT TTTATGCTGACCAACTTTAAAAGCATTATGAGTCAGAACGTGCAGAGCCTGCGCTTTAGCGGCACC GCGGCGTGGGAAATGGCGAGCGTGGCGTGCGGCCGCCTGGATTGCTTTTATGAATGGGGCATTCA TCCGTGGGATATTGCGGCGGCGAGCATTATGATTACCGAAGCGGGCGGCGTGGTGGTGGATCCGA GCGGCGGCCCGGTGGATATGGAAGGCCGCCGCGTGCTGTGCGGCGGTCAGAAAATTGTGGATAT TGTGAGCAAACTGCTGATTCCGAAACAGCCGCCGACCGATCAGCAGTAA

5	https://www.ncbi.nlm.nih.gov/protein/KAF2073217.1?report=genbank&log$=prottop&blast_rank=5&RID=NX9S2ZWC013	Polysphondylium violaceum (G1PP)	ATGGCGATGGAAATTAGCTATGCGGATTATCTGCAGAGCGCGATTGATGTGGCGAAAGATATTGG CCCGACCATTCTGAAAAACTATAACAGCCGCGAACGCAAAATTGAATATAAAGGCGCGATTGATCT GGTGACCGATACCGATAAAGCGGTGGAAGAACATATTATTAAAACCCTGACCGGCAAATATCCGC ATACCAAAATTCTGGGCGAAGAAAGCACCAAAGATGGCATTTATAACTGGGGCAACGAACCGACC TGGATTATTGATCCGATTGATGGCACCACCAACTTTGTGCATCGCTTTCCGCTGTTTTGCGTGAGCA TTGCGCTGAGCATTAACAAAGAAATTGTGGCGGCGGTGGTGTTTTGCCCGGTGCTGAACGAACTGT TTACCGCGACCAAAGGCGGCGGCGCGTTTCTGAACGGCGAACAGATTTTTGTGAGCAGCGTGGAT ACCCTGAGCAACGCGATGGTGGCGACCAACGTGGGCTATGATCGCACCAAAGCGGGCATTGATCA TATGCTGGGCAACCTGGATAACATTCTGAATCAGAACGTGCAAGCGATTCGCTTTAGCGGCACCGC GGCGTGGGAAATGGCGAGCGTGAGCTGCGGCCGCCTGGATTGCTTTTATGAATGGGGCATTCATC CGTGGGATATTGCGGCGAGCACCCTGCTGATTCGCGAAGCGGGCGGCGTGGTGGTGGATCCGAG CGGCAGCGATGTGAACATGGAAGGCCGCCGCGTGCTGTGCGGCGGCAAACCGATTGTGGATCTGC TGAGCAAACTGCTGCTGCCGAAACCGAGCAGCAACTAA
6	https://www.ncbi.nlm.nih.gov/protein/XP_020429789.1?report=genbank&log$=prottop&blast_rank=6&RID=NX9S2ZWC013	Heterostelium album PN500 (G1PP)	ATGGCGGATACCCCGAACAACATTAGCTATGAACAGTATCTGCAGAGCGCGGTGAAAATTGCGCG CGAAGTGGGCCCGAAAATTCTGGAAAACTATAGCAACCGCGATAAAAAAGTGGAATATAAAGGCG CGGTGGATCTGGTGACCGAAACCGATAAAAACACCGAAGAAGCGATTATTAAAGCGCTGACCACC GAATATCCGCATACCAAAATTCTGGGCGAAGAAAGCACCAAAGATGGCATTTATAACTGGGGCAA CGAACCGACCTGGATTATTGATCCGATTGATGGCACCACCAACTTTGTGCATCGCTTTCCGCTGTTT TGCGTGGCGATTGCGCTGAGCATTAACAAAGAAATTGTGGTGGGCATGATTTATAGCCCGGTGCT GAACGAAATGTTTACCGCGACCAAAGGCGGCGGCAGCTATCTGAACGGCGTGAAACTGGATGTGA CCAACGTGGATAACCTGCAGCAAGCGGTGATTGCGACCAACGTGGGCTATGATCGCAGCGAATAT GGCGTGGAATTTATGCTGAGCAACATTAAAAACATTCTGAGTCAGAACGTGCAGAGCATTCGCATG AGCGGCACCGCGGCGTGGGAAATGGCGAGCGTGGCGTGCGGCCGCCTGGATTGCTTTTATGAATG GGGCATTCATCCGTGGGATATTGCGGCGAGCAGCATTCTGATCACCGAAGCGGGCGGCGTTGTGC TGGATCCGAGCGGCGGACCGGTGCAGATGGAAGGCCGCCGCGTGCTGTGCGGCAACCCGAAAAT TTGCGATCTGGTGAGCAAATGCCTGGGCGTGCGCAAACCGCTGTAA
NA	https://rest.uniprot.org/uniprotkb/P19926.fasta	E.coli agp	ATGAACAAAACGCTAATCGCCGCAGCTGTGGCAGGGATAGTTTTACTCGCTTCAAACGCTCAGGCA CAAACCGTACCGGAAGGCTATCAGCTACAGCAAGTGCTCATGATGAGCCGCCATAACTTACGTGCG CCGCTGGCGAACAATGGCAGTGTGCTGGAGCAGTCGACGCCGAATAAATGGCCAGAATGGGACG TCCCCGGTGGGCAACTCACCACCAAAGGTGGCGTGCTCGAAGTGTATATGGGCCATTACATGCGTG AATGGCTGGCAGAGCAGGGGATGGTGAAATCGGGGGAATGCCCGCCGCCGTACACCGTTTATGCC TATGCCAATAGTCTGCAACGTACCGTTGCGACCGCACAGTTCTTTATTACCGGCGCATTCCCGGGGT GTGATATTCCTGTGCATCACCAGGAAAAAATGGGCACTATGGACCCAACCTTTAACCCGGTGATCA CCGATGATTCCGCCGCATTCAGTGAACAGGCGGTGGCGGCAATGGAGAAAAGAGCTCAGCAAACTC CAGCTTACCGACAGCTACCAGCTACTGGAAAAAATCGTTAACTATAAAGATTCCCCTGCCTGTAAAG AGAAACAACAGTGTTCGCTGGTGGATGGCAAAAATACCTTTAGCGCCAAGTATCAACAAGAACCA GGTGTTTCCGGGCCGCTGAAAGTCGGCAACTCGCTGGTAGATGCGTTTACTTTGCAATATTACGAA GGTTTTCCGATGGATCAGGTGGCCTGGGGAGAAATCAAATCTGACCAGCAGTGGAAGGTGTTGTC GAAGCTTAAAAACGGCTACCAGGACAGCCTGTTTACCTCACCGGAAGTGGCGCGCAATGTTGCGA AACCGCTGGTCAGTTATATCGACAAAGCTCTGGTCACCGATCGCACCAGCGCACCGAAAATTACAG TGTTGGTTGGGCACGACTCCAACATTGCCTCTCTGTTAACGGCGCTGGATTTCAAACCGTATCAGTT GCATGACCAGAACGAACGCACGCCGATTGGCGGCAAAATCGTTTTCCAGCGTTGGCATGACAGCA AAGCCAATCGCGATTTGATGAAAATTGAATATGTGTATCAGAGTGCGGAACAGTTACGTAATGCCG ATGCGTTAACCCTGCAGGCTCCTGCGCAGCGTGTGACGCTGGAATTAAGCGGTTGCCCGATAGAC GCTGATGGTTTCTGCCCGATGGATAAGTTTGATAGCGTGTTGAATGAAGCGGTGAAATAA
NA	https://rest.uniprot.org/uniprotkb/P00634.fasta	E.coli PhoA	ATGAAACAAAGCACTATTGCACTGGCACTCTTACCGTTACTGTTTACCCCTGTGACAAAAGCCCGGA CACCAGAAATGCCTGTTCTGGAAAACCGGGCTGCTCAGGGCGATATTACTGCACCCGGCGGTGCTC GCCGTTTAACGGGTGATCAGACTGCCGCTCTGCGTGATTCTCTTAGCGATAAACCTGCAAAAAATA TTATTTTGCTGATTGGCGATGGGATGGGGGACTCGGAAATTACTGCCGCACGTAATTATGCCGAAG GTGCGGGCGGCTTTTTTAAAGGTATAGATGCCTTACCGCTTACCGGGCAATACACTCACTATGCGCT GAATAAAAAAACCGGCAAACCGGACTACGTCACCGACTCGGCTGCATCAGCAACCGCCTGGTCAA CCGGTGTCAAAACCTATAACGGCGCGCTGGGCGTCGATATTCACGAAAAAGATCACCCAACGATTC TGGAAATGGCAAAAGCCGCAGGTCTGGCGACCGGTAACGTTTCTACCGCAGAGTTGCAGGATGCC ACGCCCGCTGCGCTGGTGGCACATGTGACCTCGCGCAAATGCTACGGTCCGAGCGCGACCAGTGA AAAATGTCCGGGTAACGCTCTGGAAAAAGGCGGAAAAGGATCGATTACCGAACAGCTGCTTAACG CTCGTGCCGACGTTACGCTTGGCGGCGGCGCAAAAACCTTTGCTGAAACGGCAACCGCTGGTGAA TGGCAGGGAAAAACGCTGCGTGAACAGGCACAGGCGCGTGGTTATCAGTTGGTGAGCGATGCTG CCTCACTGAATTCGGTGACGGAAGCGAATCAGCAAAAACCCCTGCTTGGCCTGTTTGCTGACGGCA ATATGCCAGTGCGCTGGCTAGGACCGAAAGCAACGTACCATGGCAATATCGATAAGCCCGCAGTC ACCTGTACGCCAAATCCGCAACGTAATGACAGTGTACCAACCCTGGCGCAGATGACCGACAAAGCC ATTGAATTGTTGAGTAAAAATGAGAAAGGCTTTTTCCTGCAAGTTGAAGGTGCGTCAATCGATAAA CAGGATCATGCTGCGAATCCTTGTGGGCAAATTGGCGAGACGGTCGATCTCGATGAAGCCGTACA ACGGGCGCTGGAATTCGCTAAAAAGGAGGGTAACACGCTGGTCATAGTCACCGCTGATCACGCCC ACGCCAGCCAGATTGTTGCGCCGGATACCAAAGCTCCGGGCCTCACCCAGGCGCTAAATACCAAAG ATGGCGCAGTGATGGTGATGAGTTACGGGAACTCCGAAGAGGATTCACAAGAACATACCGGCAGT CAGTTGCGTATTGCGGCGTATGGCCCGCATGCCGCCAATGTTGTTGGACTGACCGACCAGACCGAT CTCTTCTACACCATGAAAGCCGCTCTGGGGCTGAAATAA

Table 7. Plasmid and Genomic Part Sequences

Part	Source	DNA Sequence
BCD18	(Mutalik et al., 2013)	GGGCCCAAGTTCACTTAAAAAGGAGATCAACAATGAAAGCAATTTTCGTACTGAAACATC TTAATCATGCGACGGAGCGTTTCTA
BCD14	(Mutalik et al., 2013)	GGGCCCAAGTTCACTTAAAAAGGAGATCAACAATGAAAGCAATTTTCGTACTGAAACATC TTAATCATGCGGTGGAGGGTTTCTA
BCD7	(Mutalik et al., 2013)	GGGCCCAAGTTCACTTAAAAAGGAGATCAACAATGAAAGCAATTTTCGTACTGAAACATCTTAATCATGCTGGGGAGGGTTT CTA
FAB46	(Mutalik et al., 2013)	GCTCAAAAAGAGTATTGACTTCGCATCTTTTTGTACCTATAATAGATTCAT
pCC1BAC	https://novoprolabs.com/vector/Vgm4tani	See Source
pET-28a(+)	https://www.addgene.org/vector-database/2565/	See Source

References

14.2: Classes of Monosaccharides. (2014, July 3). Chemistry LibreTexts. https://chem.libretexts.org/Courses/Sacramento_City_College/SCC%3A_Chem_309_-_General_Organic_and_Biochemistry_(Bennett)/Text/14%3A_Carbohydrates/14.2%3A_Classes_of_Monosaccharides

Axelsen, K. (2017). *Keto-L-tagatose (CHEBI:134275)*. https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:134275

Bell, L. N. (2015). Chapter 74—Tagatose Stability in Beverages as Impacted by Composition and Thermal Processing. In V. Preedy (Ed.), *Processing and Impact on Active Components in Food* (pp. 613–618). Academic Press. https://doi.org/10.1016/B978-0-12-404699-3.00074-3

Bertelsen, H., Jensen, B. B., & Buemann, B. (1999). D-Tagatose—A Novel Low-Calorie Bulk Sweetener with Prebiotic Properties. *Low-Calories Sweeteners: Present and Future*, *85*, 98–109. https://doi.org/10.1159/000059685

Bober, J. R., & Nair, N. U. (2019). Galactose to tagatose isomerization at moderate temperatures with high conversion and productivity. *Nature Communications*, *10*(1), 4548. https://doi.org/10.1038/s41467-019-12497-8

Bouffard, G. G., Rudd, K. E., & Adhya, S. L. (1994). Dependence of Lactose Metabolism upon Mutarotase Encoded in the gal Operon in Escherichia coli. *Journal of Molecular Biology*, *244*(3), 269–278. https://doi.org/10.1006/jmbi.1994.1728

Dai, Y., Li, M., Jiang, B., Zhang, T., & Chen, J. (2021). Whole-cell biosynthesis of d-tagatose from maltodextrin by engineered Escherichia coli with multi-enzyme co-expression system. Enzyme and Microbial Technology, 145, 109747. https://doi.org/10.1016/j.enzmictec.2021.109747

Donner, T. W., Magder, L. S., & Zarbalian, K. (2010). Dietary supplementation with d-tagatose in subjects with type 2 diabetes leads to weight loss and raises high-density lipoprotein cholesterol. *Nutrition Research (New York, N.Y.)*, *30*(12), 801–806. https://doi.org/10.1016/j.nutres.2010.09.007

Dotzel, M. (2002, December 2). *Food Labeling: Health Claims; D-tagatose and Dental Caries*. Federal Register. https://www.federalregister.gov/documents/2002/12/02/02-30474/food-labeling-health-claims-d-tagatose-and-dental-caries

Ensor, M., Banfield, A. B., Smith, R. R., Williams, J., & Lodder, R. A. (2015). Safety and Efficacy of D-Tagatose in Glycemic Control in Subjects with Type 2 Diabetes. *Journal of Endocrinology, Diabetes & Obesity*, *3*(1), 1065.

Flamholz, A., Noor, E., Bar-Even, A., & Milo, R. (2012). eQuilibrator—The biochemical thermodynamics calculator. *Nucleic Acids Research*, *40*(D1), D770–D775. https://doi.org/10.1093/nar/gkr874

Fry, J. C. (2012). 3—Natural low-calorie sweeteners. In D. Baines & R. Seal (Eds.), *Natural Food Additives, Ingredients and Flavourings* (pp. 41–75). Woodhead Publishing. https://doi.org/10.1533/9780857095725.1.41

Guerrero-Wyss, M., Durán Agüero, S., & Angarita Dávila, L. (2018). D-Tagatose Is a Promising Sweetener to Control Glycaemia: A New Functional Food. *BioMed Research International*, *2018*, 8718053. https://doi.org/10.1155/2018/8718053

Karp, P. D., Weaver, D., Paley, S., Fulcher, C., Kubo, A., Kothari, A., Krummenacker, M., Subhraveti, P., Weerasinghe, D., Gama-Castro, S., Huerta, A. M., Muñiz-Rascado, L., Bonavides-Martinez, C., Weiss, V., Peralta-Gil, M., Santos-Zavaleta, A., Schröder, I., Mackie, A., Gunsalus, R., … Paulsen, I. (2014). The EcoCyc Database. *EcoSal Plus*, *6*(1), ecosalplus.ESP-0009-2013. https://doi.org/10.1128/ecosalplus.ESP-0009-2013

Kim, J.-H., Lim, B.-C., Yeom, S.-J., Kim, Y.-S., Kim, H.-J., Lee, J.-K., Lee, S.-H., Kim, S.-W., & Oh, D.-K. (2008). Differential Selectivity of the Escherichia coli Cell Membrane Shifts the Equilibrium for the Enzyme-Catalyzed Isomerization of Galactose to Tagatose. *Applied and Environmental Microbiology*, *74*(8), 2307–2313. https://doi.org/10.1128/AEM.02691-07

Kim, P. (2004). Current studies on biological tagatose production using l-arabinose isomerase: A review and future perspective. *Applied Microbiology and Biotechnology*, *65*(3). https://doi.org/10.1007/s00253-004-1665-8

Lee, S.-H., Hong, S.-H., Kim, K.-R., & Oh, D.-K. (2017). High-yield production of pure tagatose from fructose by a three-step enzymatic cascade reaction. *Biotechnology Letters*, *39*(8), 1141–1148. https://doi.org/10.1007/s10529-017-2340-3

Liu, J.-J., Zhang, G.-C., Kwak, S., Oh, E. J., Yun, E. J., Chomvong, K., Cate, J. H. D., & Jin, Y.-S. (2019). Overcoming the thermodynamic equilibrium of an isomerization reaction through oxidoreductive reactions for biotransformation. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-09288-6

Mei, W., Wang, L., Zang, Y., Zheng, Z., & Ouyang, J. (2016). Characterization of an L-arabinose isomerase from Bacillus coagulans NL01 and its application for D-tagatose production. *BMC Biotechnology*, *16*(1), 55. https://doi.org/10.1186/s12896-016-0286-5

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., & Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, *10*(4), Article 4. https://doi.org/10.1038/nmeth.2404

Nam, Y.-W., Nishimoto, M., Arakawa, T., Kitaoka, M., & Fushinobu, S. (2019). Structural basis for broad substrate specificity of UDP-glucose 4-epimerase in the human milk oligosaccharide catabolic pathway of Bifidobacterium longum. *Scientific Reports*, *9*(1), 11081. https://doi.org/10.1038/s41598-019-47591-w

Rojo-Poveda, O., Barbosa-Pereira, L., Orden, D., Stévigny, C., Zeppa, G., & Bertolino, M. (2020). Physical Properties and Consumer Evaluation of Cocoa Bean Shell-Functionalized Biscuits Adapted for Diabetic Consumers by the Replacement of Sucrose with Tagatose. *Foods*, *9*(6), 814. https://doi.org/10.3390/foods9060814

Shintani, T. (2019). Food Industrial Production of Monosaccharides Using Microbial, Enzymatic, and Chemical Methods. *Fermentation*, *5*(2), Article 2. https://doi.org/10.3390/fermentation5020047

Synthesis of Tagatose_Chemicalbook. (2021). https://www.chemicalbook.com/Article/Synthesis-of-Tagatose.htm

Tagatose. (n.d.). Retrieved July 31, 2022, from https://www.bionity.com/en/encyclopedia/Tagatose.html

The FDHA [@TheFDHA]. (2018, September 18). *#FDHACares #Tips4Diabetes Glycemic index gives you an idea about how fast your body converts carbs in a food into glucose 2 foods w/same amount of carbohydrates can have different glycemic index numbers. The smaller the number, the less impact the food has on your blood sugar. Https://t.co/Q2PxapTto7* [Tweet]. Twitter. https://twitter.com/TheFDHA/status/1042065684740681729/photo/1

Vera, C., & Illanes, A. (2016). Chapter 3—Lactose-Derived Nondigestible Oligosaccharides and Other High Added-Value Products. In A. Illanes, C. Guerrero, C. Vera, L. Wilson, R. Conejeros, & F. Scott (Eds.), *Lactose-Derived Prebiotics* (pp. 87–110). Academic Press. https://doi.org/10.1016/B978-0-12-802724-0.00003-2

Walsh, M. K. (2007). 4—Immobilized enzyme technology for food applications. In R. Rastall (Ed.), *Novel Enzyme Technology for Food Applications* (pp. 60–84). Woodhead Publishing. https://doi.org/10.1533/9781845693718.1.60

Xu, Y. (2011). *Process for manufacturing tagatose* (World Intellectual Property Organization Patent No. WO2011150556A1). https://patents.google.com/patent/WO2011150556A1/en

Zumbé, A., Lee, A., & Storey, D. (2001). Polyols in confectionery: The route to sugar-free, reduced sugar and reduced calorie confectionery. *The British Journal of Nutrition, 85 Suppl 1*, S31-45. https://doi.org/10.1079/BJN2000260