Title
Microbial Production of Natural Sugars for A Healthier Future

Author
Zhang, Angela Xin Yu

Publication Date
2021

Peer reviewed|Thesis/dissertation

Microbial Production of Natural Sugars for A Healthier Future

By

ANGELA XIN YU ZHANG
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____
Dr. Shota Atsumi, Chair

_____
Dr. Xi Chen

_____
Dr. Justin Siegel

Committee in Charge

2021

Dedication

I would like to dedicate this dissertation to the following people: my advisor Dr. Shota Atsumi, who saw the potential in me and taught me poker wisdom of when to go all-in for our scientific pursuits; my collaborators and funding agencies that entrusted me to push forward our research goals; my family, for showing their love by bringing me home-cooked food; my partner John McArthur, who has always advocated for me and taught me how to do so for myself; and all my friends, faculty and staff at UC Davis that have supported me through the last five years of graduate school. With encouragement from this community, I developed interpersonal and technical skills that enabled me to succeed and be resilient through the highs and lows of life.

Table of Contents

Microbial Production of Natural Sugars for A Healthier Future

Abstract

As our society increases its demand for natural sugars in the food and health industry, we need to establish alternative methods for the sustainable and economical production of these chemical commodities. Over the last 30 years, *Escherichia coli* has emerged as a malleable micro-organism that can be engineered as an efficient whole cell catalyst for chemical production. The first chapter of this thesis describes the establishment of *E. coli* as a heterotrophic production host and provides a brief overview of compounds that are commonly produced in this organism. To overcome the limitations of compound extraction, difficulties in using traditional total synthesis schemes, and costly in-vitro enzymatic processes to produce carbohydrates, the second and third chapters describe the development of metabolic engineering strategies for a fucosylated human milk oligosaccharide, lactodifucotetraose (LDFT), and a rare sugar D-psicose. Lastly, the appendix discusses the characterization of mutations acquired during adaptive laboratory evolution processes that enhances *E. coli*'s tolerance towards isobutyl acetate.

By
Angela Xin Yu Zhang
Doctor of Philosophy in Chemistry
University of California, Davis
Professor Shota Atsumi, Chair

Chapter 1: Introduction – *Escherichia coli* as a Chemical Production Platform

1.1 Introduction

As of 2021, the world population has reached 7.9 billion and by 2050, the world population is anticipated to exceed 10 billion. Our society's increasing demands for petroleum-derived chemicals and fuels and natural products from plants and animals will eventually exceed the finite supply of fossil fuels, fresh water, and arable land available on Earth required to produce essential commodities. Due to this supply and demand conundrum, alternative avenues of sustainable chemical production need to be established to continue supporting our expanding population and global trade economy.

In the last 30 years, metabolic engineering of microorganisms has emerged as a viable method for producing a wide array of chemicals, such as amino acids[1], biofuels[2,3], polymer precursors[4,5], and pharmaceuticals[6,7]. This interdisciplinary field first began with combining the power of genetic engineering, knowledge of existing metabolic pathways and their enzymatic reactions, and expression of recombinant DNA to rationally modify an organism's native cellular processes to overproduce a desired chemical[8]. Subsequent advancements of synthetic biology tools[9,10], molecular cloning[11], and genome editing techniques[12,13] have helped to improve the design and construction of a desired chemical production pathway and host strains. Computational and experimental flux models[14–16] and omics studies[17–19] have also aided in understanding the effect of chemical production on native metabolic pathways and regulatory systems. The holistic consideration of all the described aspects enables cost-effective and sustainable biotransformation of inexpensive feedstocks into valuable products with high yield and rates[20].

1.2 *Escherichia coli* as a Heterotrophic Host

Of the prokaryotic microorganisms used in the metabolic engineering field, *E. coli* is one of the most well-studied hosts[21]. This Gram-negative facultative anaerobe is a workhorse chassis in the molecular biology field, owing to its fast growth rate on inexpensive culture media, lack of biofilm formation, ease of genetic manipulation, and its ability to express proteins encoded by recombinant DNA[22]. As a heterotroph, *E. coli* can feed off organic compounds to grow and replicate and its native cellular machinery has been utilized to convert simple sugar substrates into bulk and fine chemicals. Four *E. coli* strain isolates, K-12, B, C and W, are classified under the biosafety Risk Group 1, meaning they are non-pathogenic and are safe for biotechnology applications. Extensive characterization of *E. coli*'s physiology[23,24], genome[25–27], transcriptome[28,29], proteome[28,30,31], and metabolome[32,33] has led to the curation of comprehensive organism databases such as EcoCyc[34], PortEco[35], ECMDB[36], and KEGG[37], which are invaluable resources to the metabolic engineering field. Along with the communal sourcing of recombinant DNA parts and genetic engineering tools through iGEM and Addgene[38], *E. coli* is inarguably the most favored host organism for exploratory research in academia and the biotechnology industry.

1.3 Synthetic Biology Parts

Efficient chemical production in *E. coli* requires methods to control the expression of desired genes. Plasmid expression vectors enables metabolic engineers to regulate them on a transcriptional and translational level. They contain the minimum of the following elements: promoter, ribosomal binding site, terminator, origin of replication and a selective marker. Promoters enable transcription of a desired gene and are available

in various forms: chemically-inducible promoters, such as the isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible P_{LlacO1} and anhydrotetracycline (aTc)-inducible P_{LtetO1}, which allow tunable expression of genes by varying the concentration of the small molecule inducer[39]; light, temperature and pH-activated promotersthat enable gene expression by altering culturing conditions without the addition of additional chemicals that may increase the cost of fermentation or complicate downstream purification processes[40–42]; constitutive promoters that are recognized by *E. coli*'s σ^{70} or σ^{S} RNA polymerase for gene expression during the exponential or stationary growth phases[43]. Libraries of ribosomal binding sites, which are translation initiation sequences, have also been engineered and measured with 1000-fold dynamic range[44,45]. Terminators, which serves to stop transcription of a desired gene and to prevent unintended transcription of neighboring gene sequences, are available in rho-independent and rho-dependent options[46,47]. Origins of replication are AT-rich DNA sequences that recruit transcription factors for the propagation of expression vectors to daughter cells and control the number of copies of the vector a cell carries[48]. Antibiotic resistance genes are often used as selection markers, which incentivize *E. coli* to maintain the replication of the vector when grown in the presence of the respective antibiotic[49]. In special cases when antibiotics cannot be used, growth-essential genes are used in the presence of a gene product inhibitor[50,51].

1.4 Molecular Cloning Strategies

Various DNA construction methods have aided the personalization of gene expression vectors. The most classic technique is restriction enzyme cloning, where endonucleases that recognize specific DNA sequences are used to create 5', 3', or

blunt overhangs for T4 DNA ligase to fuse together in-vitro[52]. Gateway cloning takes advantage of the lambda phage integration *attB* recognition sites for site-specific recombination[53]. Gibson cloning allows for assembling multiple overlapping DNA inserts with a mixture of a 5' exonuclease, a DNA polymerase and a DNA ligase in a single isothermal reaction[54]. Golden Gate Assembly uses Type IIS restriction enzymes that cut outside their recognition sites to avoid the retention of restriction enzyme recognition sites in a final vector[55]. Sequence and ligation-independent cloning overcomes the need of recognition sites by employing a 3' exonuclease, T4 DNA polymerase, to generate 5' overhangs in insert and vector sequences[56]. These fragments anneal in vitro and are transformed into *E. coli* for ligation. QuikChange™ site-directed mutagenesis uses PCR amplification to create small base pair replacements, deletions and insertions to a region of interest in a vector and a mixture of a kinase, ligase and restriction enzyme DpnI to re-circularize the PCR product[57]. Cost, time, complexity of desired recombinant DNA modifications all should be considered in identifying the best fit cloning technique to use.

1.5 Genomic Modification Techniques

Historically, chemical mutagens, UV radiation, and transposon mutagenesis have been used to generate random mutations as a means of strain improvement[58,59]. These strategies have the disadvantages of undesired genomic mutations that may lead to growth hindrance and unpredictable phenotypes. In targeted metabolic engineering to reduce carbon flux into an undesired metabolic pathway or to stabilize the expression of heterologous genes, precise and predictable genome engineering is required for host modifications. *In vivo* homologous recombination genomic modification techniques have

been established to delete or insert genes of interest using phage-encoded recombinases or endonucleases[12]. The λ-Red recombinase system employs three enzymes, Exo, Beta, and Gam, respectively, to digest, stabilize and protect the integration of a double-stranded linear DNA substrate, which encodes for a selective marker flanked two flippase-recognition sequences and homology sequences to a genomic region of interest[60]. Expression of a flippase subsequently removes the selective marker, leaving behind a short DNA sequence encoding for a nonfunctional peptide. This technique has been used in the creation of the Keio Collection, a library of 3985 single-gene deletions of all nonessential genes in *E. coli* strain BW25113 that has helped to assess unknown gene functions and regulatory networks[61].

For seamless and directed editing, the type II CRISPR/Cas system from *Streptococcus pyogenes* has been modified into a CRISPR/Cas9-mediated genomic engineering technology[62]. Cas9, an endonuclease that recognizes NGG DNA sequences, is directed to a genomic region of interest by a small guide RNA with 20 bp homology to the target location. The endonuclease creates a double strand DNA break and in the presence of a double strand linear DNA repair fragment that shares homology to the cleaved genomic region, *E. coli*'s natural DNA repair machinery integrates the DNA via homologous recombination. The efficiency of the system was further enhanced by combining the expression of Exo, Beta, and Gam of the λ-Red recombinase system to stabilize the donor DNA[63].

1.6 Chemicals Produced in *Escherichia coli*

1.6.1 Amino acids

Amino acids are organic monomer building blocks of proteins. Of the 20 L-amino acids that are required for proper human biological function, 11 can be endogenously made[64]. The remaining 9 amino acids termed as essential amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) must be obtained through food. Since the advent discovery of naturally abundant L-glutamate production in *Corynebacterium* in 1957[65], microorganisms have been harnessed to overproduce amino acids as additives to livestock feed, dietary supplements, cosmetics, and as precursors for chemical synthesis[66–69].

E. coli's native amino acid production pathways are tightly controlled via enzymatic feedback inhibition and transcriptional activators and repressors[70] and a significant amount of work has been focused on deregulating these bottlenecks for amino acid production by overexpressing feedback resistant enzymes and deleting transcriptional regulators[71–73]. From the pentose phosphate pathway, L-histidine can be produced up to 66.5 g/L under fed-batch fermentation, with a productivity of 1.5 g/L/h[74]. Derived from chorismate of the shikimate pathway, aromatic amino acids L-phenylalanine, L-tryptophan and L-tyrosine have been produced at 72.9 g/L[75], 44.0 g/L[76], and 43.14 g/L[77], respectively, in fed-batch fermentation. Recent work on metabolic engineering of *E. coli* to produce L-valine, a hydrophobic branched-chain amino acid, used transcriptomic analysis to identify genes of L-valine global regulators and exporters for overexpression and in silico knockout simulations to determine genes to remove for enhancing pyruvate and NADPH levels. 7.6 g/L L-valine was produced in the mutant[78]. Subsequent flux balance analysis identified ATP availability as a limiting factor for L-valine production and by employing fed-batch culturing strategy with acetate

assimilation for ATP production, 32.3 g/L L-valine was produced[79]. Concerns over the environmental impacts of chemically producing the sulfur-containing L-methionine have also pushed advancements in *E. coli* fermentation to improve titers to 12.8 g/L[80].

1.6.2 Biofuels

Fossil fuels are energy-rich hydrocarbons produced from the anaerobic decomposition of ancient organisms. Found in the forms of petroleum, coal, and natural gas, the combustion of these hydrocarbons produces 85% of the world's energy required for transportation and heating[81]. While fossil fuels have played an instrumental role in the revolutionizing the global industry, it is a nonrenewable resource and its combustion releases carbon dioxide into the atmosphere that is driving global climate change[82]. To phase-out the use of fossil fuels in our society, alternative production methods of fuel sources not only need to be renewable, sustainable, but also cost-effective to incentivize the global industry in this energy transition.

In the early 1990s, researchers took interested in metabolic engineering of *E. coli* to produce biofuels due to its natural ability to produce ethanol, a highly combustible solvent. As a first-generation biofuel, ethanol was initially synthesized from corn-derived glucose through *E. coli*'s native anaerobic fermentation process at a low yield of 0.26 g/g of glucose. This endogenous pathway is inefficient due to the formation of acetate and formate as byproducts and the imbalanced use of nicotinamide adenine dinucleotide (NADH)[83]. By installing a heterologous fermentation pathway consisting of pyruvate decarboxylase *pdc* and alcohol dehydrogenase II *adhB* from *Zymomonas mobilis*, acetate and formate production was eliminated and ethanol production reached the theoretical maximum production yield of 0.54 g/g of glucose[84]. With this established

bioethanol platform, researchers transitioned to using lignocellulosic biomass from agricultural waste as a cheaper and more sustainable starting material to avoid the competition of using edible food sources for fuel production[85]. In the second generation bioethanol production, enzymatic hydrolysis of lignocellulose releases hexoses, glucose and galactose, and pentoses, xylose and arabinose, that can be catabolized by *E. coli*[85]. From mixed sugar feeding, engineered ethanologenic recombinant *E. coli* can produce up to 45 g/L of ethanol[86].

Longer-chained alcohols, such as 1-propanol, isopropanol, 1-butanol, and isobutanol have also been production targets due to their higher blending capability, compatibility to combustion engines, and versatile industrial application[87–90]. In fermentative and non-fermentative pathways, acetate and 2-ketoacids produced from L-threonine, L-valine, and L-leucine amino acid pathways can be converted into their respective alcohol product by 2-keto-acid decarboxylases and alcohol dehydrogenases[88,91]. Recombinant *E. coli* strains have successfully produced 1-propanol at a yield of 0.15 g/g of glucose[92], isopropanol at 0.26 g/g of glucose[91], 1-butanol at 0.066 g/g of glucose[90], and isobutanol at 0.35 g/g of glucose[88].

1.3.3 Bioplastic Polymer Precursors

By 2050, plastic production will account for 20% of global annual oil consumption, coming in second to energy production[93]. Due to concerns over the atmospheric pollution of oil refinement and environmental contamination of microplastic degradation, bio-based plastics are promising renewable and biodegradable alternatives. Polyhydroxybutyrate (PHB), a natural accumulating polyester under conditions of excess carbon or limited oxygen, nitrogen, phosphorus and sulfur, has

similar material properties as petroleum-derived plastics polyethylene and polypropylene[4]. Although natural producers *Ralstonia eutropha* and *Alcaligenes eutropha* can accumulate up to 90% of their dry cell weight in PHB, they exhibit poor fermentation growth rates, which makes them unsuitable hosts for industrial application[94]. Due to its fast growth rate and robust production of acetyl-CoA, recombinant *E. coli* expressing three PHB biosynthetic genes *phaABC*, which encode for a 3-ketothiolase, acetoacetyl-CoA reductase, and a PHB synthase, from *Alcaligenes eutropha* produced 80 g/L of PBH from 20 g/L glucose within 42 h under pH-stat fed-batch conditions[95].

Polybutylene succinate (PBS) is also another biodegradable thermoplastic polymer with comparable flexibility and thermostability to polypropelene[96]. Although there currently is no natural biosynthetic pathway identified for PBS, it can be chemically synthesized by polycondensation of two monomers succinic acid and 1,4-butanediol. Succinate, a dicarboxylic acid metabolite of the citric acid cycle, is natively produced in small quantities along with lactate, formate, acetate and pyruvate under anaerobic conditions. Genome modifications to remove fermentation of organic acids, installation of a glyoxylate shunt to circumvent α-ketoglutarate production, and expression of pyruvate carboxylase *pyc* from *Lactococcus lactis* has enabled accumulation of succinate under aerobic conditions up to 58 g/L from glucose[97]. Unlike succinate, 1,4-butandiol is not a natural metabolite and required biopathway prediction algorithms and genome-scale metabolic modeling of *E. coli* to elucidate its production from common metabolic intermediates[98]. An anaerobic fermentation route through the citric acid cycle

was chosen to provide NAD(P)H for the reduction of intermediates into 1,4-butandiol and the optimized strain produced 18 g/L of 1,4-butandiol[98].

1.6.4 Biopharmaceuticals

Aside from small molecule products, *E. coli* is also an efficient production host for therapeutic proteins. During the late 1970s, the historical production of recombinant human insulin in *E. coli* was a hallmark breakthrough in medicine and revolutionized the access to Type I and II diabetes treatment[99]. Traditionally extracted from cattle and bovine pancreas, insulin production was inefficient, costly, burdened the food-supply chain, and caused side effects due to the slight differences in amino acid residues between human and animal protein constructs[99,100]. Plasmid-based expression of the A and B chains and C-terminal peptide of human insulin in *E. coli* produces the proinsulin precursor at 46 g/L in high density fed-batch cultures. Downstream enzymatic removal of the C-peptide and solubilization of the protein inclusion bodies results in 1.3 mg of bioactive insulin per gram of proinsulin[101]. Work in reducing protein aggregation and construction of properly folded A and B chains without C-peptide assistance has increased insulin production to 520 mg/L[102].

Techniques used in the industrial production of insulin have since paved the way for other pharmaceutical peptides and proteins for the treatment of various endocrine-related diseases. Human growth hormone, a short-chained peptide previously obtained from human cadaver tissue to stimulate proper muscle development, is produced through recombinant *E. coli* expression at 100 mg/L[103]. Interleukin 1 receptor antagonist, a non-glycosylated cytokine used in treating rheumatoid arthritis, is produced at 0.43 g/L in 1 L bioreactors and up to 12 g in optimized 50 L fermentation

runs[104]. Large scale fermentation and purification of human parathyroid hormone, which is used to increase bone density to fight osteoporosis, can produce over 300 mg/L of functional protein up to 99% purity[105].

1.6.5 Carbohydrates

While sugars are commonly degraded as substrates for small molecule chemical production, they can also be converted into complex carbohydrates. Gram-negative bacteria like *E. coli* natively encode for the production of two cell wall polymers: colanic acid, an extracellular polysaccharide comprised of D-glucose, L-fucose, D-glucuronate, and D-galactose[106]; and murein, composed of glycosaminoglycans interlinked with short peptides[107]. Synthesis of colanic acid and murein requires sugars to be in their nucleotide-activated forms for their polymerization and all respective sugar building blocks can be derived from D-glucose-6-phosphate. Researchers have taken advantage of these native nucleotide-sugar pathways to synthesize hyaluronic acid, a lubricating biopolymer that can be applied to the health and cosmetics fields. With recombinant expression of hyaluronic acid synthase from *Streptococcus pyogenes* and overexpression of the native UDP-glucuronic acid and UDP-N-acetyl-glucosamine production pathways, hyaluronic acid is produced at 190 mg/L[108] from glucose.

In recent work, *E. coli* has also been engineered to produce chondroitin sulfate, a sulfated glycosaminoglycan found in joint tissues used to treat osteoarthritis[109]. Like hyaluronic acid production, two nucleotide-activated sugars UDP-glucuronic acid and UDP-N-acetyl-galactosamine are required for the synthesis of the chondroitin precursor[110]. Overexpression of UDP-N-acetyl-glucosamine epimerase and chondroitin synthase from *E. coli* strain O5:K4:H4 converts UDP-N-acetyl-glucosamine into UDP-N-

acetyl-galactosamine and condenses UDP-glucuronic acid and UDP-N-acetyl-galactosamine into chondroitin, respectively[110]. With recombinant expression of human chondroitin-4-O-sulfotransferase, 27 μg of chondroitin sulfate per gram of *E. coli* dry cell weight was produced[110].

To expand the portfolio of carbohydrate synthesis in *E. coli*, the production of human milk oligosaccharides and rare sugars are explored in Chapters 2 and 3 of this dissertation.

1.7 References

1. D'Este, M., Alvarado-Morales, M. & Angelidaki, I. Amino acids production focusing on fermentation technologies - A review. *Biotechnol. Adv.* **36**, 14–25 (2018).
2. Liao, J. C., Mi, L., Pontrelli, S. & Luo, S. Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nat. Rev. Microbiol.* **14**, 288–304 (2016).
3. Peralta-Yahya, P. P. & Keasling, J. D. Advanced biofuel production in microbes. *Biotechnol. J.* **5**, 147–162 (2010).
4. Jambunathan, P. & Zhang, K. Engineered biosynthesis of biodegradable polymers. *J. Ind. Microbiol. Biotechnol.* **43**, 1037–1058 (2016).
5. Yang, J. E., Choi, S. Y., Shin, J. H., Park, S. J. & Lee, S. Y. Microbial production of lactate-containing polyesters. *Microb. Biotechnol.* **6**, 621–636 (2013).
6. Sanchez-Garcia, L. *et al.* Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microb. Cell Fact.* **15**, 33 (2016).
7. Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E. & Villaverde, A. Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* **8**, 17 (2009).
8. Woolston, B. M., Edgar, S. & Stephanopoulos, G. Metabolic engineering: past and future. *Annu. Rev. Chem. Biomol. Eng.* **4**, 259–288 (2013).
9. Yang, J., Kim, B., Kim, G. Y., Jung, G. Y. & Seo, S. W. Synthetic biology for evolutionary engineering: from perturbation of genotype to acquisition of desired phenotype. *Biotechnol. Biofuels* **12**, 113 (2019).
10. McCarty, N. S. & Ledesma-Amaro, R. Synthetic Biology Tools to Engineer Microbial Communities for Biotechnology. *Trends Biotechnol.* **37**, 181–197 (2019).
11. Ashwini, M., Murugan, S. B., Balamurugan, S. & Sathishkumar, R. Advances in molecular cloning. *Mol. Biol.* **50**, 1–6 (2016).
12. Sathesh-Prabu, C. & Lee, S. K. Genome Editing Tools for Escherichia coli and Their Application in Metabolic Engineering and Synthetic Biology. *Emerging Areas in Bioengineering* 307–319 (2018) doi:https://doi.org/10.1002/9783527803293.ch17.
13. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).

14. Weaver, D. S., Keseler, I. M., Mackie, A., Paulsen, I. T. & Karp, P. D. A genome-scale metabolic flux model of Escherichia coli K–12 derived from the EcoCyc database. *BMC Syst. Biol.* **8**, 79 (2014).
15. Fang, X., Lloyd, C. J. & Palsson, B. O. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev. Microbiol.* **18**, 731–743 (2020).
16. Long, C. P. & Antoniewicz, M. R. Metabolic flux analysis of Escherichia coli knockouts: lessons from the Keio collection and future outlook. *Curr. Opin. Biotechnol.* **28**, 127–133 (2014).
17. Bardozzo, F., Lió, P. & Tagliaferri, R. A study on multi-omic oscillations in Escherichia coli metabolic networks. *BMC Bioinformatics* **19**, 194 (2018).
18. Kim, M., Rai, N., Zorraquino, V. & Tagkopoulos, I. Multi-omics integration accurately predicts cellular state in unexplored conditions for Escherichia coli. *Nat. Commun.* **7**, 13090 (2016).
19. Ishii, N. & Tomita, M. Multi-Omics Data-Driven Systems Biology of E. coli BT - Systems Biology and Biotechnology of Escherichia coli. in (ed. Lee, S. Y.) 41–57 (Springer Netherlands, 2009). doi:10.1007/978-1-4020-9394-4_3.
20. Ko, Y.-S. *et al.* Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem. Soc. Rev.* **49**, 4615–4636 (2020).
21. Blount, Z. D. The unexhausted potential of E. coli. *Elife* **4**, e05826 (2015).
22. Pontrelli, S. *et al.* Escherichia coli as a host for metabolic engineering. *Metab. Eng.* **50**, 16–46 (2018).
23. Alteri, C. J. & Mobley, H. L. T. Escherichia coli physiology and metabolism dictates adaptation to diverse host microenvironments. *Curr. Opin. Microbiol.* **15**, 3–9 (2012).
24. Guennadi, S., Danièle, J.-P. & Richard, D. Escherichia coli Physiology in Luria-Bertani Broth. *J. Bacteriol.* **189**, 8746–8749 (2007).
25. Riley, M. *et al.* Escherichia coli K-12: a cooperatively developed annotation snapshot—2005 . *Nucleic Acids Res.* **34**, 1–9 (2006).
26. A., G. E. C. *et al.* The Essential Genome of Escherichia coli K-12. *MBio* **9**, e02096-17 (2021).
27. Schneider, D. *et al.* Genomic comparisons among Escherichia coli strains B, K-12, and O157:H7 using IS elements as molecular markers. *BMC Microbiol.* **2**, 18 (2002).
28. Marisch, K. *et al.* A Comparative Analysis of Industrial Escherichia coli K–12 and B Strains in High-Glucose Batch Cultivations on Process-, Transcriptome- and Proteome Level. *PLoS One* **8**, e70516 (2013).
29. Sastry, A. V *et al.* The Escherichia coli transcriptome mostly consists of independently regulated modules. *Nat. Commun.* **10**, 5536 (2019).
30. Han, M.-J. Exploring the proteomic characteristics of the Escherichia coli B and K-12 strains in different cellular compartments. *J. Biosci. Bioeng.* **122**, 1–9 (2016).
31. Soufi, B., Krug, K., Harst, A. & Macek, B. Characterization of the E. coli proteome and its modifications during growth and ethanol stress . *Frontiers in Microbiology* vol. 6 103 (2015).
32. Zhang, W., Li, F. & Nie, L. Integrating multiple ‘omics’ analysis for microbial

biology: application and methodologies. *Microbiology* **156**, 287–301 (2010).
33. Rochfort, S. Metabolomics reviewed: a new 'omics' platform technology for systems biology and implications for natural products research. *J. Nat. Prod.* **68**, 1813–1820 (2005).
34. Karp, P. D. *et al.* The EcoCyc Database. *EcoSal Plus* **8**, 10.1128/ecosalplus.ESP-0006–2018 (2018).
35. Hu, J. C. *et al.* PortEco: a resource for exploring bacterial biology through high-throughput data and analysis tools. *Nucleic Acids Res.* **42**, D677–D684 (2014).
36. Guo, A. C. *et al.* ECMDB: the E. coli Metabolome Database. *Nucleic Acids Res.* **41**, D625–D630 (2013).
37. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
38. Timmons, J. J. & Densmore, D. Repository-based plasmid design. *PLoS One* **15**, e0223935 (2020).
39. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
40. Rodrigues, J. L. & Rodrigues, L. R. Potential Applications of the Escherichia coli Heat Shock Response in Synthetic Biology. *Trends Biotechnol.* **36**, 186–198 (2018).
41. Liu, Z. *et al.* Programming Bacteria With Light-Sensors and Applications in Synthetic Biology. *Front. Microbiol.* **9**, 2692 (2018).
42. San, K. Y., Bennett, G. N., Chou, C. H. & Aristidou, A. A. An optimization study of a pH-inducible promoter system for high-level recombinant protein production in Escherichia coli. *Ann. N. Y. Acad. Sci.* **721**, 268–276 (1994).
43. Shimada, T., Yamazaki, Y., Tanaka, K. & Ishihama, A. The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of Escherichia coli. *PLoS One* **9**, e90447 (2014).
44. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360 (2013).
45. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
46. Chen, J., Morita, T. & Gottesman, S. Regulation of Transcription Termination of Small RNAs and by Small RNAs: Molecular Mechanisms and Biological Functions . *Frontiers in Cellular and Infection Microbiology* vol. 9 201 (2019).
47. Ciampi, M. S. Rho-dependent terminators and transcription termination. *Microbiology* **152**, 2515–2528 (2006).
48. del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Díaz-Orejas, R. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**, 434–464 (1998).
49. Alexeyev, M. F., Shokolenko, I. N. & Croughan, T. P. Improved antibiotic-resistance gene cassettes and omega elements for Escherichia coli vector construction and in vitro deletion/insertion mutagenesis. *Gene* **160**, 63–67 (1995).
50. Goh, S. & Good, L. Plasmid selection in Escherichia coli using an endogenous

essential gene marker. *BMC Biotechnol.* **8**, 61 (2008).
51. Jang, C.-W. & Magnuson, T. A novel selection marker for efficient DNA cloning and recombineering in E. coli. *PLoS One* **8**, e57075–e57075 (2013).
52. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of Biologically Functional Bacterial Plasmids In Vitro *Proc. Natl. Acad. Sci.* **70**, 3240 LP – 3244 (1973).
53. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Res.* **10**, 1788–1795 (2000).
54. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
55. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **3**, e3647 (2008).
56. Li, M. Z. & Elledge, S. J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51–59 (2012).
57. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).
58. Foster, P. L. In vivo mutagenesis. *Methods Enzymol.* **204**, 114–125 (1991).
59. Cain, A. K. *et al.* A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* **21**, 526–540 (2020).
60. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4**, 206–223 (2009).
61. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008-2006.0008 (2006).
62. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
63. Bassalo, M. C. *et al.* Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synth. Biol.* **5**, 561–568 (2016).
64. Hou, Y., Yao, K., Yin, Y. & Wu, G. Endogenous Synthesis of Amino Acids Limits Growth, Lactation, and Reproduction in Animals. *Adv. Nutr.* **7**, 331–342 (2016).
65. Kinoshita, S., Nakayama, K. & Akita, S. Taxonomical Study of Glutamic Acid Accumulating Bacteria, Micrococcus glutamicus nov. sp. *Bull. Agric. Chem. Soc. Japan* **22**, 176–185 (1958).
66. Sanchez, S., Rodríguez-Sanoja, R., Ramos, A. & Demain, A. L. Our microbes not only produce antibiotics, they also overproduce amino acids. *J. Antibiot. (Tokyo).* (2017) doi:10.1038/ja.2017.142.
67. Leuchtenberger, W., Huthmacher, K. & Drauz, K. Biotechnological production of amino acids and derivatives: current status and prospects. *Appl. Microbiol. Biotechnol.* **69**, 1–8 (2005).
68. Uneyama, H., Kobayashi, H. & Tonouchi, N. New Functions and Potential Applications of Amino Acids. *Adv. Biochem. Eng. Biotechnol.* **159**, 273–287 (2017).
69. Wendisch, V. F., Bott, M. & Eikmanns, B. J. Metabolic engineering of Escherichia coli and Corynebacterium glutamicum for biotechnological production of organic

acids and amino acids. *Curr. Opin. Microbiol.* **9**, 268–274 (2006).
70. Nakamori, S. Early History of the Breeding of Amino Acid-Producing Strains. *Adv. Biochem. Eng. Biotechnol.* **159**, 35–53 (2017).
71. Rodriguez, A. *et al.* Engineering Escherichia coli to overproduce aromatic amino acids and derived compounds. *Microb. Cell Fact.* **13**, 126 (2014).
72. Usuda, Y. & Kurahashi, O. Effects of deregulation of methionine biosynthesis on methionine excretion in Escherichia coli. *Appl. Environ. Microbiol.* **71**, 3228–3234 (2005).
73. Niu, H. *et al.* Metabolic engineering for improving l-tryptophan production in Escherichia coli. *J. Ind. Microbiol. Biotechnol.* **46**, 55–65 (2019).
74. Wu, H. *et al.* Highly Efficient Production of l-Histidine from Glucose by Metabolically Engineered Escherichia coli. *ACS Synth. Biol.* **9**, 1813–1822 (2020).
75. Liu, Y. *et al.* Genetic engineering of Escherichia coli to improve L-phenylalanine production. *BMC Biotechnol.* **18**, 5 (2018).
76. Liu, L., Duan, X. & Wu, J. L-Tryptophan Production in Escherichia coli Improved by Weakening the Pta-AckA Pathway. *PLoS One* **11**, e0158200 (2016).
77. Kim, B., Binkley, R., Kim, H. U. & Lee, S. Y. Metabolic engineering of Escherichia coli for the enhanced production of l-tyrosine. *Biotechnol. Bioeng.* **115**, 2554–2564 (2018).
78. Park, J. H., Lee, K. H., Kim, T. Y. & Lee, S. Y. Metabolic engineering of Escherichia coli for the production of l-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc. Natl. Acad. Sci.* **104**, 7797 LP – 7802 (2007).
79. Park, J. H., Kim, T. Y., Lee, K. H. & Lee, S. Y. Fed-batch culture of Escherichia coli for L-valine production based on in silico flux response analysis. *Biotechnol. Bioeng.* **108**, 934–946 (2011).
80. Zhou, H.-Y. *et al.* Enhanced L-methionine production by genetically engineered Escherichia coli through fermentation optimization. *3 Biotech* **9**, 96 (2019).
81. *bp Statistical Review of World Energy*. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2021-full-report.pdf (2021).
82. Wuebbles, D. J. & Jain, A. K. Concerns about climate change and the role of fossil fuel use. *Fuel Process. Technol.* **71**, 99–119 (2001).
83. Jarboe, L. R., Grabar, T. B., Yomano, L. P., Shanmugan, K. T. & Ingram, L. O. Development of ethanologenic bacteria. *Adv. Biochem. Eng. Biotechnol.* **108**, 237–261 (2007).
84. Ohta, K., Beall, D. S., Mejia, J. P., Shanmugam, K. T. & Ingram, L. O. Genetic improvement of Escherichia coli for ethanol production: chromosomal integration of Zymomonas mobilis genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Appl. Environ. Microbiol.* **57**, 893–900 (1991).
85. Aditiya, H. B., Mahlia, T. M. I., Chong, W. T., Nur, H. & Sebayang, A. H. Second generation bioethanol production: A critical review. *Renew. Sustain. Energy Rev.* **66**, 631–653 (2016).
86. Asghari, A., Bothast, R. J., Doran, J. B. & Ingram, L. O. Ethanol production from

hemicellulose hydrolysates of agricultural residues using genetically engineered Escherichia coli strain KO11. *J. Ind. Microbiol. Biotechnol.* **16**, 42–47 (1996).
87. Atsumi, S. *et al.* Metabolic engineering of Escherichia coli for 1-butanol production. *Metab. Eng.* **10**, 305–311 (2008).
88. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).
89. Inokuma, K., Liao, J. C., Okamoto, M. & Hanai, T. Improvement of isopropanol production by metabolically engineered Escherichia coli using gas stripping. *J. Biosci. Bioeng.* **110**, 696–701 (2010).
90. Shen, C. R. & Liao, J. C. Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. *Metab. Eng.* **10**, 312–320 (2008).
91. Hanai, T., Atsumi, S. & Liao, J. C. Engineered synthetic pathway for isopropanol production in Escherichia coli. *Appl. Environ. Microbiol.* **73**, 7814–7818 (2007).
92. Shen, C. R. & Liao, J. C. Synergy as design principle for metabolic engineering of 1-propanol production in Escherichia coli. *Metab. Eng.* **17**, 12–22 (2013).
93. World Economic Forum. *The new plastics economy: rethinking the future of plastics*. (2016).
94. L., M. L. & W., H. G. Metabolic Engineering of Poly(3-Hydroxyalkanoates): From DNA to Plastic. *Microbiol. Mol. Biol. Rev.* **63**, 21–53 (1999).
95. Kim, B. S., Lee, S. Y. & Chang, H. N. Production of poly-β-hydroxybutyrate by fed-batch culture of recombinantEscherichia coli. *Biotechnol. Lett.* **14**, 811–816 (1992).
96. Rafiqah, S. A. *et al.* A Review on Properties and Application of Bio-Based Poly(Butylene Succinate). *Polymers (Basel).* **13**, (2021).
97. Lin, H., Bennett, G. N. & San, K.-Y. Metabolic engineering of aerobic succinate production systems in Escherichia coli to improve process productivity and achieve the maximum theoretical succinate yield. *Metab. Eng.* **7**, 116–127 (2005).
98. Yim, H. *et al.* Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7**, 445–452 (2011).
99. Baeshen, N. A. *et al.* Cell factories for insulin production. *Microb. Cell Fact.* **13**, 141 (2014).
100. Fineberg, S. E. *et al.* Immunological Responses to Exogenous Insulin. *Endocr. Rev.* **28**, 625–652 (2007).
101. Redwan, E. M., Matar, S. M., El-Aziz, G. A. & Serour, E. A. Synthesis of the Human Insulin Gene: Protein Expression, Scaling Up and Bioactivity. *Prep. Biochem. Biotechnol.* **38**, 24–39 (2007).
102. Govender, K. *et al.* A novel and more efficient biosynthesis approach for human insulin production in Escherichia coli (E. coli). *AMB Express* **10**, 43 (2020).
103. Rezaei, M. & Zarkesh-Esfahani, S. H. Optimization of production of recombinant human growth hormone in Escherichia coli. *J. Res. Med. Sci.* **17**, 681–685 (2012).
104. Zanette, D. *et al.* Human IL-1 receptor antagonist from Escherichia coli: large-scale microbial growth and protein purification. *J. Biotechnol.* **64**, 187–196 (1998).
105. Liu, Q. *et al.* Large scale preparation of recombinant human parathyroid hormone 1–84 from Escherichia coli. *Protein Expr. Purif.* **54**, 212–219 (2007).

106. Stevenson, G., Andrianopoulos, K., Hobbs, M. & Reeves, P. R. Organization of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J. Bacteriol.* **178**, 4885–4893 (1996).
107. Vollmer, W. & Bertsche, U. Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. *Biochim. Biophys. Acta - Biomembr.* **1778**, 1714–1734 (2008).
108. Yu, H. & Stephanopoulos, G. Metabolic engineering of Escherichia coli for biosynthesis of hyaluronic acid. *Metab. Eng.* **10**, 24–32 (2008).
109. Henrotin, Y., Mathy, M., Sanchez, C. & Lambert, C. Chondroitin sulfate in the treatment of osteoarthritis: from in vitro studies to clinical recommendations. *Ther. Adv. Musculoskelet. Dis.* **2**, 335–348 (2010).
110. Badri, A. *et al.* Complete biosynthesis of a sulfated chondroitin in Escherichia coli. *Nat. Commun.* **12**, 1389 (2021).

Chapter 2: Microbial Production of Human Milk Oligosaccharide Lactodifucotetraose (LDFT)

This chapter is reproduced with permission from:
Zhang A, Sun L, Bai Y, Yu H, McArthur JB, Chen X, Atsumi S.
"Microbial Production of Human Milk Oligosaccharide Lactodifucotetraose"
Metab Eng. 2021. Volume 66, 12-20.
doi:10.1016/j.ymben.2021.03.014

2.1 Introduction

Human milk oligosaccharides (HMOs) are a class of over 200 compounds present at 20-23 g/L in colostrum and 12-14 g/L in mature milk[1–4]. Unlike their common precursor lactose, HMOs are indigestible by human infants and instead improve neonatal health by serving as effective antimicrobials and antivirals, prebiotics, and regulators of inflammatory immune cell-response cascades[4–9]. These and other potential benefits of HMOs make them attractive targets of study for preventing or treating diseases in both children and adults[4]. The bioactive properties of HMOs have motivated efforts to define mechanistic effects of individual compounds[10–13], but the sources of HMOs are limited and their large-scale isolation for such studies is exceedingly difficult. While production of individual HMOs using *in vitro* enzymatic reactions has been successful[14–21], these methods require supplementation of stoichiometric amounts of ATP and other cofactors that increase the production cost and may complicate the purification process of the oligosaccharide products.

Microbial production is a viable alternative method to produce HMOs. Whole cell biocatalysts are self-maintaining systems and do not require an exogenous supply of expensive cofactors. Enzymatic reactions in cells can also achieve high regio- and stereo-specific production of structurally complex molecules. Several simple HMOs including 2'-FL, 3-FL, lacto-*N*-triose II, lacto-*N*-tetraose (LNT), and lacto-*N*-neotraose

(LNnT) have been produced in engineered microorganisms[22–27]. Linear HMO backbones such as lactose, LNT, and LNnT can be combinatorially glycosylated at multiple sites with fucose and sialic acid to further produce HMOs of higher structural complexity. While *in vitro* enzymatic synthesis can construct these decorated HMOs by strategically producing each intermediate HMO structure in individual reaction systems, microbial production of multi-glycosylated HMOs in a microbial host has not been demonstrated.

We have established a method for producing a difucosylated tetrasaccharide lactodifucotetraose (LDFT) in *E. coli*. LDFT is one of the most abundant fucosylated HMOs and is produced at an average of 0.43 g/L over the first year of lactation by secretory mothers[28]. Its structure consists of a core lactose unit that is fucosylated at the C2' and C3 positions. Studies have shown that LDFT is effective in preventing *Campylobacter jejuni*-associated diarrhea and suppressing platelet-induced inflammatory processes in neonates[29,30]. Its activity as a gastrointestinal and immunological modulator has motivated further research into its potential therapeutic applications. However, the high cost and limited availability of LDFT in the market ($66 - 140/mg; Biosynth Carbosynth, Elicityl) are barriers to these biological studies.

LDFT can be synthesized from lactose and L-fucose in a two-step fucosylation process using an α1–2-fucosyltransferase and an α1–3-fucosyltransferase. While monofucosylation of lactose with a single fucosyltransferase for the microbial production of 2'-FL and 3-FL has been studied, the effects of implementing an α1–2-fucosyltransferase and an α1–3-fucosyltransferase together in a cellular system to produce a difucosylated HMO has not been reported. As lactose is a suitable acceptor

substrate for both fucosyltransferases, both 2'-FL and 3-FL can be produced as mono-fucosylated products in the first fucosylation step of the system with the presence of both fucosyltransferases. It was shown previously that while an α1–3/4-fucosyltransferase from *Helicobacter pylori* (Hp3/4FT) can use both non-fucosylated and α1–2-fucosylated galactosyl oligosaccharides as substrates[19,21], α1–2-fucosyltransferases from *Escherichia coli* O126 (EcWbgL)[21,31] and *Thermosynechococcus elongates* [18] are selective towards lactose and other non-fucosylated galactosyl oligosaccharide acceptor substrates. Therefore, to produce LDFT in high yields, it is essential to pair an α1–2-fucosyltransferase with high activity towards lactose with an α1–3-fucosyltransferase with higher activity towards 2'-FL than lactose so that there is minimal production of 3-FL as a side product.

In this study, we created a system in *E. coli* using two fucosyltransferases that preferentially fucosylates lactose to form a 2'-FL intermediate that is further fucosylated to produce the target LDFT. We assessed various promoter expression systems to establish heterologous expression of the desired biosynthetic pathway. LDFT production was decoupled from bacterial growth by removing catabolic pathways of starting substrates and by maintaining cell density with glycerol, an inexpensive carbon source that does not activate carbon catabolite repression of lactose and L-fucose transporters[32,33]. To enhance intracellular availability of substrates, the lactose and L-fucose transporter genes, *lacY* and *fucP*, were additionally expressed from plasmids. With additional fine-tuning of the expression levels of individual glycosyltransferase genes, we produced 5.1 g/L of LDFT from 3 g/L lactose, achieving 91% of the theoretical maximum yield of LDFT in 24 h.

2.2 Results

2.2.1 Pathway design for LDFT production in *E. coli*

HMO production does not naturally occur in *E. coli*, therefore we employed the following three enzymes for the production of LDFT: a bifunctional L-fucokinase/GDP-L-fucose pyrophosphorylase (BfFKP) from *Bacteroides fragilis*[34], an α1-2-fucosyltransferase (EcWbgL) from *E. coli* O126[21,31], and α1-3/4-fucosyltransferase (Hp3/4FT) from *Helicobacter pylori* UA948[19,35]. Acceptor substrate specificity studies of both EcWbgL and Hp3/4FT have been reported[19,21,31,36]. EcWbgL exhibits high activity towards non-fucosylated acceptor substrates, such as lactose, *N*-acetyllactosamine (LacNAc), and lactulose, and no activity towards 3-FL. Hp3/4FT has been shown to be highly active towards LacNAc and 2'-fucosyl-LacNAc with low activity towards lactose. The acceptor preferences of the fucosyltransferases allow sequential fucosylation of lactose for the formation of LDFT in the presence of both fucosyltransferases. BfFKP uses one ATP and GTP to convert L-fucose to GDP-fucose, which is taken as a donor substrate by EcWbgL to fucosylate lactose at the C2' position, forming the intermediate 2'-FL (Figure 2.2.1). Due to its structural similarity to 2'-fucosyl-LacNAc, 2'-FL was hypothesized to be a suitable acceptor substrate for fucosylation by Hp3/4FT to produce LDFT, which is expected to be secreted to the supernatant by native membrane exporter SetA[37].

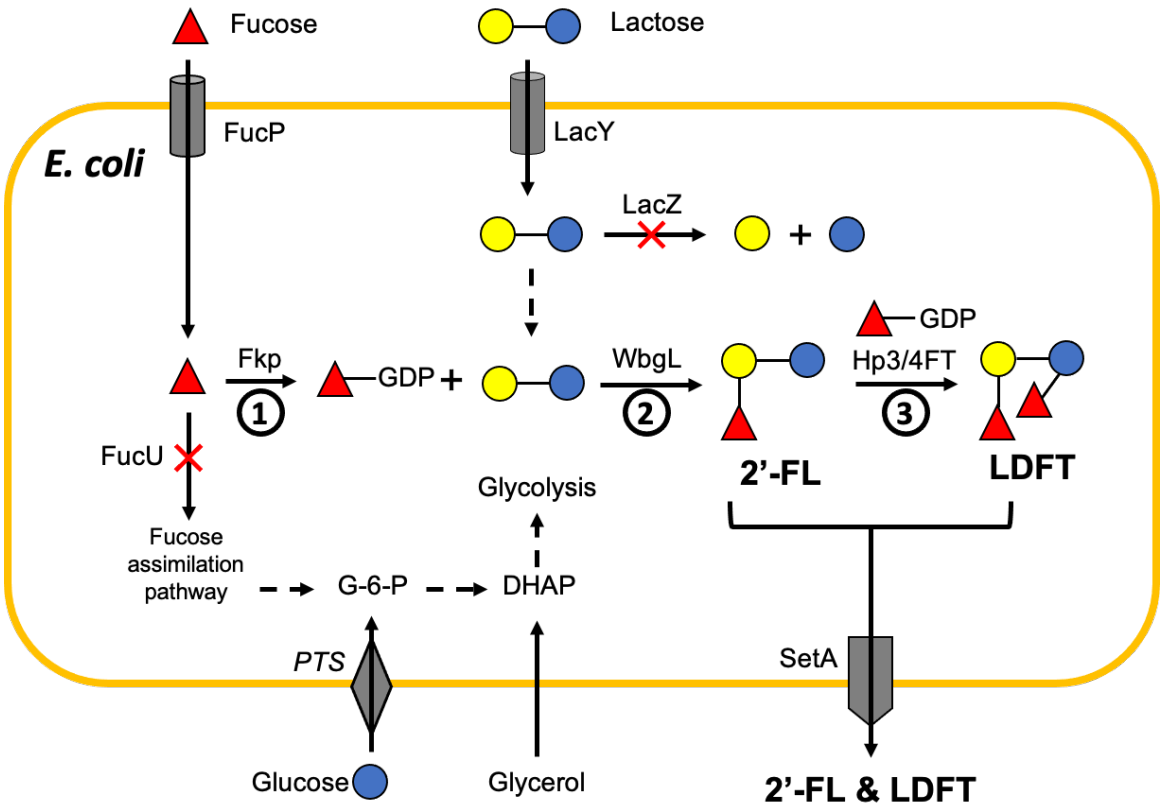


Figure 2.2.1 Pathway design for LDFT production in *E. coli*. Fucose (red triangle) and lactose (glucose moiety, blue diagonal striped circle; galactose moiety, yellow filled circle) are transported into the cytosol through sugar transporters FucP and LacY, respectively. The *fucU* and *lacZ* genes are deleted to prevent substrate assimilation into central carbon metabolism. Fucose is converted to donor substrate GDP-fucose by BfFKP. EcWbgL glycosylates lactose at the C2' position with GDP-fucose to form 2'-FL. Hp3/4FT glycosylates 2'-FL at the C3 position with GDP-fucose to form LDFT. G-6-P: glucose-6-phosphate, DHAP: dihydroxyacetone phosphate, PTS: phosphotransferase system, SetA: sugar efflux transporter A.

We were initially concerned that the relatively low soluble expression level of recombinant fucosyltransferases might cause bottlenecks for synthesizing fucosylated HMOs in microbial hosts[38]. In this study, we truncated the C-terminal 34-amino acid hydrophobic sequence of Hp3/4FT to increase its solubility[19]. To increase the expression of fucosyltransferases, we selected *E. coli* B strain BL21 Star (DE3) (Table S2) as an LDFT production host. BL21 Star (DE3) is widely used for recombinant

protein expression and is capable of high expression via the two-step IPTG-inducible T7 bacteriophage promoter[39]. The *fkp* and *wbgL* genes were cloned together into an expression vector under a T7-promoter (P_{T7}, pAL1779, Table S3) and the truncated *Hp3/4ft* was cloned into a second expression vector under P_{T7} (pAL1817, Table S3). Lactose and L-fucose were used as starting substrates for LDFT production, but *E. coli* is known to catabolize these two sugars for growth. To maximize LDFT production, assimilation of L-fucose and lactose for cellular growth should be minimized. Therefore, we evaluated the strain's ability to grow on these two carbon sources to determine which carbon assimilating pathways to remove. Although the BL21 Star (DE3) encodes all genes involved in L-fucose degradation, the strain was not able to grow on L-fucose as the sole carbon source (Figure S1a). The strain was able to grow on lactose as the sole carbon source (Figure S1a). When the *lacZ* gene encoding for a β-galactosidase was deleted in the strain (Table S1: Strain 1), lactose did not enable growth anymore (Figure S1).

The two plasmids containing the LDFT production pathway (pAL1779 and pAL1817, Table S3) were introduced into Strain 1 to form Strain 2 (Table S1). To determine the optimal carbon source for growth and production, Strain 2 was grown in parallel with glucose, a common feedstock known for its catabolite repression towards lactose importation[40], and glycerol, an inexpensive feedstock that does not cause catabolite repression. Under both of these culturing conditions, Strain 2 did not produce LDFT nor its precursor, 2'-FL. To examine the expression from P_{T7}, the plasmid containing *sfgfp* under P_{T7} (pAL1843, Table S3) was introduced into BL21 Star (DE3) and Strain 1 to form Strains 3 and 4, respectively (Table S1). Strain 3 produced a strong

fluorescent signal after IPTG induction while Strain 4 did not produce fluorescence signal in either induction conditions, suggesting that T7 RNA polymerase expression was lacking (Figure S1b). Sequencing of the *attB* integration locus in Strain 1 revealed an excision of the λDE3 lysogen containing P_{lacUV5}:*lacZ*α–*T7rnap*. Several attempts were made to remove *lacZ* from BL21 Star (DE3) without off-target modifications to the λDE3 lysogen but resulted in either no gene deletion or cell death.

2.2.2 LDFT production in *E. coli* B strains

Due to difficulties in genetically modifying BL21 Star (DE3), we integrated P_{lacUV5}:*T7rnap* into the *E. coli* K-12 derivative strains, BW25113 Z1 and MG1655 Z1 (Table S2). The Z1 fragment containing *lacI*^{q}, *tetR*, and *spec*^{r} was integrated into the *attB* site of these strains. Many regions in the *E. coli* genome are stable and high-efficiency integration sites for heterologous genes[41], therefore we chose intergenic locus *ss9* as the insertion site for P_{lacUV5}:*T7rnap*. The P_{lacUV5}:*T7rnap* cassette was integrated into *ss9* of BW25113 Z1 and MG1655 Z1 to form Strains 5 and 6, respectively (Table S1).

We introduced pAL1834 containing P_{T7}:*sfgfp* into Strains 5 and 6 to form Strains 7 and 8, respectively (Table S1) to assess the repression and induction efficiencies of P_{T7} through a fluorescence assay. Tight repression of GFP expression without IPTG was observed in Strains 7 and 8 (Figure S2a). IPTG induction in Strains 7 and 8 increased GFP fluorescence 95-fold and 440-fold, respectively (Figure S2a). Strain 6 was chosen as the base strain for further genetic modification due to its tighter repression and stronger inducibility of P_{T7}. We tested the growth of Strain 6 on L-fucose and lactose. It was able to grow on fucose or lactose as a sole carbon source (Figure

S2b). To remove L-fucose and lactose assimilation, we deleted *fucU* encoding an L-fucose mutarotase and *lacZ* to form Strain 10 (Table S1). Strain 10 was not able to grow on L-fucose or lactose as a sole carbon source (Fig. S2b).

The LDFT production plasmids (pAL1779 and pAL1817, Table S3) were introduced into Strain 10 to form Strain 11 (Table S1). Strain 11 was grown to test LDFT production from lactose and L-fucose. Glucose or glycerol was used to maintain cellular growth. Under both conditions, LDFT was not produced in Strain 11. This led us to examine the T7 RNA polymerase expression system in Strain 10. pAL1834 containing P_{T7}:*sfgfp* was introduced into Strain 10 to form Strain 12 (Table S1). Strain 12 produced strong GFP fluorescence without IPTG induction, indicating the expression from P_{T7} was leaky in Strain 12 (Figure S2c). In Strain 10, we found a mutation in the promoter region of the P_{lacUV5}:*T7rnap* cassette. We have attempted several times the deletion of *lacZ* in Strain 9 without incurring P_{lacUV5} mutations but were unsuccessful. Due to the similarity of the *lacZ* promoter to P_{lacUV5}, we assume that the mutations in P_{lacUV5} are correlated with the CRISPR-Cas9-mediated gene removal of *lacZ*.

2.2.3. Introduction of the T7 RNAP gene into K-12 derivative strains

To avoid the potential sequence similarity issues observed for P_{lacUV5} and the native *lacZ* promoter, we introduce the three modifications into MG1655 Z1 in a different order. We first deleted *fucU* and *lacZ* in MG1655 Z1 to form Strain 13 (Δ*fucU*) and Strain 14 (Δ*fucU* Δ*lacZ*)) and then integrated P_{lacUV5}:*T7rnap* into the *ss9* locus to form Strain 15 (Table S1). Strain 15 was unable to grow on L-fucose or lactose as a sole carbon source (Figure S2d). Although the P_{lacUV5}:*T7rnap* cassette in Strain 15 had no mutations, Strain 15 with pAL1834 harboring P_{T7}:*sfgfp* (Table S1: Strain 16) showed

leaky GFP expression without IPTG. To determine if other *lac*-based promoters are deregulated by our strain modifications, we introduced pAL421 containing P_{LlacO1}:*sfgfp* into MG1655 z1, Strains 14 and 15 to form Strains 17, 18, and 19, respectively (Table S1) to assess the regulation of the lac-based promoter in these strains. The expressions from P_{LlacO1} without IPTG were well repressed in Strains 17, 18 and 19 (Figure S3a). Next, pAL2045 containing P_{lacUV5}:*sfgfp* was introduced into MG1655 *Z1*, Strains 13 and 14 to form Strains 20, 21, and 22, respectively (Table S1). The expression of *sfgfp* in Strains 21 and 22 was leakier than that in Strain 20 (Figure S3b), suggesting that the deletion of *fucU* caused the leaky P_{lacUV5} expression.

2.2.4. Production of LDFT in K-12 derivative strains

Rather than pursuing alternative promoters for *T7rnap*, we decided to use other induction systems for the LDFT biosynthetic pathway genes. The *fkp* and *wbgL* genes were cloned under P_{LlacO1} (pAL1759, Table S3) and the *Hp3/4ft* gene was cloned under an aTc-inducible promoter P_{LtetO1} (pAL1760, Table S3)[42]. The LDFT production plasmids (pAL1759 and pAL1760) were introduced to Strain 14 to form Strain 23 (Table 1). Strain 23 was grown in M9P containing L-fucose and lactose with glucose or glycerol. After 24 h, Strain 23 produced 0.08 g/L 2'-FL and 0.16 g/L LDFT under the glycerol conditions, but neither were produced under the glucose conditions (Figure 2).

2.2.5. Enhancing substrate levels by overexpressing transporter genes

Intracellular availability of L-fucose and lactose is important for efficient LDFT production. We hypothesized that expression of the substrate transporter genes would increase the substrate supply and improve LDFT production. Therefore, we expressed the lactose and L-fucose membrane symporter genes, *lacY* and *fucP,* under a

constitutive promoter (iGEM part No. BBa_K1824896, Table S3). The *lacY* gene was expressed from the *fkp-wbgL* plasmid pAL2027 (Table S3). The LDFT production plasmids with *lacY* (pAL2027 and pAL1760) were introduced into Strain 14 to form Strain 24 (Table S1) but the overexpression of *lacY* did not improve LDFT production (Figure 2.2.3). The *fucP* gene was expressed from the *fkp-wbgL* plasmid pAL2028 (Table S3). The LDFT production plasmids with *fucP* (pAL2028 and pAL1760) were introduced into Strain 14 to form Strain 25 (Table 1). After 24 h, Strain 25 produced 0.9 g/L LDFT, a 6.9-fold improvement compared to Strain 23.

Next, both *lacY* and *fucU* were expressed from the *fkp-wbgL* plasmid pAL2029 (Table S3). The LDFT-production plasmids with *lacY* and *fucU* (pAL2029 and pAL1760)

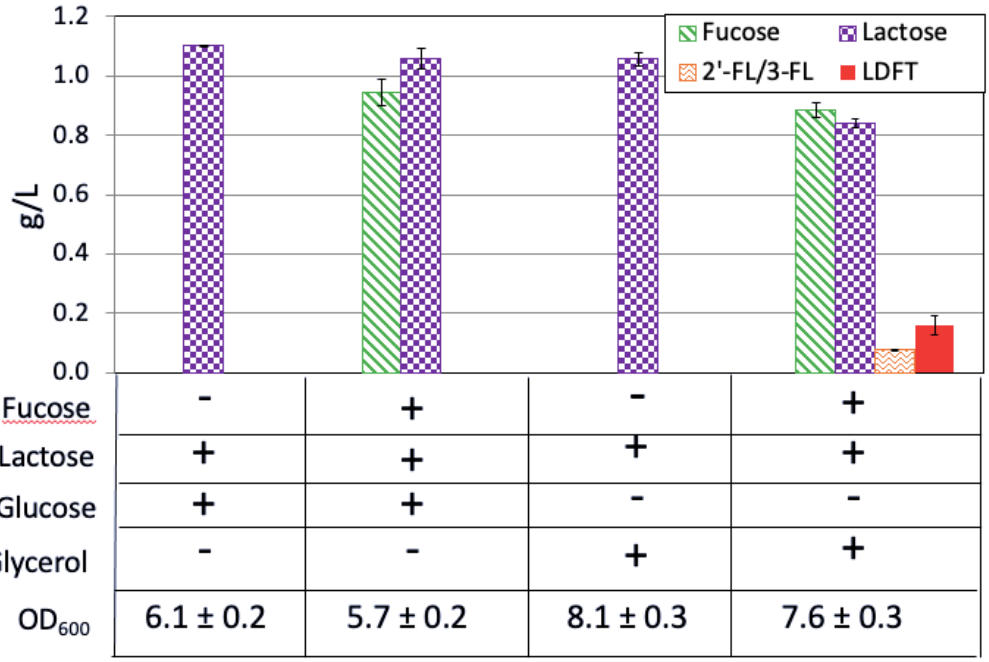


Figure 2.2.2 Effects of carbon sources on LDFT production. Strain 23 (MG1655 Z1 ΔfucU ΔlacZ with the LDFT production plasmids, Table 1) was grown in M9P with 5 g/L glucose or 10 g/L glycerol at 30 °C for 24 h. Cultures were supplemented with 1 g/L lactose and with or without 1 g/L fucose and induced with 1 mM IPTG and 100 ng/mL aTc. L-Fucose concentration (green diagonal), lactose concentration (purple checkered), monofucosides (2'-FL/3-FL) concentration (orange zigzag) and LDFT concentration (red filled) were measured at 24 h. Error bars indicate s.d. (n = 3 biological replicates).

were introduced into Strain 14 to form Strain 26 (Table S1). Strain 26 produced 1.1 g/L LDFT after 24 h, representing 59% of the theoretical maximum yield (TMY) from lactose and accumulated 0.17 g/L 2'-FL and/or 3-FL (Figure 2.2.3). As the HPLC and the MS methods used were unable to discriminate between the two mono-fucosylated lactose, the combined concentrations of 2'-FL and 3-FL are reported in this paper.

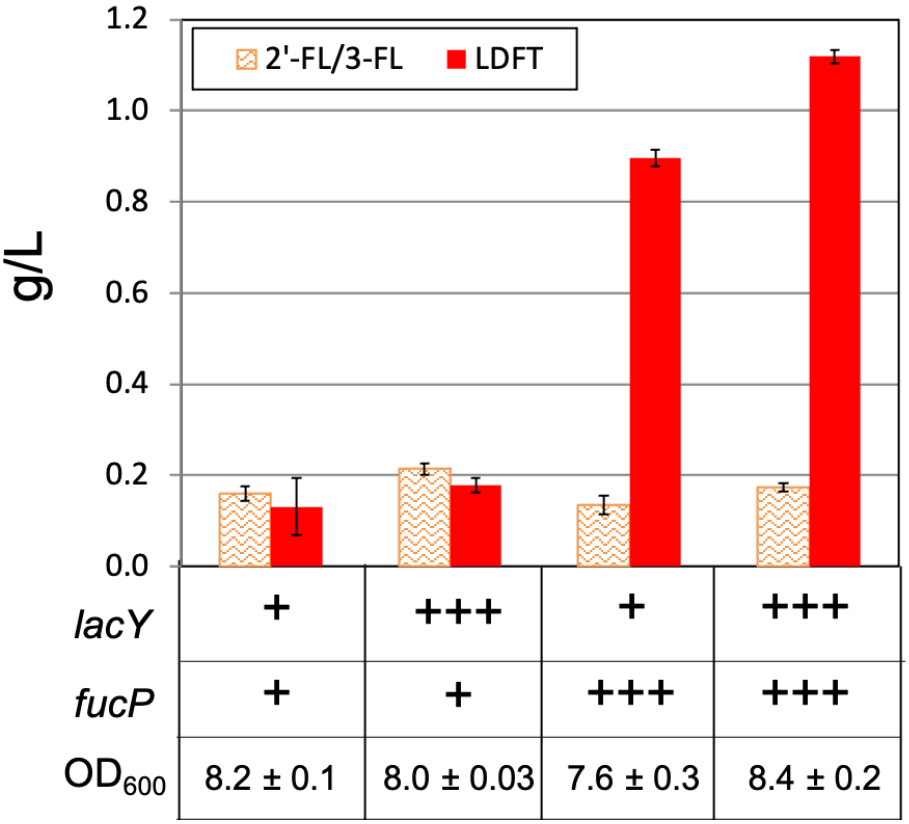


Figure 2.2.3 Additional expression of lactose and fucose permease genes to enhance lactose and fucose availabilities. Strain 14 (MG1655 Z1 ΔfucU ΔlacZ, Table 1) was used as a host strain. Strain 23 (Strain 14 with the LDFT production plasmids), Strain 24 (Strain 14 with the LDFT production plasmids with lacY), Strain 25 (Strain 14 with the LDFT production plasmids with fucU), and Strain 26 (Strain 14 with the LDFT production plasmids with lacY and fucU) were grown in M9P with 10 g/L glycerol at 30 °C for 24 h. Cultures were supplemented with 1 g/L lactose and 1 g/L L-fucose and induced with 1 mM IPTG and 100 ng/mL aTc. Growth and production of monofucosides (2'-FL/3-FL, orange zigzag) and LDFT (red filled) were determined at 24 h. + indicates the corresponding gene was expressed from the genome and +++ indicates the corresponding gene was additionally expressed from a plasmid. Error bars indicate s.d. (n = 3 biological replicates).

2.2.6 Tuning of the expression levels of the LDFT biosynthetic pathway genes

To fine-tune the nucleotide activation of L-fucose and the fucosylation reactions, we screened a range of IPTG concentrations (0, 25, 50, 100, and 1000 μM) for the expression of P_{LlacO1}:*fkp-wbgL* in the presence of 100 ng/mL aTc for induction of P_{LtetO1}:*Hp3/4ft* . The best growth, greatest lactose and L-fucose consumption, and the highest level LDFT production (1.6 g/L, 89% of TMY) was observed with 50 μM IPTG (Figure 2.2.4a). A range of aTc concentrations (0, 25, 50 and 100 ng/mL) were tested for the LDFT production in the presence of 50 μM IPTG to determine if adjusting Hp3/4FT expression levels could improve LDFT production. Strain 26 produced more LDFT with higher concentrations of aTc (Figure 2.2.4b). Thus, the induction condition with 50 μM IPTG and 100 ng/mL aTc was used for further studies.

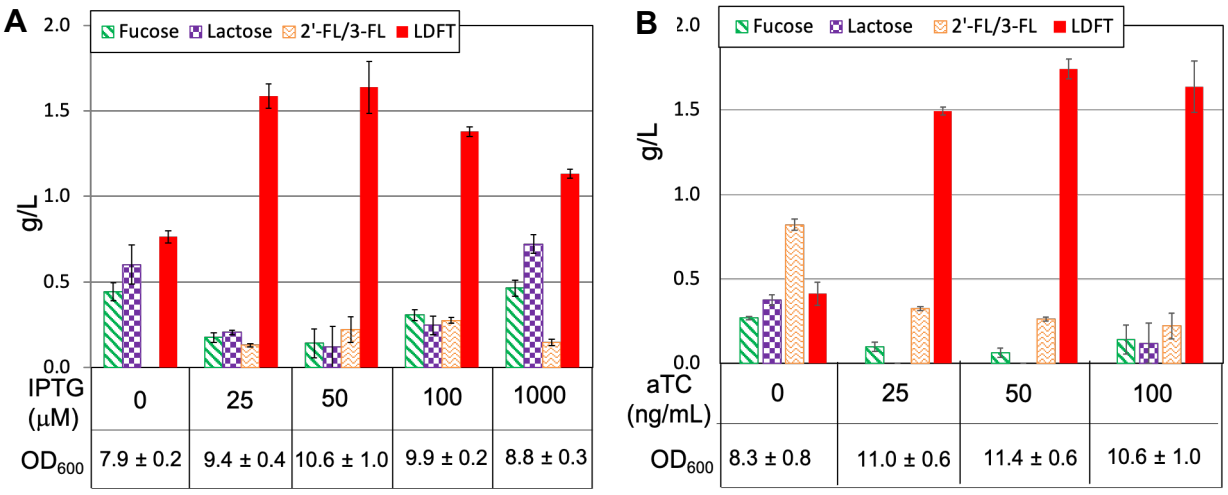


Figure 2.2.4 Effects of IPTG and aTc concentrations on LDFT production. (A) Strain 26 (MG1655 Z1 ΔfucU ΔlacZ with the LDFT production plasmids with lacY and fucU) was grown in M9P with 10 g/L glycerol at 30 °C for 24 h. Cultures were supplemented with 1 g/L lactose and 1 g/L fucose and induced with 100 ng/mL aTc and various concentrations of IPTG (0, 25, 50, 100 and 1000 mM). (B) Strain 26 was grown as described in (a) except with 50 mM IPTG and various concentrations of aTc (0, 25, 50, 100 ng/mL). OD_{600}, L-Fucose concentration (green diagonal), lactose concentration (purple checkered), monofucosides (2'-FL/3-FL) concentration (orange zigzag) and LDFT concentration (red filled) were measured at 24 h. Error bars indicate s.d. (n ≥ 3 biological replicates).

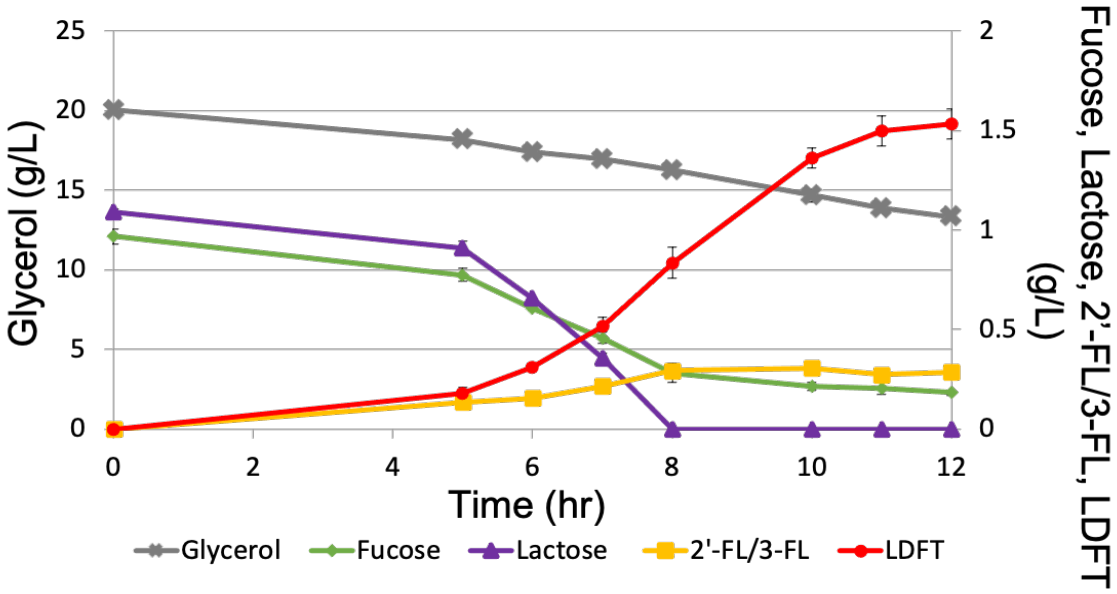


Figure 2.2.5 LDFT production in Strain 26. Strain 26 (MG1655 z1 ΔfucU ΔlacZ with the LDFT production plasmids with lacY and fucU) was grown in M9P with 20 g/L glycerol at 30 °C for 12 h. Cultures were supplemented with 1 g/L lactose and 1 g/L fucose and induced with 50 mM IPTG and 100 ng/mL aTc. Glycerol concentration (gray cross), fucose concentration (green diamond), lactose concentration (purple triangle), monofucosides (2'-FL/3-FL) concentration (orange square) and LDFT concentration (red circle) were monitored during the experiment. Error bars indicate s.d. (n = 3 biological replicates).

2.2.7 Characterization of LDFT production

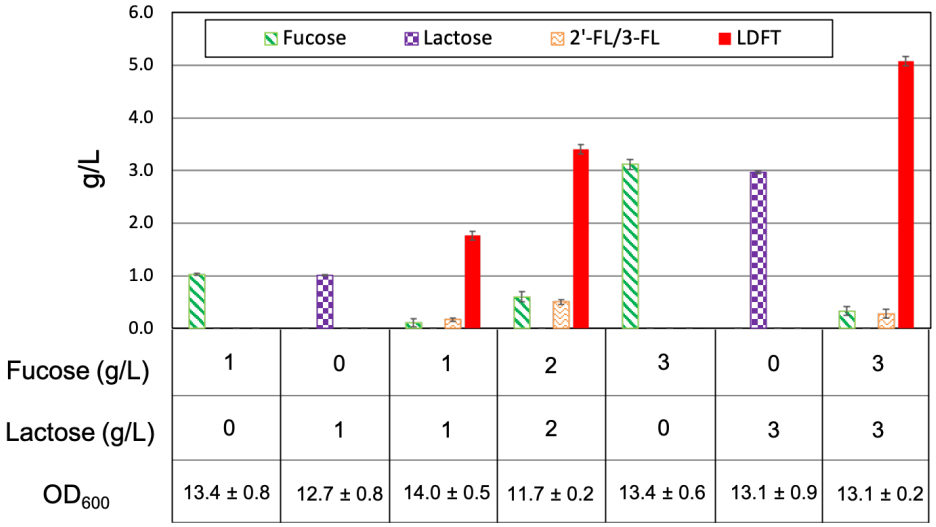
The LDFT production profile in Strain 26 was characterized for 12 h post-induction by monitoring substrate, intermediate, side product, and LDFT levels using HPLC (Fig. 5). LDFT was first detected at 5 h, and between 5 to 10 h the production rate was 0.24 g/L/h (Fig. 5). Monofucosides (2'-FL/3-FL) were accumulated up to 0.3 g/L until lactose was depleted at 8 h and remained constant at ~0.3 g/L between 8 to 12 h. The lack of monofucoside consumption after 8 h indicated that most of the remaining monofucoside was 3-FL, which was the side product produced by Hp3/4FT from lactose that cannot be fucosylated further by EcWbgL to produce LDFT.

When EcWbgL and Hp3/4FT are expressed at the same time, both enzymes can compete to fucosylate lactose into 2'-FL and 3-FL, respectively. In the presence of lactose and 2'-FL, Hp3/4FT can also convert the respective acceptor substrates into 3-FL and LDFT. We hypothesized that the delayed induction of *Hp3/4ft* would decrease the competition between EcWbgL and Hp3/4FT for lactose and decrease the production of the side product, 3-FL. Therefore, we tested the delaying of the Hp3/4FT expression by adding 100 ng/mL aTc at 2, 4, and 6 h. However, the delayed expressions of *Hp3/4ft* resulted in increased monofucoside accumulation and decreased LDFT production (Figure S4). This increase in monofucoside in the supernatant suggests that 2'-FL formed by EcWbgL may be secreted to media and its reimport may be limited, which decreases the substrate availability of Hp3/4FT for LDFT production.

To examine the import efficiency of 2'-FL, we fed 2'-FL to the production cultures. The *wbgL* gene was removed from pAL2029 to form pAL2059 (Table S3). pAL2059 and pAL1760 were introduced into Strain 14 to form Strain 27 (Table S1). Strain 27 was grown in M9P with 10 g/L glycerol. Cultures were induced with 50 μM IPTG and 100 ng/mL aTc and supplemented with 1.42 g/L of 2'-FL (mole equivalent to 1 g/L lactose) and 0.5 g/L L-fucose. Lactose was not fed to the cultures and *wbgL* was not present in system, making it unlikely for Strain 27 to produce 2'-FL and 3-FL. Under these conditions, LDFT should be produced only from the fed 2'-FL. Strain 27 produced only 0.4 g/L LDFT in 24 h, further supporting that the import of 2'-FL is not efficient in *E. coli* (Fig. S5).

2.2.9 LDFT production with higher substrate concentrations

Strain 26 consumed 1 g/L lactose within 8 h and LDFT production reached completion at 12 h post-induction (Figure 2.2.5). To evaluate LDFT production with higher substrate concentrations, Strain 26 was grown in M9P with 20 g/L glycerol and various amounts of lactose and L-fucose (1, 2, or 3 g/L) for 24 h. In conditions with only lactose or fucose as the added substrate, Strain 26 did not produce any detectable amounts of fucosides. In the presence of both substrates, the increase in LDFT yield was proportional to the increase of substrate concentrations (Figure 2.2.6). Strain 26 consumed 3.0 g/L lactose and 2.6 g/L L-fucose and produced 5.1 g/L LDFT as the major product in 24 h with minor accumulation of monofucosides. LDFT was produced at 91% of TMY.



Fucose (g/L)	1	0	1	2	3	0	3
Lactose (g/L)	0	1	1	2	0	3	3
OD_{600}	13.4 ± 0.8	12.7 ± 0.8	14.0 ± 0.5	11.7 ± 0.2	13.4 ± 0.6	13.1 ± 0.9	13.1 ± 0.2

Figure 2.2.6 LDFT production in Strain 26 with various concentrations of lactose and fucose. Strain 26 (MG1655 Z1 ΔfucU ΔlacZ with the LDFT production plasmids with lacY and fucU) was grown in M9P with 20 g/L glycerol at 30 °C for 24 h. Cultures were supplemented with lactose and L-fucose (1, 2, and 3 g/L of each) and induced with 50 mM IPTG and 100 ng/mL aTc. OD_{600}, fucose concentration (green diagonal), lactose concentration (purple dotted), 2'-FL concentration (orange wave) and LDFT concentration (red filled) were measured at 24 h. Error bars indicate s.d. (n ≥ 3 biological replicates).

2.3 Discussion

LDFT has been identified as an effective gastrointestinal and immunological modulator and has the potential to be developed to treat human diseases. Its high cost and limited commercial access make LDFT a desirable target for production in microbial hosts. Systems developed in *E. coli*, *B. subtilis*, and *S. cerevisiae* have successfully produced HMOs such as 2'-FL, 3-FL, LNT, and LNnT, which represent only a small fraction of over 200 naturally occurring HMOs. Developing microbial production systems dedicated to synthesizing HMOs with a higher structural complexity is still challenging. In this study, we established a microbial system that specifically and efficiently produces LDFT.

The greatest challenge of this study was pairing an α1–2-fucosyltransferase with an α1–3-fucosyltransferase that can efficiently produce LDFT with minimal accumulation of monofucoside intermediates. We selected EcWbgL to drive lactose fucosylation into 2'-FL because it expresses well in *E. coli* and has been characterized to prefer β1-4-linked galactose substrates, such as lactose and LacNAc[31]. From acceptor substrate screenings of α1–3-fucosylatransferases, Hp3/4FT was annotated with high activity towards 2'-fucosyl LacNAc, which suggested 2'-FL may also be a suitable acceptor for Hp3/4FT[19,36]. In our characterization of LDFT production, we showed Hp3/4FT had preferential activity towards 2'-FL over lactose and LDFT was formed as the dominant product (Figure 2.2.5). The presence of residual monofucosides indicates possible formation of the side product 3-FL, which is an unsuitable acceptor for EcWbgL[31]. Fortunately, monofucoside titers were relatively low and can be separated from LDFT in downstream purification processes. We can continue to screen

α1–3-fucosyltransferases for lower activity towards lactose and also pursue protein engineering strategies to expand α1–2-fucosyltransferase's acceptor substrate range to 3-FL so that this side product can be fucosylated into LDFT.

The rate of LDFT formation was dictated by carbon catabolite repression (CCR) and the activity of sugar transporters, which firmly control the import of carbohydrates across the inner membrane[43]. It has been shown that import of glucose through the phosphotransferase system inhibits transcription of *lac* operon genes, including *lacY*. From our experiments, glucose conditions led to suppressed LDFT production while glycerol conditions resulted in improved LDFT production. This suggests glucose inhibits lactose import whereas glycerol allows for lactose import through sufficient *lacY* expression. Although glucose is a traditional carbon feedstock for microbial fermentation, it is unsuitable for HMO production systems that use lactose as a substrate. In the absence of CCR, LDFT production was still limited by the native expression levels of *lacY* and *fucP* (Figure 2.2.3). Additional expression of *fucP* increased LDFT production by 6.9-fold to 0.9 g/L (Figure 2.2.3), indicating that L-fucose import was one of the bottlenecks for LDFT production. While native expression levels of *lacY* without CCR were adequate for supplying lactose, overexpression of *lacY* and *fucU* further balanced the donor-acceptor substrate ratio and improved LDFT titers to 1.1 g/L in 24 h (Figure 2.2.3).

Lastly, balancing expression levels of the LDFT biosynthetic pathway genes (*fkp*, *wbgL,* and *Hp3/4ft*) was critical for efficient LDFT production. Decreasing expression of *fkp* reduces excessive ATP and GTP consumption in GDP-fucose production, potentially relieving the metabolic burden of regenerating nucleotide cofactors (Figure

2.2.4a). Decreasing expression of *wbgL* helps synchronize 2'-FL production with Hp3/4FT's slower turnover rate, streamlining 2'-FL towards LDFT production (Fig 4a). Decreasing or delaying *Hp3/4ft* expression causes build-up of 2'-FL, which is rapidly exported from the cell (Figure 2.2.4b). It has been hypothesized that LacY is an importer for 2'-FL[44], but enhanced *lacY* expression was still insufficient for LDFT production from 2'-FL feeding (Figure S5). Expression of additional heterologous importers may improve 2-FL transport. Fucosyllactose transporters have been identified in gut prebiotic *Bifidobacterium* species and are ideal candidates for screening in further studies to improve LDFT production[45].

Due to concerns about strain virulence for the production of bioactive compounds, the HMO production technologies can be translated to nonpathogenic generally-recognized-as-safe (GRAS) strains such as *Bacillus subtilis*, *Corynebacterium glutamicum*, and *Saccharomyces cerevisiae*[46–48]. Advancements in GRAS strains' synthetic biology toolbox such as genome editing, vector expression systems, and tuning of gene expression has improved their industrial application in producing nutraceuticals, food additives and biofuels. Some of these GRAS hosts also enable post-translational modification of enzymes and localization of proteins into organelles or on membranes. Development of GRAS HMO fucosylation systems would also forge production routes for other fucosylated compounds for pharmaceutical research.

2.4 Methods

2.4.1 Reagents

All enzymes involved in the molecular cloning experiments were purchased from New England Biolabs (NEB). All synthetic oligonucleotides were synthesized by

Integrated DNA Technologies. Sanger sequencing was provided by Genewiz. D-Lactose was purchased from Sigma-Aldrich. L-Fucose was purchased from V-Labs, Inc. An analytical standard of 2'-FL was purchased from Carbosynth.

For synthesizing 3-FL, 8 mg lactose, L-fucose (1.3 equiv.), adenosine 5'-triphosphate (ATP, 1.3 equiv.), and guanidine 5'-triphosphate (GTP, 1.3 equiv.) were dissolved in 2.3 mL of 100 mM Tris-HCl buffer (pH 7.5) containing 20 mM $MgCl_2$, 0.35 mg *Bacteroides fragilis* bifunctional L-fucokinase/GDP-L-fucose pyrophosphorylase (BfFKP)[34], 0.15 mg *Pasteurella multocida* inorganic pyrophosphatase (PmPpA)[49], and 0.3 mg Hp3/4FT. The reaction mixture was incubated at 30 °C at 100 rpm for 16 h. The product formation was monitored by liquid chromatography-mass spectrometry (LCMS) (Shimadzu). When all lactose was converted to 3'-FL, the reaction was stopped by adding an equivalent volume of ice-cold ethanol. The mixture was kept at 4 °C for 30 min then centrifuged at 6,900 g for 30 min. The precipitates were removed and the supernatant was concentrated with a rotary evaporator and then passed through a Dowex® 1×8 ion exchange column. The partially purified product was obtained by elution with water. The eluate was concentrated, passed through a Bio-Gel P-2 gel filtration column, and eluted with water. The fractions containing the pure 3-FT product were collected and lyophilized.

To synthesize the LDFT standard, 8 mg lactose, L-fucose (1.2 equiv.), ATP (1.2 equiv.), and GTP (1.2 equiv.) were dissolved in 2.3 mL of 100 mM Tris-HCl buffer (pH 7.5) containing 20 mM $MgCl_2$, 0.3 mg BfFKP, 0.1 mg PmPpA, and 0.2 mg *Helicobacter mustelae* α1–2-fucosyltransferase (Hm2FT) [50]. The reaction mixture was incubated at 30 °C at 100 rpm for 16 h. The product formation was monitored by LCMS. When all

lactose was converted to 2'-FL, the reaction mixture was concentrated and applied to the next fucosylation step without purification. In the second step, the reaction mixture containing 10 mM 2'-FL formed from the previous step, L-fucose (1.2 equiv.), ATP (1.2 equiv.), and GTP (1.2 equiv.) in 2.3 mL of 100 mM Tris-HCl buffer (pH 7.5) containing 20 mM $MgCl_2$, 0.35 mg BfFKP, 0.15 mg PmPpA, and 0.3 mg Hp3/4FT. The reaction mixture was incubated at 30 °C at 100 rpm for 16 h. When all 2'-FL was converted to LDFT as monitored by LCMS, the reaction was stopped by adding an equal volume of ice-cold ethanol. The mixture was kept at 4 °C for 30 min and then centrifuged at 6,900 g for 30 min. The precipitates were removed and the supernatant was concentrated with a rotary evaporator and then passed through a Dowex® 1×8 ion exchange column. The partially purified product was obtained by elution with water. The eluate was concentrated, passed through a Bio-Gel P-2 gel filtration column, and eluted with water. The fractions containing the pure LDFT product were collected and lyophilized.

2.4.2 Strains and plasmids

All strains used in this study are listed in Tables 1 and S1. All plasmids and primers are listed on Tables S2 and S3. Gene deletions and integrations were constructed using CRISPR-Cas9-mediated homologous recombination[51]. Linear DNA repair fragments for gene deletions were constructed by PCR assembly or amplification from genomic DNA using primers listed in Tables S2 and S4. The linear DNA repair fragment for *ss9*::P_{lacUV5}:*T7rnap* was PCR amplified from repair plasmid pAL1856 constructed from pSS9 template (Addgene plasmid #71655)[41] listed in Table S1 and S4. All genomic modifications were PCR and sequence verified.

Plasmids for sfGFP fluorescence assays, LDFT production, and 3-FL production were constructed using sequence and ligation independent cloning (SLIC)[52]. Plasmids encoding sgRNAs for CRISPR-Cas9-mediated homologous recombination were constructed with Q5 site-directed mutagenesis using a modified template pTargetF (Addgene plasmid # 62226). Templates used for DNA amplification and cloning are listed in Table S5. All plasmids were verified by PCR and Sanger sequencing.

2.4.3. Culture conditions

Overnight cultures were grown at 37 °C, 250 rpm, in 3 mL of Luria-Bertani (LB) media with appropriate antibiotics. Antibiotic concentrations were as follows: spectinomycin (50 μg/mL), ampicillin (200 μg/mL), and kanamycin (50 μg/mL). Growth assays were carried out in M9 minimal medium (33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 9.4 mM NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$) including 1000 × A5 trace metal mix (2.86 g H_3BO_3, 1.81 g $MnCl_2{\cdot}4H_2O$, 0.079 g $CuSO_4{\cdot}5H_2O$, 49.4 mg $Co(NO_3)_2{\cdot}6H_2O$ per liter water). LDFT production was carried out in M9 minimal medium supplemented with 5 g/L yeast extract (M9P). Optical densities were measured at 600 nm (OD_{600}) with a Synergy H1 hybrid plate reader (BioTek Instruments, Inc.).

2.4.4. Growth Assays

Overnight cultures were inoculated at 1% in 3 mL of M9 minimal medium supplemented with 1 g/L D-lactose or 1 g/L L-fucose. Cultures were grown at 37 °C, 250 rpm, for 24 h and OD_{600} was measured.

2.4.5. Fluorescence Assays

Overnight cultures were inoculated at 1% in 3 mL of LB media and grown at 37 °C, 250 rpm, until OD_{600} reached 0.4–0.6. Cultures were respectively induced with IPTG

(1.0 mM) and grown at 37 °C, 250 rpm, for 24 h. Fluorescence emission was measured at 510 nm with a Synergy H1 hybrid plate reader (BioTek Instruments, Inc.).

2.4.6. LDFT production

Overnight cultures were inoculated at 1% in 3 mL of M9P supplemented with 5 g/L glucose, 10 g/L glycerol, or 20 g/L glycerol. Cultures were grown at 37 °C, 250 rpm, until OD_{600} reached 0.4–0.6. Appropriate concentrations of lactose, L-fucose, IPTG, and anhydrotetracycline (aTc) were added and the cultures were grown at 30 °C, 250 rpm, for 24 h. The produced LDFT was confirmed by high resolution electrospray ionization mass spectrometry using a Thermo Electron LTQ-Orbitrap Hybrid MS at the Mass Spectrometry Facility in the University of California, Davis.

2.4.7. HPLC Analysis

To measure glycerol, L-fucose, lactose, 2'-FL, 3-FL, and LDFT, cell culture supernatant was analyzed using HPLC (Shimadzu) equipped with a refractive index detector (RID) 10 A and a Luna Omega HILIC Sugar column (Phenomenex). The mobile phase consisted of 100% 70:30 HPLC-grade acetonitrile:MilliQ water was run at a flow rate of 1.0 mL/min for 12 min, with the column oven at 35 °C and RID cell temperature at 40 °C.

To prepare samples for HPLC analysis, 125 μL of culture was collected and spun down at 17,000 g for 5 min. 15 μL of culture supernatant or compound standard in water was diluted with 45 μL of MilliQ water and 180 μL of acetonitrile. The mixture was vortexed and spun down at 17,000 g for 5 min. 40 μL of each sample was injected into the column for analysis.

2.5 Supplementary Information

Table 1 Strain list

Strain no.	*E. coli* strain	Plasmid	Key Genotype	Reference
1	AL3535		As BL21 Star (DE3), but Δ*lacZ*	This study
2	AL3535	pAL1779/pAL1817	Δ*lacZ, P*T7:*fkp-wbgL*, *P*T7:*Hp3/4ft*	This study
3	BL21 Star (DE3)	pAL1834	*P*T7:*sfgfp*	This study
4	AL3535	pAL1834	Δ*lacZ, P*T7: *sfgfp*	This study
5	AL3600		As AL62, but *ss9*::*P*lacUV5:*T7rnap*	This study
6	AL3601		As AL1050, but *ss9::PlacUV5*:*T7rnap*	This study
7	AL3600	pAL1834	*P*T7:*sfgfp*	This study
8	AL3601	pAL1834	*P*T7:*sfgfp*	This study
9	AL3606		As Strain 6, but Δ*fucU*	This study
10	AL3659		As Strain 9, but Δ*lacZ*	This study
11	AL3659	pAL1779/pAL1817	Δ*fucU,* Δ*lacZ, P*T7:*fkp-wbgL*	This study
12	AL3659	pAL1834	Δ*fucU,* Δ*lacZ, P*T7:*sfgfp*	This study
13	AL3585		As AL1050, but Δ*fucU*	This study
14	AL3664		As Strain 13, but Δ*lacZ*	This study
15	AL3732		As Strain 14, but *ss9*::*P*lacUV5:*T7rnap*	This study
16	AL3732	pAL1834	Δ*fucU,* Δ*lacZ, ss9*::*P*lacUV5:*T7rnap, P*T7:*sfgfp*	This study
17	AL1050	pAL421	*P*LlacO1:*sfgfp*	This study
18	AL3664	pAL421	Δ*fucU,* Δ*lacZ, P*LlacO1:*sfgfp*	This study
19	AL3732	pAL421	Δ*fucU,* Δ*lacZ, ss9*::*P*lacUV5:*T7rnap, P*LlacO1:*sfgfp*	This study
20	AL1050	pAL2054	*P*lacUV5:*sfgfp*	This study
21	AL3585	pAL2054	Δ*fucU, P*lacUV5:*sfgfp*	This study
22	AL3664	pAL2054	Δ*fucU,* Δ*lacZ, P*lacUV5:*sfgfp*	This study
23	AL3664	pAL1759/pAL1760	Δ*fucU,* Δ*lacZ, P*LlacO1:*fkp-wbgL, P*LtetO1:*Hp3/4ft*	This study
24	AL3664	pAL2027/pAL1760	Δ*fucU,* Δ*lacZ, P*LlacO1:*fkp-wbgL, BBa_K1824896:lacY, P*LtetO1:*Hp3/4ft*	This study
25	AL3664	pAL2028/pAL1760	Δ*fucU,* Δ*lacZ, P*LlacO1:*fkp-wbgL, BBa_K1824896:fucP, P*LtetO1:*Hp3/4ft*	This study
26	AL3664	pAL2029/pAL1760	Δ*fucU,* Δ*lacZ, P*LlacO1:*fkp-wbgL, BBa_K1824896:lacY-fucP, P*LtetO1:*Hp3/4ft*	This study
27	AL3664	pAL2059/pAL1760	Δ*fucU,* Δ*lacZ, P*LlacO1:*fkp, BBa_K1824896:lacY-fucP, P*LtetO1:*Hp3/4ft*	This study

Table S2. Strains used in this study

Strain	Genotype	Source
XL-1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [*F´ proAB lacIq ZΔM15 Tn10* (*tet*r)]	Agilent (Santa Clara, CA)

BL21 Star (DE3) (AL15)	F$^-$ *omp*T *hsd*S_B (r_B^-, m_B^-) *gal dcm rne*131 (DE3)	ThermoFisher (Waltham, MA)
BW25113 *Z1* (AL62)	*lacI*$^+$*rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ *rph-1* Δ*(araB–D)567* Δ*(rhaD–B)568* Δ*lacZ4787(::rrnB-3) hsdR514 rph-1 attB::lacI*^{q} *tetR spec*^{r}	This study
MG1655 *Z1* (AL1050)	*F- lambda- ilvG- rfb-50 rph-1 attB::lacI*^{q} *tetR spec*^{r}	Rodriguez et al. 2015[53]
AL3271	As BW25113, but F′ [*proAB lacIq ZΔM15 Tn10* (*tet*^{r})] Δ*fucU*	This study
AL3535	As BL21 Star (DE3), but Δ*lacZ*	This study
AL3585	As AL1050, but Δ*fucU*	This study
AL3600	As AL62, but *ss9*::P_{lacUV5}:*T7rnap*	This study
AL3601	As AL1050, but *ss9*::P_{lacUV5}:*T7rnap*	This study
AL3606	As AL3601, but Δ*fucU*	This study
AL3659	As AL3606, but Δ*lacZ*	This study
AL3664	As AL3585, but Δ*lacZ*	This study
AL3732	As AL3664, but *ss9*::P_{lacUV5}:*T7rnap*	This study

Table S3. Plasmids used in this study

Plasmid	Genotype	Source
pCas	P_{cas}:*cas9* P_{araB}:*Red lacI*^{q} P_{trc}:*sgRNA pMB1 repA101*(Ts) *kan*^{r}	Addgene #62225 [51]
pTargetF	*sgRNA-pmB1 pMB1 spec*^{r}	Addgene #62226 [51]
pss9 integration template	HR1$^{\#}$-P_{T7A1}:*gfpUV*-HR2$^{\wedge}$ *pBR322 tet*^{r}	Addgene #71655 [41]
pAL421	P_{LlacO1}:*sfgfp* ColE1 *amp*^{r}	This study
pAL631	P_{LlacO1}:*sfgfp* ColE1 *kan*^{r}	This study
pAL1023	P_{LtetO1} ColA *kan*^{r}	This study
pAL1354	P_{LlacO1} ColE1 *amp*^{r}	This study
pAL1687	P_{T7}:*fkp* pBR322 *amp*^{r}	This study
pAL1688	P_{T7}:*wbgL* pBR322 *amp*^{r}	This study
pAL1689	P_{T7}:*Hp3/4ft* pBR322 *amp*^{r}	This study
pAL1759	P_{LlacO1}:*fkp-wbgL* ColE1 *amp*^{r}	This study
pAL1760	P_{LtetO1}:*Hp3/4ft* ColA *kan*^{r}	This study
pAL1762	*sgRNA-ss9* pMB1 *spec*^{r}	This study
pAL1779	P_{T7}:*fkp-wbgL* pBR322 *amp*^{r}	This study
pAL1817	P_{T7}:*Hp3/4ft* ColA *kan*^{r}	This study
pAL1783	HR1-P_{T7A1}:*gfpUV*-HR2 pBR322 *amp*^{r}	This study
pAL1834	P_{T7}:*sfgfp* pBR322 *amp*^{r}	This study
pAL1845	Δ*lacZ* HR1$^{\#}$-HR2$^{\wedge}$ ColE1 *amp*^{r}	This study
pAL1846	*sgRNA-lacZ* pMB1 *spec*^{r}	This study
pAL1851	*sgRNA-lacZ* pMB1 *amp*^{r}	This study
pAL1853	*sgRNA-ss9* pMB1 *amp*^{r}	This study

pAL1854	P_{lacUV5}:*lacZα-T7rnap* ColE1 *amp^r*	This study
pAL1855	P_{lacUV5}:*T7rnap* ColE1 *amp^r*	This study
pAL1856	HR1-P_{lacUV5}:*T7rnap*-HR2 pBR322 *amp^r*	This study
pAL1864	*sgRNA-fucU* pMB1 *amp^r*	This study
pAL2026	*BBa_K1824896*, $P_L$lacO1:fkp-wbgL colE1 amp^r*	This study
pAL2027	*BBa_K1824896:lacY, P_{LlacO1}:fkp-wbgL* ColE1 *amp^r*	This study
pAL2028	*BBa_K1824896:fucP, P_{LlacO1}:fkp-wbgL* ColE1 *amp^r*	This study
pAL2029	*BBa_K1824896:lacY-fucP, P_{LlacO1}:fkp-wbgL* ColE1 *amp^r*	This study
pAL2054	P_{lacUV5}*:sfgfp* ColE1 *amp^r*	This study
pAL2059	*BBa_K1824896:lacY, P_{LlacO1}:fkp* ColE1 *amp^r*	This study

#upstream homologous region, ^downstream homologous region, *iGEM part #: BBa_K1824896

Table S4. Oligonucleotides used in this study

Name	Sequence 5' → 3'
AZ52	GTCTTGTCGATCAGGATGATC
AZ55	CGAGCCCGTATAAACTGAAAGC
AZ56	CTAGGTCTAGGGCGGCGGATTTG
AZ57	CGTAAGATACTGACAGAAAACGC
AZ60	GGAGGAAGGAAAGAATATCTGG
AZ61	GTGACTTTATTGGCTGCTATTCC
AZ64	CAAATAGGGGTTCCGCGCACAT
AZ65	GATATGACTGTTCTCGATCCA
AZ82	CCCTGGCAAATGTTGATTGA
AZ83	CAGGCTGTTACCAAAGAAGT
AZ105	CGGCCTTATTGTCTCTCTGC
AZ154	CCTAGGTCTAGGGCGGCGGATTTG
AZ155	CATTATAACATTCTTCAAGCAGCC
AZ224	AATTCATTAAAGAGGAGAAAAGATATACCATGGGCAGCAG
AZ225	CATATGTATATCTCCTTCTTTTATGATCGTGATACTTGGAATC
AZ226	AAGAAGGAGATATACATATGAGCATTATTCG
AZ227	TTAGCAGCCGGATCTCAGTG
AZ228	CACTGAGATCCGGCTGCTAAGGTACCTAATCTAGAGGCATC
AZ229	TTTCTCCTCTTTAATGAATTCGGTCAGTGCGTCC
AZ230	AATTCATTAAAGAGGAGAAACATATGTTCCAACCGCTGCTG
AZ231	CTCTAGAGTCATTAGGTACCGCTTTGTTAGCAGCCGGATC
AZ233	TTTCTCCTCTTTAATGAATTCGG
AZ259	GAATTCGGTCAGTGCGTCCTGCTG
AZ274	AAGGATCCGGCTGCTAACAAAAGGAGATATACATATGAGC

AZ275	ACTCAGCTTCCTTTCGGGCTAGCAGCCGGATCTCAGTG
AZ276	AGCCCGAAAGGAAGCTGAGTTGGCTGCTG
AZ277	TTGTTAGCAGCCGGATCCTTATGATCGTGATACTTG
AZ293	ATGATTGAACAAGATGGATTGCACGC
AZ294	AGGAGAGCGTTCACCGACAAAACGCCAGCAACGCGG
AZ295	AATCCATCTTGTTCAATCATACTCTTCCTTTTTCAATATTATTGAAGC ATTTATCAGGG
AZ307	CACTTTACTACCCACGCCGC
AZ308	GACTGGCAGCAACAGGTGGC
AZ309	GTTGAGCTACAGGCGGTCAG
AZ310	ATTTACTAACTGGAAGAGGC
AZ311	CATTGAGTCAACCGGAATGG
AZ312	AAACCAATCGGTAAGGAAGG
AZ313	TTTTACCGTTCACGCGCTGG
AZ336	TGGTGCCGCGCGGCAGCCATATGGGTCATCACCACCATCATC
AZ337	TTCGGGCTAGCAGCCGGATCTTATTTGTACAGTTCGTCCATGCCG
AZ338	GATCCGGCTGCTAGCCCGAAAGGAAGCTGAGTTGGCTG
AZ339	ATGGCTGCCGCGCGGCACCAG
AZ340	AATGCGCGCCATTACCGAGTCCG
AZ341	AGCTGTTTCCTGTGTGAAATTGTTATCCGC
AZ342	ATTTCACACAGGAAACAGCTTAATAACCGGGCAGGCCATGTCTG
AZ343	ACTTTCTCAATAAATGCCTCTACTGCTGGCGCACC
AZ344	GAGGCATTTATTGAGAAAGTTAATCTAGAGGCATCAAATAAAACGA AAGGCTCAGTCG
AZ345	ACTCGGTAATGGCGCGCATTGGTCAGTGCGTCCTGCTGATG
AZ347	GCCGACACCAGTTTTAGAGCTAGAAATAG
AZ348	TCCGCCGCCTACTAGTATTATACCTAGGACTGAG
AZ359	CAGCGGTGGAGTGCAATGTCATGAGTATTCAACATTTCCG
AZ360	ATCGACTGGCGAGCGGCATCTTACCAATGCTTAATCAGTG
AZ361	GATGCCGCTCGCCAGTCGATTGGC
AZ362	GACATTGCACTCCACCGCTGATGAC
AZ364	TCCGGATTTACTAACTGGAAGAGGCACTAAATG
AZ365	AGCTGTTTCCTGTGTGAAATTGTTATCCGCTC
AZ366	CCTTTCGTCTTCACCTCGAGTCACTCATTAGGCACCCCAGGC
AZ367	GGTACCTTAGCAGCCGGATCTTACGCGAACGCGAAGTCCGAC
AZ368	GATCCGGCTGCTAAGGTACCTAATCTAGAGGC
AZ369	CTCGAGGTGAAGACGAAAGGGCCT
AZ370	AGTTGATATGTCAAACAGGTTCACTCATTAGGCACCCCAGGC
AZ371	CGGCGCTCAGTTGGAATTCAACAACAGATAAAACGAAAGGCC
AZ372	TGAATTCCAACTGAGCGCCGGTC
AZ373	ACCTGTTTGACATATCAACTGCGCC
AZ384	GTGATGATGGGTTTTAGAGCTAGAAATAGC
AZ385	CAGCGGCGGTACTAGTATTATACCTAGGAC
AZ403	ACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG
AZ411	CGCGCGGCACACTAGTATTATACCTAGGAC

AZ710	GTGCCACCTGACGTCTAAGACTAGTACTCTAGTATTTCTCCTCTTTA
AZ711	GCTACTAGAGTACTAGAGTACTAGAGATTAAAGAGGAGAAATACTA GAGTACTAGTCTTA
AZ712	TCTCTAGTACTCTAGTACTCTAGTAGCTAGCACTGTACCTAGGACT GAGCTAGCCGT
AZ713	ACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTTGACGG CTAGCTCAGTCC
AZ714	TGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTT CACCTCGAGAAT
AZ715	TTGTTATCCGCTCACAATGTCAATTGTTATCCGCTCACAATTCTCGA GGTGAAGACGAA
AZ716	CAATTGACATTGTGAGCGGATAACAAG
AZ717	TCTTAGACGTCAGGTGGCACTTTTCG
AZ718	GTGCCACCTGACGTCTAAGATTAAGCGACTTCATTCACCTG
AZ719	GAGAAATACTAGAGTACTAGATGTACTATTTAAAAAACACAAACTTT TGGATG
AZ720	CTAGTACTCTAGTATTTCTCCTCTTTAATCTCTAGTAC
AZ721	GTGCCACCTGACGTCTAAGATCAGTTAGTTGCCGTTTGAGAAC
AZ722	GAGAAATACTAGAGTACTAGATGGGAAACACATCAATACAAACGCA GAG
AZ723	GAAAGAGGGGACAAACTAGTATGGGAAACACATCAATACAAACG
AZ724	TTGTCCCCTCTTTCTCTAGATTAAGCGACTTCATTCACCTGACG
AZ819	CTAACTGGAAGAGGCACTAAATGGGTCATCACCACCATCATCACG
AZ820	GGTACCTTAGCAGCCGGATCTTATTTGTACAGTTCGTCCATGCCG
AZ821	GATCCGGCTGCTAAGGTACCTAATC
AZ822	TTAGTGCCTCTTCCAGTTAGTAAATCCGG
AZ851	CTGCTAAGGTACCTAATCTAGAGGCATC
AZ852	CCGGATCTTATGATCGTGATACTTGGAATC
JO232	GGTTCCGCGCACATTTCCC
MMM40	GAGTCAGTGAGCGAGGAAGC
MMM131	GCTTGGTTGAGAATACGCCG
MMM132	GCCTACGATTACGCATGGCTTG
SD62	GGCCCTTTCGTCTTCACCTCGAG
SL005	AACGCAGTCAGGCACCGTGTATGAGTATTCAACATTTCCG
SL006	GAGGTGCCGCCGGCTTCCATTTACCAATGCTTAATCAGTG
SL007	ATGGAAGCCGGCGGCACCTC
SL008	ACACGGTGCCTGACTGCGTTAGC
YT167	TAATGACTCTAGAGGCATCAAATAA
YT054	TTGTCGGTGAACGCTCTCCTG
YT400	ATGGGTCATCACCACCATCATCA
YT430	CCAGTAGTAGGTTGAGGCCGTTGAG
YT092	CTACTCAGGAGAGCGTTCAC
YT101	GCTTCCCAACCTTACCAGAG
YTC427	CAAGCAGCAGATTACGCGCAG

Table S5. Guide for CRISPR-Cas9-mediate gene deletions and insertions

	pTargetF		PCR Linear Repair Fragment	
Modification	Plasmid	20 bp sgRNA sequence 5'→ 3'	Primers	Template
Δ*fucU*	pAL1864	ACCGCCGCTGGTGATGATGG	AZ82 (F), AZ83 (R)	AL3271 gDNA
Δ*lacZ*	pAL1851	AGGCGGCGGAGCCGACACCA	AZ340 (F), AZ343 (R)	pAL1845
ss9::*PlacUV5*:*T7rnap*	pAL1853	TCTGGCGCAGTTGATATGTA	MMM131 (F), MMM132 (R)	pAL1856

Table S6. Plasmid construction guide

	PCR for Vector			PCR for Insert(s)			
Plasmid	Primer (F)	Primer (R)	Template	Primer (F)	Primer (R)	Template	Sequence of Interest
pAL1759	AZ228	AZ229	pAL1354	AZ224	AZ225	pAL1687	*fkp*
				AZ226	AZ227	pAL1688	*wbgL*
pAL1760	YT167	AZ233	pAL1023	AZ230	AZ231	pAL1689	*Hp3/4ft*
pAL1779	AZ276	AZ277	pAL1687	AZ274	AZ275	pAL1688	*wbgL*
pAL1817	AZ294	AZ295	pAL1689	AZ293	YT054	pAL1023	*ColA-kan*^{r}
pAL1762*	MMM139	MMM140	pTargetF				
pAL1783	SL007	SL008	pss9	SL005	SL006	pAL1354	*amp*^{r}
pAL1845	AZ344	AZ345	pAL1354	AZ340	AZ341	AL1050 gDNA	400bp upstream HR1 *lacZ*
				AZ342	AZ343	AL1050 gDNA	400bp downstream HR2 *lacZ*
pAL1854	AZ368	AZ369	pAL1759	AZ366	AZ367	BL21 Star (DE3) gDNA	P_{lacUV5}:*lacZα T7rnap*
pAL1855*	AZ364	AZ365	pAL1854				
pAL1856	AZ372	AZ373	pAL1783	AZ370	AZ371	pAL1855	P_{lacUV5}:*T7rnap*
pAL1846*	AZ347	AZ348	pTargetT				
pAL1851	AZ361	AZ362	pAL1846	AZ359	AZ360	pAL1687	*amp*^{r}
pAL1853	AZ361	AZ362	pAL1762	AZ359	AZ360	pAL1687	*amp*^{r}
pAL1864*	AZ384	AZ385	pAL1851				
pAL1834	AZ338	AZ339	pAL1687	AZ336	AZ337	pAL421	*sfgfp*
pAL2026	AZ716	AZ717	pAL1759	AZ710, AZ712, AZ714	AZ711, AZ713, AZ715	N/A	*BBa_K1824896*
pAL2027	AZ720	AZ717	pAL2026	AZ718	AZ719	AL1050 gDNA	*lacY*
pAL2028	AZ721	AZ722	pAL2026	AZ718	AZ719	AL1050 gDNA	*fucP*
pAL2029	AZ724	AZ717	pAL2027	AZ721	AZ723	AL1050 gDNA	*fucP*
pAL2054	AZ821	AZ822	pAL1855	AZ819	AZ820	pAL631	*sfgfp*
pAL2059*	AZ851	AZ852	pAL2029				

*Q5-site directed mutagenesis (NEB).

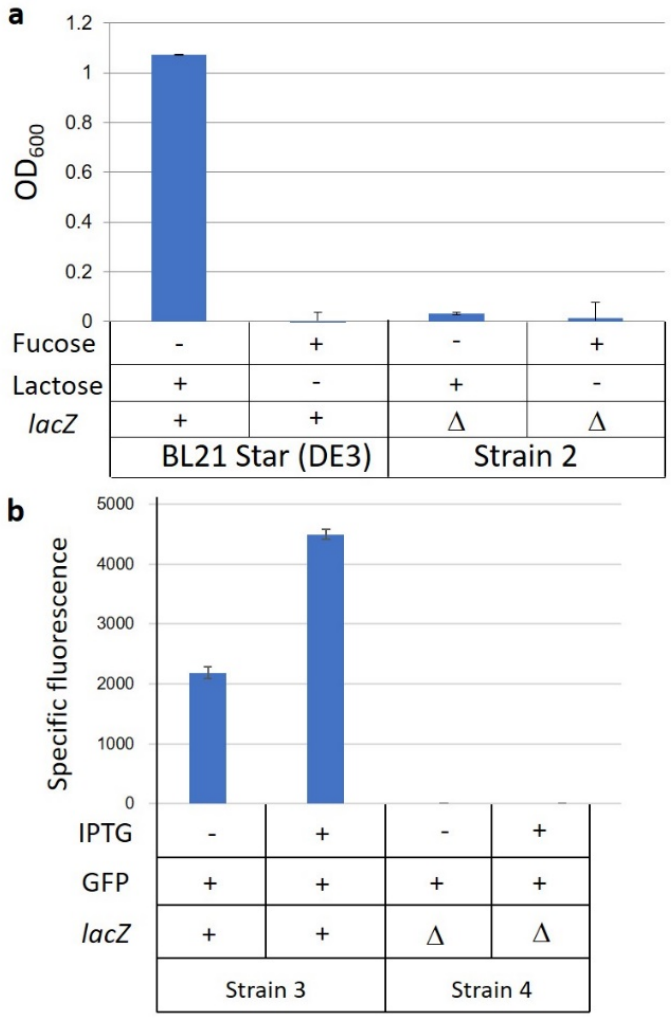


Figure S1 Modifications in *E. coli* B-strains for LDFT production. (**a**) Growth of BL21 Star (DE3) and Δ*lacZ* mutant (Strain 2, Table 1) in M9 minimal media with or without 1 g/L L-fucose or D-lactose. (**b**) Expression of P_{T7}:*sfgfp* in BL21 Star (DE3) (Strain 3) and Δ*lacZ* mutant (Strain 4, Table S1) in LB-media. Cultures were induced with or without 1 mM IPTG, respectively. Δ indicates gene was removed from the genome. Error bars indicate s.d. (n = 3 biological replicates).

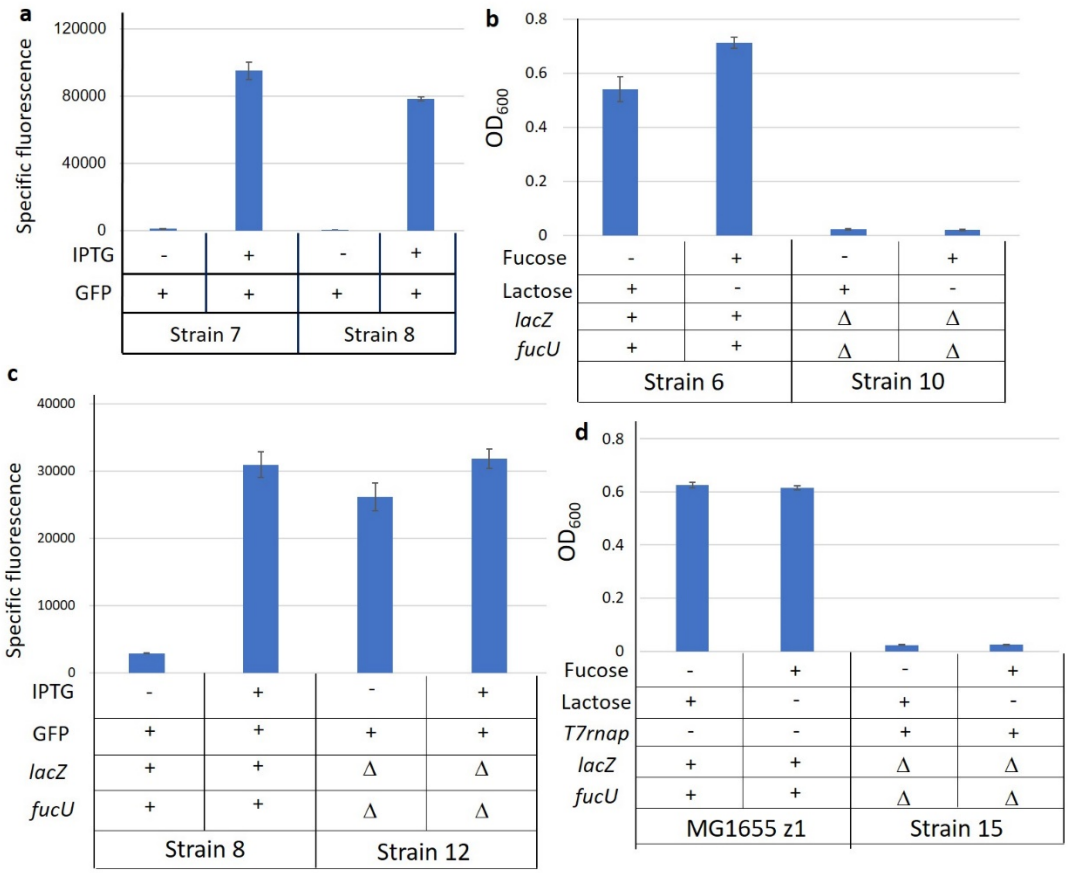


Figure S2 Installation of the T7 RNAP expression system into *E. coli* K-12-derivative strains. (**a**) GFP fluorescence assay in K-12 derivative strains, BW25113 *Z1* (Strain 7) and MG1655 *Z1* (Strain 8, Table S1) with *ss9*::*PlacUV5*:*T7rnap*. Cultures were induced with or without 1 mM IPTG at 37 °C for 24 h. (**b**) Growth assay of MG1655 *Z1* (Strain 6) and Strain 6 with Δ*fucU* Δ*lacZ* (Strain 10, Table S1) in M9-minimal media with or without 1 g/L L-fucose or D-lactose at 37 °C for 24 h. (**c**) Fluorescence assay to evaluate GFP expression from T7 promoter in Strain 8 and Strain 8 with ΔfucU ΔlacZ (Strain 12, Table S1). Cultures were grown in LB-media and induced with or without 1 mM IPTG at 37˚C for 24 h. (**d**) Growth assay of MG1655 *Z1* and MG1655 *Z1* with Δ*fucU* Δ*lacZ* (Strain 15, Table S1) in M9-minimal media with 1 g/L L-fucose or D-lactose. Δ indicates gene was removed from the genome. Error bars indicate s.d. (n = 3 biological replicates).

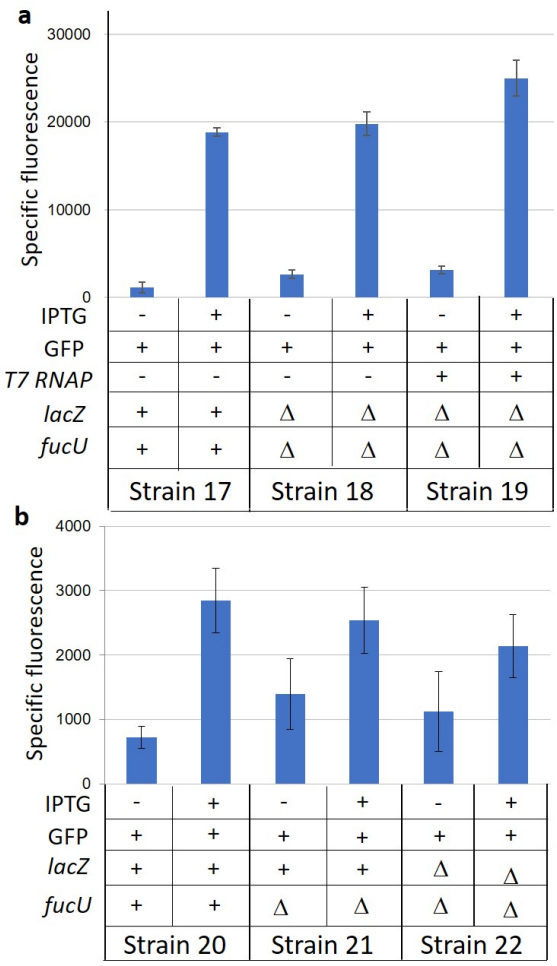


Figure S3 Fluorescence expression control under lac-promoter variants in K-12 derivative strains. Cultures were grown in LB-media and induced with or without 1 mM IPTG at 37 °C for 24 h. (**a**) GFP expression under promoter P_{LlacO1} in MG1655 *Z1* (Strain 17), MG1655 *Z1* with Δ*fucU* Δ*lacZ* (Strain 18) and Strain 18 with *ss9*::*PlacUV5*:*T7rnap* (Strain 19, Table S1). (**b**) GFP expression under promoter P_{lacUV5} in MG1655 *Z1* (Strain 20), MG1655 *Z1* with Δ*fucU* (Strain 21) and MG1655 *Z1* with Δ*fucU* Δ*lacZ* (Strain 22, Table S1). Δ indicates gene was removed from the genome. Error bars indicate s.d. (n = 3 biological replicates).

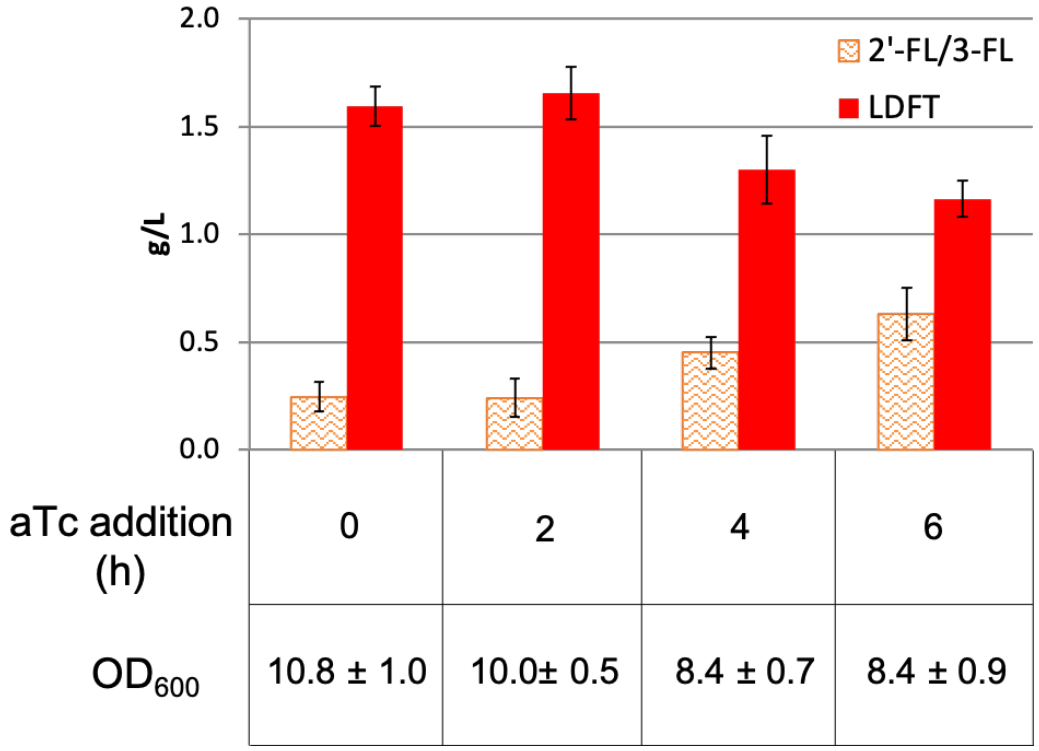


Figure S4 Delayed expression of Hp3/4ft in Strain 26. Strain 26 (MG1655 *Z1* Δ*fucU* Δ*lacZ* with the LDFT production plasmids with *lacY* and *fucU*) was grown in M9P with 10 g/L glycerol at 30 °C for 24 h. Cultures were supplemented with 1 g/L lactose and 1 g/L fucose and induced with 50 μM IPTG at 0 h. 100 ng/mL aTc was added to cultures at 0, 2, 4, and 6 h. OD_{600}, monofucosides (2'-FL/3-FL) concentration (orange zigzag) and LDFT concentration (red filled) were measured at 24 h. Error bars indicate s.d. (n = 3 biological replicates).

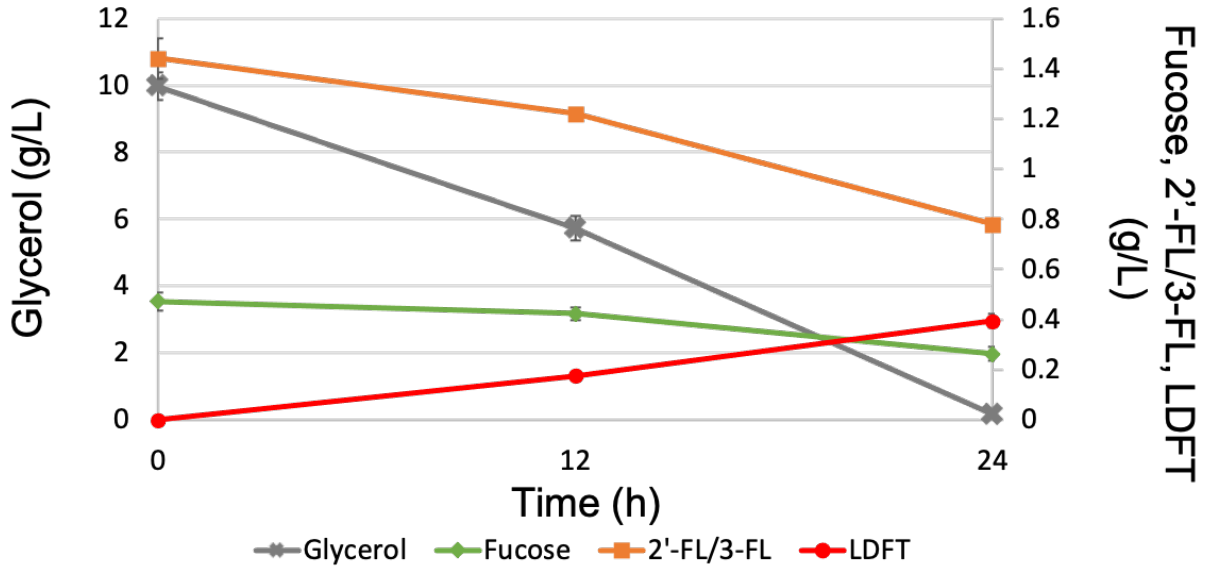


Fig. S5 LDFT production with 2'-FL feeding. The *wbgL* gene was removed from pAL2029, generating pAL2059 (Table S2). Strain 27 (MG1655 *Z1* Δ*fucU* Δ*lacZ* harboring pAL1760 (*Hp3/4ft*) and pAL2059 (*fkp*, *lacY*, and *fucU*), Table S1) was grown in M9P with 10 g/L glycerol at 30 °C for 24 h. Cultures were supplemented with 1.4 g/L 2'-FL (mole equivalent to 1 g/L lactose) and 0.5 g/L fucose and induced with 50 μM IPTG and 100 ng/mL aTc. Glycerol concentration (gray cross), fucose concentration (green diamond), 2'-FL concentration (orange square) and LDFT concentration (red circle) were monitored during the experiment. Error bars indicate s.d. (n = 3 biological replicates).

2.6 References

1. Smilowitz, J. T., Lebrilla, C. B., Mills, D. A., German, J. B. & Freeman, S. L. Breast Milk Oligosaccharides: Structure-Function Relationships in the Neonate. *Annu. Rev. Nutr.* **34**, 143–169 (2014).
2. Chen, X. Human Milk Oligosaccharides (HMOS): Structure, Function, and Enzyme-Catalyzed Synthesis. *Adv. Carbohydr. Chem. Biochem.* **72**, 113—190 (2015).
3. Yu, H. & Chen, X. CHAPTER 11 Enzymatic and Chemoenzymatic Synthesis of Human Milk Oligosaccharides (HMOS). in *Synthetic Glycomes* 254–280 (The Royal Society of Chemistry, 2019). doi:10.1039/9781788016575-00254.
4. Wiciński, M., Sawicka, E., Gębalski, J., Kubiak, K. & Malinowski, B. Human Milk Oligosaccharides: Health Benefits, Potential Applications in Infant Formulas, and Pharmacology. *Nutrients* **12**, 266 (2020).
5. Ballard, O. & Morrow, A. L. Human milk composition: nutrients and bioactive factors. *Pediatr. Clin. North Am.* **60**, 49–74 (2013).
6. Rudloff, S. & Kunz, C. Milk oligosaccharides and metabolism in infants. *Adv. Nutr.* **3**, 398S-405S (2012).
7. Kulinich, A. & Liu, L. Human milk oligosaccharides: The role in the fine-tuning of innate immune responses. *Carbohydr. Res.* **432**, 62–70 (2016).
8. Triantis, V., Bode, L. & van Neerven, R. J. J. Immunological Effects of Human Milk Oligosaccharides. *Front. Pediatr.* **6**, 190 (2018).
9. Ayechu-Muruzabal, V. *et al.* Diversity of human milk oligosaccharides and effects on early life immune development. *Frontiers in Pediatrics* vol. 6 239 (2018).
10. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).
11. Hegar, B. *et al.* The Role of Two Human Milk Oligosaccharides, 2'-Fucosyllactose and Lacto-N-Neotetraose, in Infant Nutrition. *Pediatr. Gastroenterol. Hepatol. Nutr.* **22**, 330–340 (2019).
12. Berger, P. K. *et al.* Human milk oligosaccharide 2'-fucosyllactose links feedings at 1 month to cognitive development at 24 months in infants of normal and overweight mothers. *PLoS One* **15**, e0228323 (2020).
13. Borewicz, K. *et al.* The association between breastmilk oligosaccharides and faecal microbiota in healthy breastfed infants at two, six, and twelve weeks of age. *Sci. Rep.* **10**, 4270 (2020).
14. Ágoston, K., Hederos, M., Bajza, I. & Dekany, G. Kilogram scale chemical synthesis of 2′-fucosyllactose. *Carbohydr. Res.* **476**, (2019).
15. Bandara, M. D., Stine, K. J. & Demchenko, A. V. The chemical synthesis of human milk oligosaccharides: Lacto-N-neotetraose (Galβ1→4GlcNAcβ1→3Galβ1→4Glc). *Carbohydr. Res.* **483**, 107743 (2019).
16. Bandara, M. D., Stine, K. J. & Demchenko, A. V. Chemical synthesis of human milk oligosaccharides: lacto-N-neohexaose (Galβ1 → 4GlcNAcβ1→)2 3,6Galβ1 → 4Glc. *Org. Biomol. Chem.* **18**, 1747–1753 (2020).
17. Xiao, Z. *et al.* Chemoenzymatic Synthesis of a Library of Human Milk

Oligosaccharides. *J. Org. Chem.* **81**, 5851–5865 (2016).
18. Zhao, C. *et al.* The one-pot multienzyme (OPME) synthesis of human blood group H antigens and a human milk oligosaccharide (HMOS) with highly active Thermosynechococcus elongatus α1–2-fucosyltransferase. *Chem. Commun.* **52**, 3899–3902 (2016).
19. Yu, H. *et al.* H. pylori α1–3/4-fucosyltransferase (Hp3/4FT)-catalyzed one-pot multienzyme (OPME) synthesis of Lewis antigens and human milk fucosides. *Chem. Commun.* **53**, 11012–11015 (2017).
20. Bai, J. *et al.* Biochemical characterization of Helicobacter pylori α1–3-fucosyltransferase and its application in the synthesis of fucosylated human milk oligosaccharides. *Carbohydr. Res.* **480**, 1–6 (2019).
21. McArthur, J. B., Yu, H. & Chen, X. A Bacterial β1–3-Galactosyltransferase Enables Multigram-Scale Synthesis of Human Milk Lacto-N-tetraose (LNT) and Its Fucosides. *ACS Catal.* **9**, 10721–10726 (2019).
22. Huang, D. *et al.* Metabolic engineering of Escherichia coli for the production of 2′-fucosyllactose and 3-fucosyllactose through modular pathway enhancement. *Metab. Eng.* **41**, 23–38 (2017).
23. Yu, S. *et al.* Production of a human milk oligosaccharide 2′-fucosyllactose by metabolically engineered Saccharomyces cerevisiae. *Microb. Cell Fact.* **17**, 101 (2018).
24. Choi, Y. H., Park, B. S., Seo, J.-H. & Kim, B.-G. Biosynthesis of the human milk oligosaccharide 3-fucosyllactose in metabolically engineered Escherichia coli via the salvage pathway through increasing GTP synthesis and β-galactosidase modification. *Biotechnol. Bioeng.* **116**, 3324–3332 (2019).
25. Dong, X. *et al.* Modular pathway engineering of key precursor supply pathways for lacto-N-neotetraose production in Bacillus subtilis. *Biotechnol. Biofuels* **12**, 212 (2019).
26. Baumgärtner, F., Conrad, J., Sprenger, G. A. & Albermann, C. Synthesis of the human milk oligosaccharide lacto-N-tetraose in metabolically engineered, plasmid-free E. coli. *Chembiochem* **15**, 1896–1900 (2014).
27. Liu, Y.-H. *et al.* Efficient sequential synthesis of lacto-N-triose II and lacto-N-neotetraose by a novel β-N-acetylhexosaminidase from Tyzzerella nexilis. *Food Chem.* **332**, 127438 (2020).
28. Chaturvedi, P. *et al.* Fucosylated human milk oligosaccharides vary between individuals and over the course of lactation. *Glycobiology* **11**, 365–372 (2001).
29. Orczyk-Pawiłowicz, M. & Lis-Kuberka, J. The Impact of Dietary Fucosylated Oligosaccharides and Glycoproteins of Human Milk on Infant Well-Being. *Nutrients* **12**, 1105 (2020).
30. Newburg, D. S., Tanritanir, A. C. & Chakrabarti, S. Lactodifucotetraose, a human milk oligosaccharide, attenuates platelet function and inflammatory cytokine release. *J. Thromb. Thrombolysis* **42**, 46–55 (2016).
31. Engels, L. & Elling, L. WbgL: a novel bacterial α1,2-fucosyltransferase for the synthesis of 2′-fucosyllactose. *Glycobiology* **24**, 170–178 (2014).
32. Kopp, J. *et al.* Impact of Glycerol as Carbon Source onto Specific Sugar and

Inducer Uptake Rates and Inclusion Body Productivity in E. coli BL21(DE3). *Bioeng. (Basel, Switzerland)* **5**, 1 (2017).

33. Paulsen, I. T., Chauvaux, S., Choi, P. & Saier Jr, M. H. Characterization of glucose-specific catabolite repression-resistant mutants of Bacillus subtilis: identification of a novel hexose:H+ symporter. *J. Bacteriol.* **180**, 498–504 (1998).
34. Yi, W. *et al.* Remodeling bacterial polysaccharides by metabolic pathway engineering. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4207–4212 (2009).
35. Rasko, D. A., Wang, G., Palcic, M. M. & Taylor, D. E. Cloning and Characterization of the α(1,3/4) Fucosyltransferase of Helicobacter pylori . *J. Biol. Chem.* **275**, 4988–4994 (2000).
36. Ma, B. *et al.* Purification, Kinetic Characterization, and Mapping of the Minimal Catalytic Domain and the Key Polar Groups of Helicobacter pylori α-(1,3/1,4)-Fucosyltransferases. *J. Biol. Chem.* **281**, 6385–6394 (2006).
37. Liu, J. Y., Miller, P. F., Willard, J. & Olson, E. R. Functional and Biochemical Characterization of Escherichia coli Sugar Efflux Transporters . *J. Biol. Chem.* **274**, 22977–22984 (1999).
38. Nidetzky, B., Gutmann, A. & Zhong, C. Leloir Glycosyltransferases as Biocatalysts for Chemical Production. *ACS Catal.* **8**, 6283–6300 (2018).
39. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **5**, 172 (2014).
40. Brückner, R. & Titgemeyer, F. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol. Lett.* **209**, 141–148 (2002).
41. Bassalo, M. C. *et al.* Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synth. Biol.* **5**, 561–568 (2016).
42. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
43. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613–624 (2008).
44. Shin, J. *et al.* Development of fluorescent Escherichia coli for a whole-cell sensor of 2'-fucosyllactose. *Sci. Rep.* **10**, 10514 (2020).
45. Sakanaka, M. *et al.* Evolutionary adaptation in fucosyllactose uptake systems supports bifidobacteria-infant symbiosis. *Sci. Adv.* **5**, eaaw7696 (2019).
46. Kaspar, F., Neubauer, P. & Gimpel, M. Bioactive Secondary Metabolites from Bacillus subtilis: A Comprehensive Review. *J. Nat. Prod.* **82**, 2038–2053 (2019).
47. Becker, J., Rohles, C. M. & Wittmann, C. Metabolically engineered Corynebacterium glutamicum for bio-based production of chemicals, fuels, materials, and healthcare products. *Metab. Eng.* **50**, 122–141 (2018).
48. Lian, J., Mishra, S. & Zhao, H. Recent advances in metabolic engineering of Saccharomyces cerevisiae: New tools and their applications. *Metab. Eng.* **50**, 85–108 (2018).
49. Yu, H. *et al.* Highly efficient chemoenzymatic synthesis of β1-3-linked galactosides. *Chem. Commun. (Camb).* **46**, 7507–7509 (2010).

50. Ye, J. *et al.* Reprogramming the enzymatic assembly line for site-specific fucosylation. *Nat. Catal.* **2**, 514–522 (2019).
51. Jiang, Y. *et al.* Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
52. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
53. Rodriguez, G. M., Tashiro, Y. & Atsumi, S. Expanding ester biosynthesis in Escherichia coli. *Nat. Chem. Biol.* **10**, 259–265 (2014).

Chapter 3: Microbial Conversion of D-Glucose into A Rare Sugar D-Psicose in *Escherichia coli*

3.1 Introduction

By 2026, the global market for rare sugars is anticipated to grow to 1.65 billion USD as they are used to replace conventional corn syrup or synthetic sweeteners in food product[1,2]. These rare sugars are monosaccharides that are found in low abundancy in nature and can exhibit physiological functions that modulate health rather than serving as an energy source[3–5]. This heightened traction for alternative sugar substitutes is propelled by health-conscious consumers' demand for low-calorie functional foods that are not only appetizing but can also prevent weight-related diseases[1,5,6]. To satisfy the needs of consumers, the food and beverage industry must establish economical and sustainable production methods for rare sugars of interest.

Due to its physiochemical and bioactive properties, rare sugar D-psicose is projected to dominate the alternative sugar market[2]. D-psicose is the ketohexose C3 epimer of D-fructose and is found in minute quantities in figs and wheat[4,7,8]. It only contains one-tenth of the calories of glucose but is comparable to sucrose's sweetness. Additionally, its browning potential, moisture retention, and highly solubility[6–8] makes it an excellent candidate for substituting sucrose or glucose in baked goods and beverages. It has also been shown in clinical studies that D-psicose has a low glycemic index, stabilizes insulin response in diabetics, and has the potential to combat obesity by preventing abdominal fat accumulation[9,10]. Although D-psicose can be extracted from fruit and grains, dedicating arable land to grow these food sources for sugar production directly competes with agriculture for global food supply chains[11]. The low abundancy of

D-psicose and the presence of other naturally produced carbohydrates also complicates separation and purification processes, making sugar extraction inefficient and unsustainable for long-term applications in the industry.

As an alternative method to increase its supply, two types of *in-vitro* enzymatic pathways have been established to synthesize D-psicose[12,13]. In isomerase-epimerase reaction systems, D-glucose is first converted by a C2-isomerase into D-fructose which is then converted into D-psicose by a C3-epimerase[12,1314,15]. Unfortunately, these reactions systems require temperature conditions between 40 ºC to 65 ºC to overcome the poor thermodynamics of the two highly reversible enzymatic reactions and results in low space-time yields during fermentation scale-up[12,13]. The instability of enzymes over time at high temperatures also adds to the manufacturing costs of frequently regenerating the catalysts[16]. While efforts have been made to identify or engineer more thermostable isomerases and epimerases and to stabilize the enzymes through immobilization strategies[16–19], this type of *in-vitro* enzymatic synthesis of D-psicose from D-glucose is still limited to 40% yield with significant amounts of D-fructose accumulating as an intermediate. In dihydroxyacetone phosphate (DHAP)-dependent aldolase systems, the reversible enantioselective addition of the ketone donor DHAP to the aldehyde acceptor glyceraldehyde generates psicose-1-phosphate that is subsequently dephosphorylated by a C1-phosphatase into free D-psicose[14,15]. Although DHAP-dependent aldolases are extensively characterized for rare sugar synthesis[20], its substrate DHAP is highly unstable, where it can isomerize into glyceraldehyde-3-phosphate or degrade into methyglyoxal upon deprotonation[21]. Furthermore, DHAP's

enediolate intermediate can lead to the formation of the diastereomer byproduct D-sorbose[14,15].

To overcome the thermodynamic limits and substrate instability of current *in-vitro* synthesis schemes, the Atsumi Lab and Siegel Lab have proposed a production pathway in *E. coli* that generates a thermodynamic sink to drive the D-glucose conversion into D-psicose via phosphorylated intermediates (Figure 3.1.1). First, the irreversible phosphorylation of the stable substrate D-glucose into glucose-6-phsophate (G6P) by the native glucose phosphotransferase system (PTS) occurs within *E. coli*'s cytoplasmic space[22,23]. Using *E. coli*'s native phosphoglucoisomerase Pgi of the Embden–Meyerhof–Parnas pathway (EMPP), G6P is isomerized into fructose-6-phosphate (F6P)[24,25]. By expressing a gene encoding for a C3-epimerase and a gene encoding for a C6 phosphatase, F6P can be epimerized by the C3-epimerase into psicose-6-phosphate (P6P) which is irreversibly dephosphorylated by the C6-phosphatase into D-psicose. Due to its structural similarity to D-glucose, D-psicose can theoretically be secreted by natively encoded sugar exporters to culture supernatant and isolated by downstream purification processes[26,27].

In this project, we identified and overexpressed the genes of *E. coli*'s native psicose-6-phosphate-3-epimerase AlsE[28–30] along with the native hexitol phosphatase YniC[31] for our psicose production module. Using a P_{T7} promoter expression system[32,33], we initially produced D-psicose at a titer of 0.52 g/L from 10 g/L glucose in 24 h. To streamline D-glucose carbon flux towards F6P, we genetically modified our host by knocking out the genes *pfkA*, *zwf*, and *rpiB* of the EMPP, Enter-Doudoroff pathway (EDP) and pentose phosphate pathway (PPP), and allose degradation pathway,

respectively, using a CRISPR-Cas9 editing system[34]. With fine-tuning of the P_{T7} system, this triple knockout strain produced 2.6 g/L of D-psicose from 10 g/L D-glucose in 24 h. Alternative promoter systems and high-density culturing conditions were explored and D-psicose titers were enhanced to 3.7 g/L in 24 h.

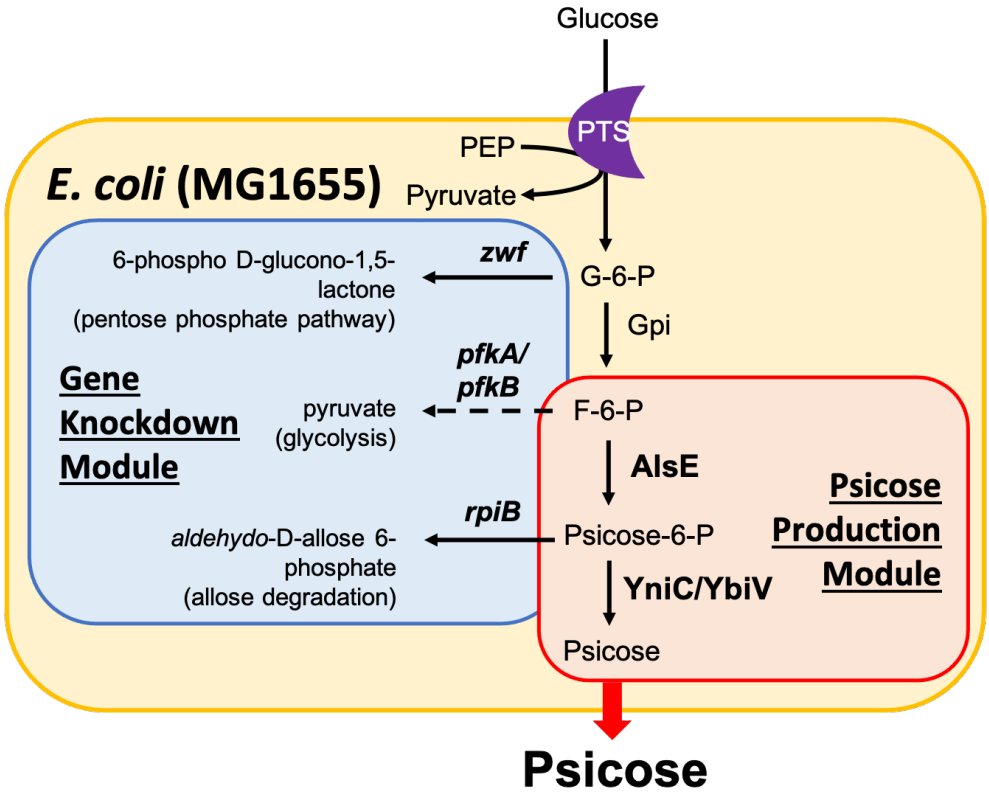


Figure 3.1.1 Pathway overview of D-psicose production. D-Glucose is simultaneously phosphorylated and transported across the inner cell membrane by the glucose phosphotransferase system (PTS) to form glucose-6-phosphate (G6P), which is isomerized by glucophosphoisomerase (Gpi) into fructose-6-phosphate (F6P). With overexpression of psicose-6-phosphate-3-epimerase (AlsE) and a phosphatase (YniC or YbiV), F6P is converted into psicose-6-phosphate (P6P) and dephosphorylated into D-psicose.

3.2 Results

3.2.1 Establishing a psicose production pathway

To produce D-psicose in *E. coli*, we needed to identify a C3 epimerase that can convert F6P into P6P and a C6 phosphatase to dephosphorylate P6P into free D-

psicose (Figure 3.1.1). Although *E. coli* does not natively produce D-psicose as a major metabolite, it does encode for the D-allose degradation pathway that produces D-psicose-6-phosphate (P6P), which is epimerized into F6P by D-psicose-6-phosphate-3-epimerase AlsE[28,30]. As for C6 phosphatases, *E. coli* encodes for a superfamily of haloacid dehalogenase (HAD)-like hydrolases that possess phosphatase activity towards various small molecules[31]. HADs YniC and YbiV were previously determined in *in-vitro assays* to have low activity towards G6P and F6P, making them of interest to test P6P's dephosphorylation into D-psicose.

For our psicose production module, we cloned the genes *alsE* and *yniC* or *ybiV*, respectively, into a pET-16b vector under a T7-promoter (P_{T7}, pAL1946 and pAL1947). To enable transcription of a P_{T7} promoter in *E. coli* strain MG1655, we integrated into the *ss9* intergenic locus the gene *T7 rnap*, which encodes for a bacteriophage T7 RNA polymerase, downstream of a P_{lacUV5} promoter. We then used P1 phage transduction to integrate a *Z1* fragment that expresses *lacI^q,* which increases the production of a lac repressor LacI^q that regulates the P_{lacUV5} promoter. These two modifications to MG1655 formed Strain AL3601 (Table S2). To test D-psicose production, plasmids pET-16b, pAL1946, and pAL1947 were introduced into AL3601, forming Strains 1, 2 and 3 (Table S1), respectively. From 10 g/L D-glucose, Strain 2 produced 0.74 g/L of D-psicose (7.4% theoretical maximum yield (TMY)) and Strain 3 produced 0.42 g/L of D-psicose in 24 h, while Strain 1 did not produce any detectable amount of D-psicose (Figure 3.2.1). Due to the higher titer of combining AlsE with YniC, we continued using the pAL1946 psicose production module in the next production experiments.

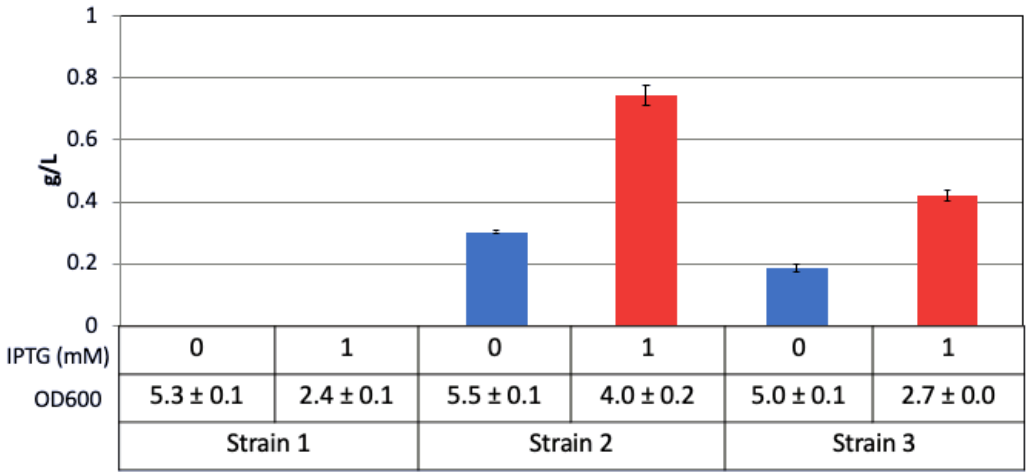


Figure 3.2.1 D-Psicose production with epimerase AlsE and phosphatases YniC or YbiV. Strains 1, 2, and 3 (Table S1) were grown in M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced with or without 1 mM IPTG. Product formation was determined at 24 h. Error bars indicate s.d. (n = 3 biological replicates).

3.2.2 Streamlining carbon flux to psicose production

To conserve intracellular F6P for psicose production, we identified metabolic branchpoints that divert F6P away to central carbon metabolism to eliminate from our production host. For decreasing the flux of F6P into the EMPP, we deleted the gene *pfkB* and *pfkA* from AL3601 (Strains 4 and 5; Table S1), which encode for phosphofructokinases PfkB and PfkA, respectively, that convert F6P into fructose-1,6-bisphosphate (F16BP)[35–37]. We introduced pAL1946 to Strains 4 and 5 to form Strains 6 and 7 (Table S1). After culturing Strains 6 and 7 with 10 g/L glucose for 24 h, we produced 0.62 g/L D-psicose in Strain 6 under the induced condition and 0.66 g/L D-psicose (6.6% TMY) in Strain 7 under the uninduced condition (Figure 3.2.2).

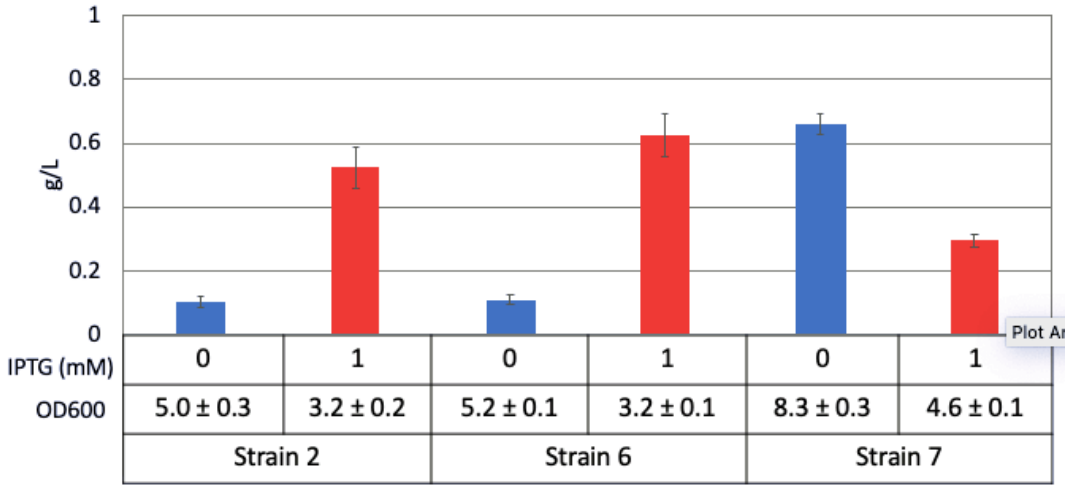


Figure 3.2.2 D-Psicose production with AlsE and YniC in phosphofructokinase knockout strains. Strains 2, 6, and 7 (Table S1) were grown in M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced with and without 1 mM IPTG. Product formation was determined at 24 h. Error bars indicate s.d. (n = 3 biological replicates).

We also identified metabolic branchpoints of G6P and P6P to remove from Strains 6 and 7 to conserve metabolites for D-psicose production. The genes *zwf*, which encodes for a glucose-6-phosphate dehydrogenase of the EDP and PPP[24,38,39], and *rpiB*, which encodes for an allose-6-phosphate isomerase of the allose degradation pathway[40–42], were deleted from Strains 6 and 7 to form Strains 8 and 9, respectively. Psicose production plasmid pAL1946 was introduced into Strains 8 and 9 to form Strains 10 and 11 (Table S1). After culturing with 10 g/L D-glucose for 24 h, Strain 10 produced 0.57 g/L of D-psicose in the induced condition whereas Strain 11 produced 1.45 g/L (14.5% TMY) of D-psicose under the uninduced condition (Figure 3.2.3).

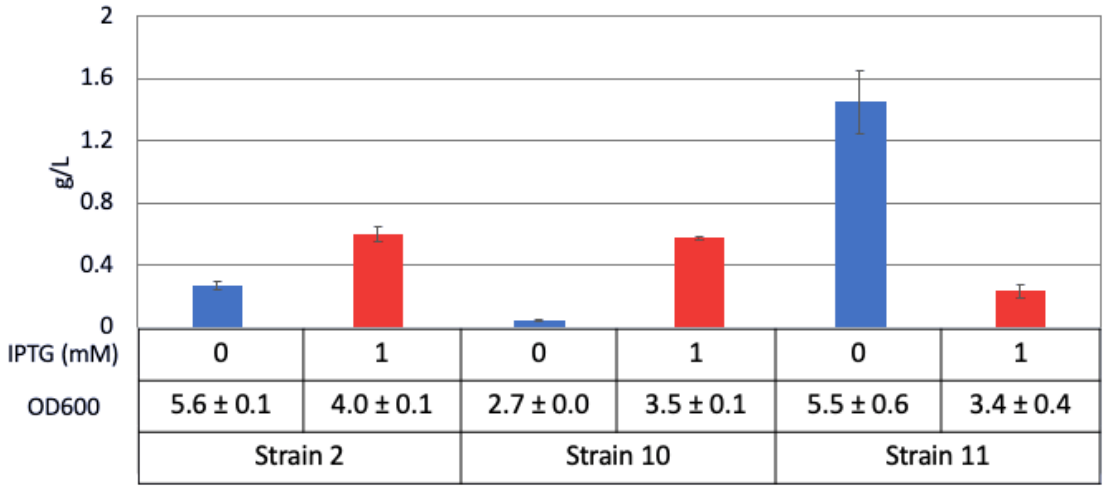


Figure 3.1.3 D-Psicose production with AlsE and YniC in *pfkA/B, zwf,* and *rpiB* triple knockout strains. Strains 2, 10, and 10 (Table S1) were grown in M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced with and without 1 mM IPTG. Product formation was determined at 24 h. Error bars indicate s.d. (n = 3 biological replicates).

3.2.3 Tuning expression level of P_{T7} promoter system for psicose pathway genes

Due to the higher production of D-psicose in Strain 11 in the absence of IPTG, we tested this strain for glucose consumption and psicose production under a gradient of IPTG concentrations (0, 1, 10, 100, and 1000 μM) to express *alsE* and *yniC*. Cultures induced with IPTG concentrations below 100 μM did not exhibit growth burden and consumed the most glucose. The best growth, glucose consumption, and D-psicose production was observed in the 10 μM IPTG condition, where Strain 11 produced 2.0 g/L (20% TMY) of D-psicose in 24 h (Figure 3.2.4A). We further screened 25 and 50 μM IPTG concentrations and observed the highest production of D-psicose in the 25 μM IPTG condition (2.6 g/L, 26% TMY) (Figure 3.2.4B).

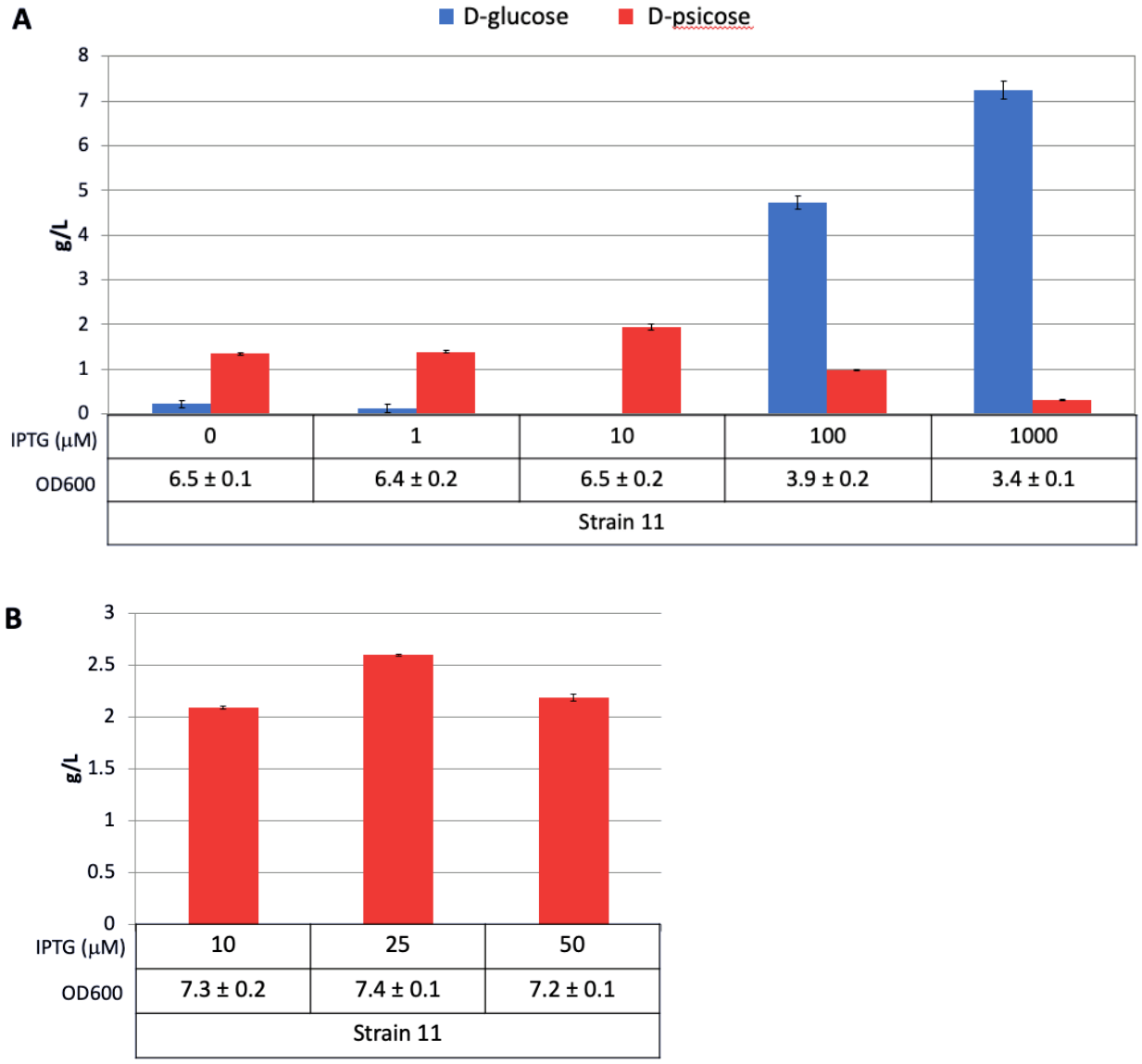


Figure 3.2.4 Effect of IPTG concentration on D-psicose production. Strain 11 was grown in M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced with various IPTG concentrations. A) 0, 1, 10, 100, and 1000 μM IPTG. Substrate remaining and product formation was determined at 24 h and are reported in g/L. Error bars indicate s.d. (n = 3 biological replicates). B) 10, 25, and 50 μM IPTG. Product formation was determined at 24 h and is reported in g/L. Error bars indicate s.d. (n = 3 biological replicates).

3.2.4 Characterizing and identifying an unknown byproduct of D-psicose production

During D-psicose production in Strain 11, we observed via a refractive index detector the accumulation of a byproduct in the culture supernatant. To determine the

source of this byproduct, we cloned each of the genes *alsE* and *yniC* separately under a P_{T7} promoter (pAL2056 and pAL2055) and introduced the respective plasmids into Strain 9 to form Strains 12 and 13 (Table S1). As a negative control, pET-16b was introduced to Strain 9 to form Strain 14 (Table S1). After culturing with 10 g/L glucose for 24 h, Strain 12 produced similar quantities of the byproduct as of Strain 11, Strain 13 produced 4-fold more of the byproduct, and Strain 14 produced 2-fold more of the byproduct (Figure 3.2.5). The basal production of the byproduct in Strain 11 was significantly decreased by overexpression of *alsE*, whereas its production was promoted by the overexpression of *yniC*, which suggests that the byproduct is formed from the substrate promiscuity of YniC towards a naturally occurring metabolite in *E. coli.*

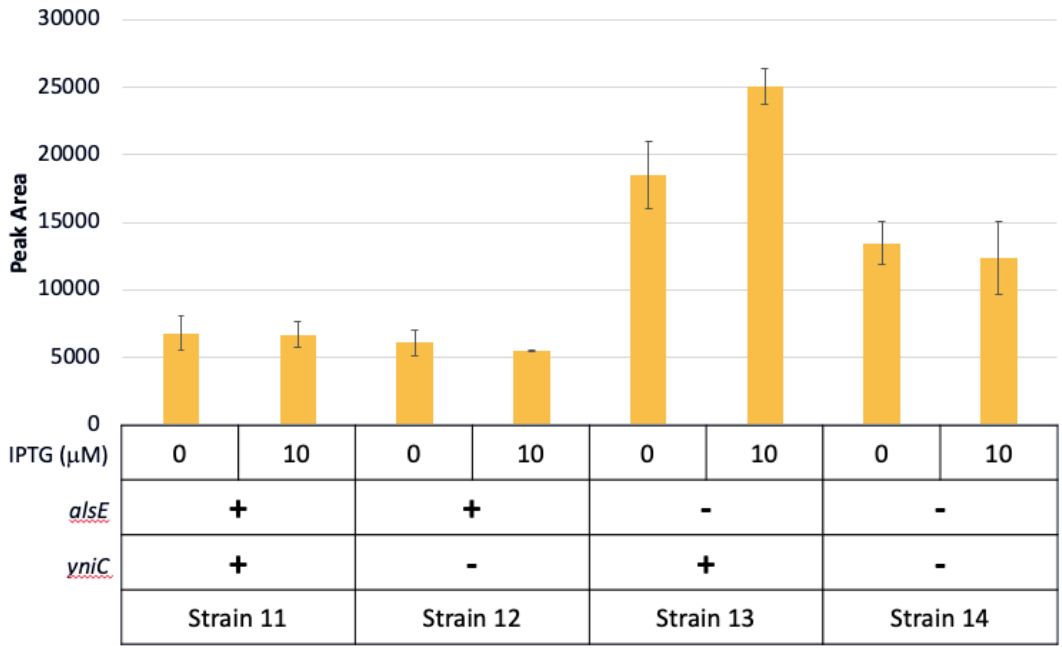


Figure 3.2.5 Effect of independent overexpression of *alsE* and *yniC* on the formation of the unknown byproduct. Strains 11, 12, 13, and 14 were grown in M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced with 0 or 10 μM IPTG. The byproduct formation was determined at 24 h and was reported in HPLC peak area. Error bars indicate s.d. (n = 3 biological replicates).

It has been shown that YniC is active towards various phosphorylated monosaccharides, but its activity towards G6P and F6P is relatively low compared to its most favored substrate 2-deoxy-glucose-6-phosphate[31,43]. The retention time of the byproduct also does not match with D-glucose or D-fructose; therefore, it is unlikely that the byproduct is the dephosphorylated species of direct intermediates of the psicose production pathway. To determine the potential phosphorylated precursor of this byproduct, we examined pathways in Strain 9 that can produce phosphorylated metabolites from G6P and F6P. Knowing that our byproduct is being retained on a column designed for separating monosaccharides and sugar alcohols, we proposed that our byproduct may be a monosaccharide produced from F6P. Of the ions remaining, the 119-ion suggested a $[M-H]^-$ adduct of C4 monosaccharide with the molecular weight of 120 g/mol and the 179-ion suggested a $[M-H]^-$ adduct of a C6 monosaccharide with the molecular weight of 180 g/mol.

With a collaboration with Elys Rodriguez (UC Davis Metabolomics Core Facility), the byproduct was determined as D-mannose, a C6 monosaccharide with the molecular weight of 180 g/mol. Mannose-6-phosphate (Man6P), which is produced from F6P by the enzyme mannose-6-phosphate isomerase ManA of the mannose degradation pathway (Figure 3.2.6), was indicated as substrate for YniC from Kuznetsova *et al.*'s substrate screening assays[31]. Although the production of D-mannose in Strains 11 and 12 are much lower than of Strains 13 and 14, F6P carbon flux has the potential to be diverted to this non-essential competing pathway. Therefore, we will explore deleting the gene *manA* from Strain 9 for future D-psicose production studies.

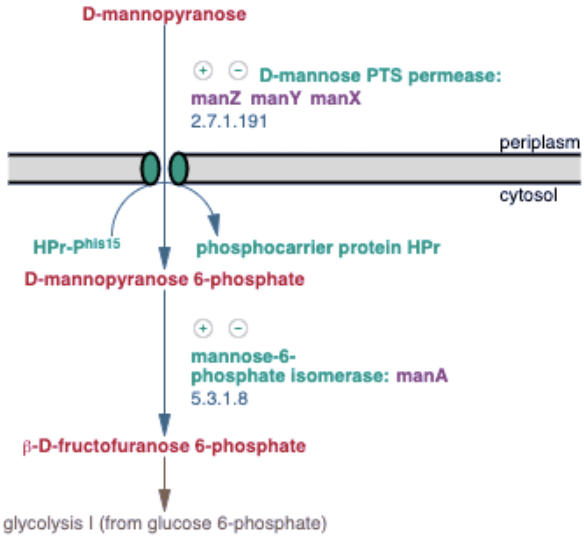
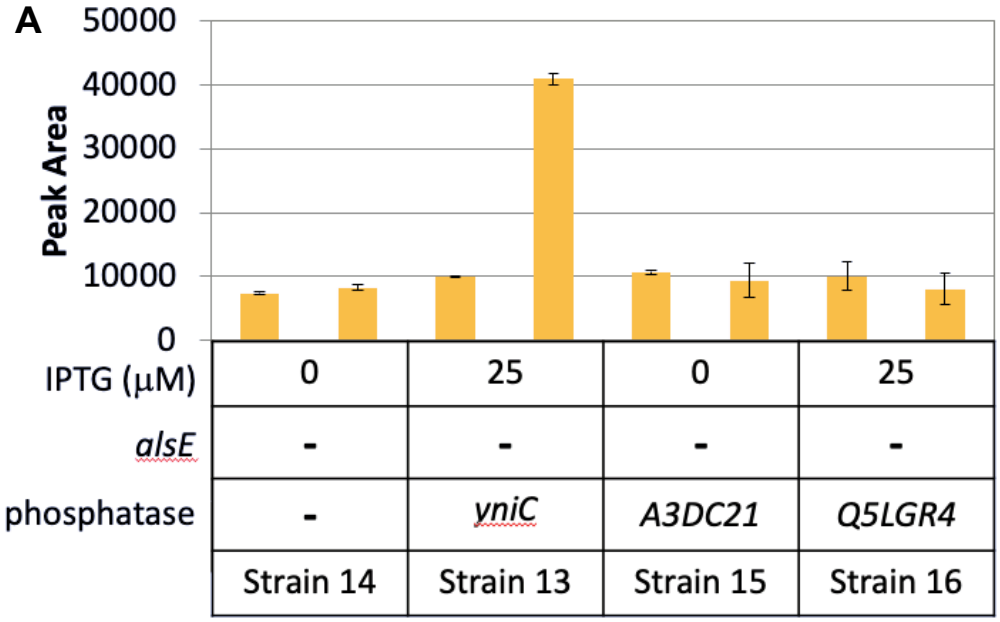


Figure 3.2.6 Mannose degradation pathway. Conventionally in the catabolic direction, D-mannose is simultaneously transported across the inner membrane and phosphorylated into mannose-6-phosphate (Man6P) by the D-mannose phosphotransferase system. In the cytoplasm, mannose-6-phosphate isomerase ManA converted Man6P into F6P, which can be assimilated into biomass production.

3.2.6 Screening phosphatases with higher P6P specificity

We have observed that in the absence of AlsE expression, the substrate promiscuity of YniC can lead to the production of other monosaccharides. Graduate students Amiruddin Bin Johan and Pam Denish in the Siegel Lab developed an *in-vitro* assay to screen phosphatases for high activity and specificity towards P6P, and they identified two phosphatases, A3DC21 of *Hungateiclostridium thermocellum* and Q5LGR4 of *Bacteroides fragilis*, that preferentially dephosphorylates P6P over G6P and F6P. We replaced *yniC* in pAL2055 with the codon-optimized genes encoding for A3DC21 and Q5LGR4 to form plasmids pAL2126 and pAL2127, respectively. Strain 9 was transformed with pAL2126 and pAL22127 to form Strains 18 and 19, respectively (Table S1). After culturing with 10 g/L D-glucose and 25 μM IPTG for 24 h, Strains 18 and 19 did not increase the production of the D-mannose relative to Strain 14 (Figure

3.2.7A). We then substituted yniC in pAL1946 with the genes for A3DC21 and Q5LGR4 to form plasmids pAL2096 and pAL2097, respectively. Plasmids pAL2096 and pAL2097 were introduced into Strain 9 to form Strains 20 and 21, respectively (Table S1). After culturing with 10 g/L D-glucose and 25 μM IPTG for 24 h, Strain 11 produced 2.5 g/L of D-psicose (24% TMY) while Strain 20 produced 2.2 g/L of D-psicose (21% TMY) and Strain 21 produced 0.94 g/L D-psicose (9% TMY) (Figure 3.2.7B). Although YniC produces D-mannose, it still has the highest activity in producing D-psicose when paired with AlsE, therefore we continued to optimize our production system with AlsE and YniC.



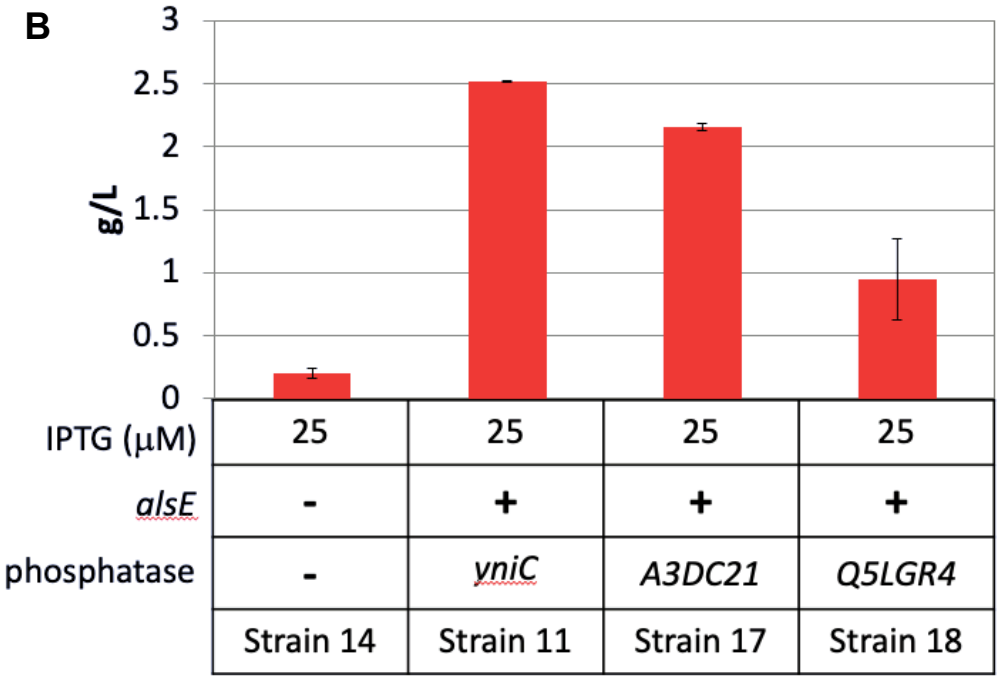


Figure 3.2.7 Screening of P6P-specific phosphatases A3DC21 and Q5LGR4. A) Effect of phosphatase specificity on the byproduct production. Strains 13, 14, 15, and 16 were grown in M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced with or without 25 μM IPTG. D-mannose production was determined at 24 h and is reported in HPLC peak area. Error bars indicate s.d. (n = 3 biological replicates). B) Effect of phosphatase specificity on D-psicose production. Strains 13, 14, 17, and 18 were grown in M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced with 25 μM IPTG. D-psicose production was determined at 24 h and is reported in g/L. Error bars indicate s.d. (n = 3 biological replicates).

3.2.7 Assessing alternative promoter systems and culturing conditions for D-psicose production

As observed with Strain 11 (Figure 3.2.4), 25 μM IPTG induction of the T7 RNA polymerase (T7 RNAP) for pAL1946 expression does not impair cell growth and D-glucose consumption, resulting the production of 2.6 g/L of D-psicose in 24 h, with a specific production of 0.35 g/L/OD. The growth burden observed in the strong expression of the bacteriophage *T7 rnap* suggests a toxic effect from allocating cellular

resources to the transcription of genes downstream of the P_{T7} promoter[44,45]. To test the expression of AlsE and YniC under a weaker promoter expression system recognized by *E. coli's* native RNA polymerase[46], we cloned *alsE* and *yniC* under the IPTG-inducible and LacI-repressible promoter P_{LlacO1}[47] to form the plasmid pAL2001. Additionally, to avoid the co-induction of T7 RNAP, we removed the gene sequence encoding for P_{lacUV5}*:T7 rnap* at *ss9* from Strain 9 to form Strain 19 (Table S1). pAL2001 was introduced into Strain 22 to form Strain 20 (Table S1). After culturing with 10 g/L D-glucose, Strain 20 produced 1.4 g/L D-psicose in 24 h (Figure 3.2.8A). Although only 4.0 g/L D-glucose was used during psicose production (Figure 3.2.8B), this P_{LlacO1} promoter expression system resulted in a 32% yield and a specific production of 0.47 g/L/OD.

Gene knockouts of key metabolic valves often hinder cellular growth rates and system productivity is limited by the formed biomass[48]. To further improve the % yield of D-glucose to D-psicose conversion, we tested the induction of pAL2001 in Strain 20 at higher cell density conditions. We cultured Strain 20 to OD_{600} ~1.0 prior to inducing with or without 1 mM IPTG. Strain 20 produced 2.0 g/L D-psicose (Figure 3.2.9A) from 6.0 g/L of consumed D-glucose (Figure 3.2.9B), resulting in a 33% yield and a specific production of 0.43 g/L/OD in 24 h. This induction condition improved glucose consumption by 2.0 g/L, but about 2/3rd of it was diverted to biomass production. We continued to increase the overall density of the cultures prior to induction by culturing 10 mL of Strain 20 to an OD_{600} ~1.0 and condensing the culture to 2.5 mL prior to inducing with 1 mM IPTG. In this high-density culture condition, Strain 20 produced 3.7 g/L D-

psicose (Figure 3.2.9C) from 10 g/L of consumed D-glucose (Figure 3.2.9D), resulting in a 37% yield and specific production of 0.39 g/L/OD in 24 h.

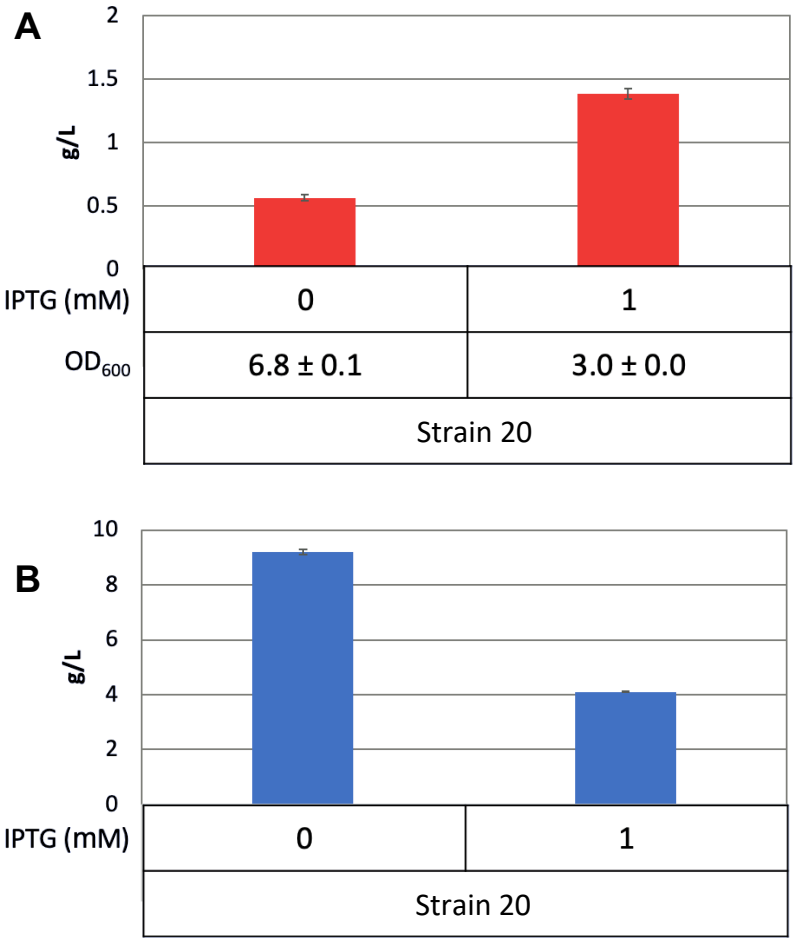


Figure 3.2.8 Effect of P_{LlacO1} promoter on D-psicose production and D-glucose consumption. Strain 20 was grown in M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced with or without 1 mM IPTG. A) D-psicose production was determined at 24 h and is reported in g/L. Error bars indicate s.d. (n = 3 biological replicates). B) D-glucose consumption was determined at 24 h and is reported in g/L. Error bars indicate s.d. (n = 3 biological replicates).

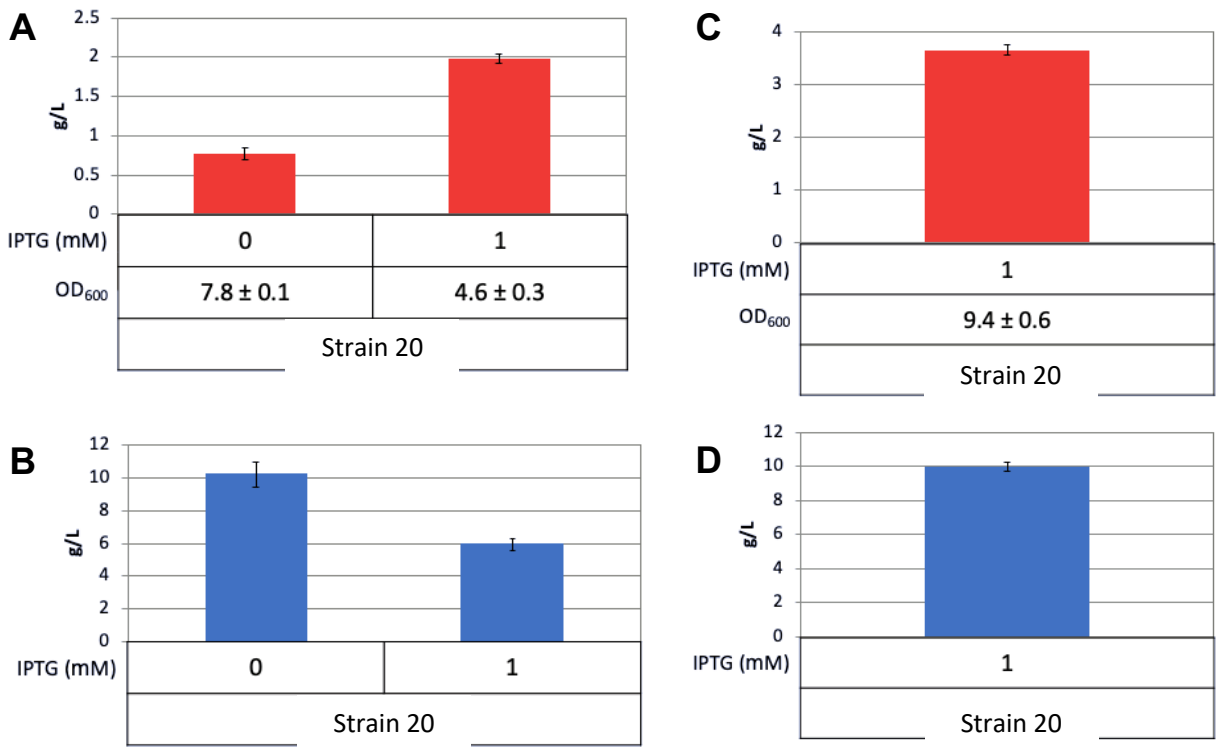


Figure 3.2.9 Effect of culture density at time of induction on D-psicose production. For OD_{600} ~1.0 induction conditions, Strain 20 was grown in 3 mL M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were induced at OD_{600} ~1.0 with or without 1 mM IPTG. A) D-psicose production was determined at 24 h and is reported in g/L. Error bars indicate s.d. (n = 3 biological replicates). B) D-glucose consumption was determined at 24 h and is reported in g/L. Error bars indicate s.d. (n = 3 biological replicates). For high-density condensed-culture induction conditions, Strain 20 was grown in 10 mL M9-minimal media supplemented with 5 g/L yeast extract and 10 g/L D-glucose. Cultures were grown to OD_{600} ~1.0, condensed to 2.5 mL, and induced at OD_{600} ~1.0 with 1 mM IPTG. C) D-psicose production was determined at 24 h and is reported in g/L. Error bars indicate s.d. (n = 3 biological replicates). D) D-glucose consumption was determined at 24 h and is reported in g/L. Error bars indicate s.d. (n = 3 biological replicates).

3.2.8 Establishing a model CRISPRi system for downregulating genes of interest

63% of D-glucose is being converted to biomass formation which significantly hinders D-psicose titers. To reduce F6P carbon flux going towards cellular growth, we are establishing CRISPR inhibition (CRISPRi)[49] to downregulate the expression of *pfkB*, the remaining phosphofructokinase of the EMPP in Strain 19. In this CRISPRi system, a mutant of Cas9 (dCas9), which has both of its endonuclease domains inactivated, is

guided to a genomic region of interest with a 20 bp single guide RNA (sgRNA) scaffold. The binding of dCas9 to the genome will then sterically hinder RNA polymerase from transcribing the gene of interest. To assess the effectiveness of sgRNA target site locations for CRISPRi, we established a fluorescent reporter system where we constructed three 20 bp sgRNAs to direct dCas9 to downregulate the expression of the green fluorescent reporter *sfgfp*[50] cloned under a P_{LlacO1} promoter. The sgRNAs targeted the coding strand of the following positions: prior to the N-terminus of *sfgfp* (pAL2066), the second operator of P_{LlacO1} (pAL2173) and the beginning of the P_{LlacO1} promoter (pAL2174). As a negative control, a sgRNA scaffold was designed without a 20 bp target sequence (pAL2063). Strain AL1050 was co-transformed with pAL1952 and respective sgRNA guides to form Strains 21, 22, 23, and 24. Strains 26, 27, and 28 had 40–60% reduction in fluorescence with expression of dCas9 relative to Strain 25 (Figure 3.2.10), indicating that sgRNAs targeting either the promoter or N-terminus of the gene of interest can be effective in recruiting dCas9 to downregulate the transcription of a gene of interest. This gene regulation strategy will be adopted for downregulating phosphofructokinase gene expression in future D-psicose production studies.

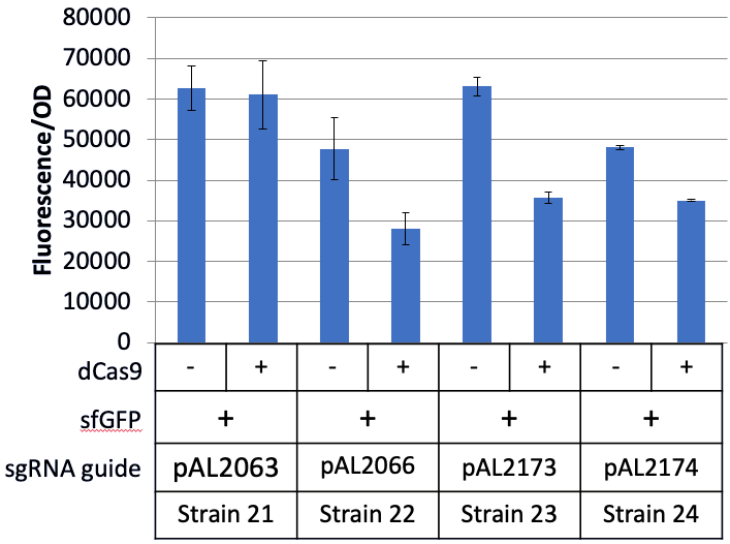


Figure 3.2.10 CRISPRi on sfGFP expression. All cultures were induced with 1 mM IPTG for sfGFP expression. Respective cultures were induced with 100 ng/mL aTc for dCas9 expression. sgRNA guides targeting the N-terminus of *sfgfp* or its P_{LlacO1} promoter recruit dCas9 to downregulate *sfgfp* transcription. Signal output is reported in specific fluorescence (510 nm emission/OD600nm) at 24 h. Error bars indicate s.d. (n = 3 biological replicates).

3.3 Discussion

D-psicose is of particular interest to the functional foods and beverage industry for its low-caloric value, taste, and health modulating effects, but its minute quantity in natural foods sources makes extraction processes too costly and inefficient to obtain D-psicose in grams-scale titers. *In-vitro* production strategies, such as isomerase-epimerase coupled systems and DHAP-aldolase phosphorylation systems, are thermodynamically limited and use unstable precursors, which hinders time-space yields in fermentation scale-up. In this study, we established a pathway in *E. coil* to produce D-psicose from D-glucose by combining native phosphorylation, isomerization, epimerization, and dephosphorylation reactions to sequester and thermodynamically drive forward our synthesis within *E. coli*'s cytoplasmic space.

While our production pathway takes advantage of *E. coli*'s native enzymes to convert D-glucose into D-psicose, D-psicose is not an abundantly synthesized product of *E. coli*'s native metabolism (Strain 1, Figure 3.2.1), suggesting that *E. coli*'s native expression levels of AlsE is insufficient to redirect F6P carbon flux. *E. coli* has evolved to actively express enzymes of central metabolism to catabolize glucose for biomass production while secondary carbon catabolic pathways, such as the D-allose degradation pathway, are not induced for expression until the secondary carbon source is present in or D-glucose is absent from the environment[30,51]. Additionally, the catalytic efficiency of central metabolic pathway reactions for F6P is higher than that of native AlsE activity[52,53], thus further suppressing D-psicose production. As seen in Strain 2, overexpression of AlsE with phosphatase YniC produces up to 0.5 g/L D-psicose in 24 h (Figure 3.2.2), supporting that enhanced C3 epimerase expression is a critical component for D-psicose production.

Improving D-psicose production required eliminating competing reactions of PPP, EMPP, and D-allose degradation pathway to redirect G6P and F6P carbon flux towards AlsE and conserving the P6P pool for YniC dephosphorylation. Serving as the first committed step of the EMPP, phosphofructokinases isozymes PfkA and PfkB irreversibly phosphorylates F6P into F16BP. It has been shown that PfkA and PfkB are responsible for about 90% and 10% of phosphofructokinase activity in *E. coli*[36], respectively, suggesting that deleting *pfkA* should significantly enhance the F6P pool for D-psicose production. By knocking out *pfkA*, Strain 7's non-induced cultures produced 0.7 g/L of D-psicose in 24 h, a 7-fold increase relative to noninduced cultures of Strain

2. This indicates that in states of high F6P accumulation, F6P can be converted to D-psicose even with minor expression levels of AlsE and YniC.

However, due to the reversible nature of Pgi, accumulated F6P in Strain 7 has the potential to be isomerized back into G6P that can be shunted towards the PPP by Zwf[25]. It has been shown that in Δ*pfkA* mutants, carbon flux ratios going through the EMPP decreases from 88% to 24% and increases through the PPP from 11% to 64%[25]. Similarly with the enhanced pool of P6P generated by AlsE, the reversible activity of RpiB can also isomerize P6P into aldehyde-D-allose-6-phosphate, thus reducing the productivity of psicose production by YniC. Thus, it was essential to incorporate the *zwf* and *rpiB* knockouts on top of the *pfkA* knockout to retain F6P and P6P intermediates to improve D-psicose production. The combination of all three gene deletions allowed Strain 11 to produce 1.4 g/L of D-psicose, a 2-fold increase relative to Strain 7. Furthermore, by combining fine-tuning the expression levels of AlsE and YniC and increasing the cell density of cultures prior to induction, D-psicose production increased to 3.7 g/L from 10 g/L D-glucose in 24 h (Figure 3.2.10C).

Phosphatase activity and specificity also played a major role in D-psicose production by controlling the level of phosphorylated intermediates derived from the enhanced F6P pool. Most phosphatases within the HAD family have broad substrate activity to small C1 and C6 phosphorylated saccharides and cofactor compounds, which inherently makes it difficult to achieve dephosphorylation of a single metabolite[31]. This substrate flexibility of HADs may be a mechanism to broadly regulate carbon flux and intracellular phosphorylated intermediates during carbohydrate metabolism[31]. In the case of YniC, although it can dephosphorylate G6P and F6P into D-glucose and D-

fructose, respectively, its activity towards these two metabolites are lower than of Man6P and E4P. However, its high activity towards Man6P, a product of the overflow of F6P through ManA into the mannose degradation pathway, generates D-mannose as a major side product when AlsE expression is insufficient to partition F6P towards D-psicose production (Figure 3.2.5). Removing *manA* may eliminate D-mannose production by YniC, but it is also possible that F6P flux will be redirected towards synthesizing other peripheral metabolites. Glucosamine-6-phosphate of UDP-*N*-acetyl-glucosamine biosynthesis, sorbitol-6-phosphate of the D-sorbitol degradation pathway, and mannitol-1-phosphate of the D-mannitol degradation pathway are a few intermediates that directly stem from F6P. These phosphorylated metabolites may be potential substrates of YniC or other natively encoded HADs in forming subsequent sugar side products during D-psicose production and would require additional gene deletions or dynamic transcriptional downregulation of these peripheral pathways by CRISPRi to prevent their production.

The production of D-psicose from the central metabolite fructose-6-phosphate also highlights the inherent challenge of balancing product formation with generating enough PEP from the lower half of the EMPP required for glucose uptake via the glucose PTS. For every mole of D-glucose that is transported across the PTS, one mole of PEP is required for the cascade phosphorylation that converts D-glucose into G6P. However, restriction of F6P carbon flux through PfkA and G6P carbon flux through Zwf limits PEP availability which can reduce the glucose uptake rate and affect D-psicose productivity. Therefore, it would be of interest to establish a PTS^- strain with overexpression of *galP*, which encodes for a D-galactose/H^+ symporter, and *glk*, which

encodes for a glucose kinase, that allows for glucose uptake and phosphorylation that is independent of the metabolite availability of lower EMPP glycolysis[54,55].

3.4 Methods

3.4.1 Reagents

All enzymes involved in the molecular cloning experiments were purchased from New England Biolabs (NEB). All synthetic oligonucleotides were synthesized by Integrated DNA Technologies. Sanger sequencing was provided by Genewiz. D-glucose, D-psicose, 75% D-erythrose, and D-mannose was purchased from Sigma-Aldrich.

3.4.2 Strains and plasmids

All strains used in this study are listed in Tables S1 and S2. All plasmids and primers are listed on Tables S3 and S4. Gene deletions and integrations were constructed using CRISPR-Cas9-mediated homologous recombination [56]. Linear DNA repair fragments for gene deletions were constructed by PCR assembly or amplification from genomic DNA using primers listed in Tables S4 and S5. All genomic modifications were PCR and sequence verified.

Plasmids for D-psicose production and CRISPRi on P_{LlacO1}*:sfgfp* were constructed using sequence and ligation independent cloning (SLIC) [57]. Plasmids encoding sgRNAs for CRISPR-Cas9-mediated homologous recombination were constructed with Q5 site-directed mutagenesis using a modified template pTargetF (Addgene plasmid # 62226). Templates used for DNA amplification and cloning are listed in Table S6. All plasmids were verified by PCR and Sanger sequencing.

3.4.3 Culture conditions

Overnight cultures were grown at 37 °C, 250 rpm, in 3 mL of Luria-Bertani (LB) media with appropriate antibiotics. Antibiotic concentrations were as follows: spectinomycin (50 µg/mL), ampicillin (200 µg/mL), and kanamycin (50 µg/mL). Growth assays were carried out in M9 minimal medium (33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 9.4 mM NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$) including 1000 × A5 trace metal mix (2.86 g H_3BO_3, 1.81 g $MnCl_2{\cdot}4H_2O$, 0.079 g $CuSO_4{\cdot}5H_2O$, 49.4 mg $Co(NO_3)_2{\cdot}6H_2O$ per liter water). Psicose production was carried out in M9 minimal medium supplemented with 5 g/L yeast extract (M9P). Optical densities were measured at 600 nm (OD_{600}) with a Synergy H1 hybrid plate reader (BioTek Instruments, Inc.).

3.4.4 D-Psicose production

Overnight cultures were inoculated at 1% in 3 mL of M9P supplemented with 10 g/L glucose. Cultures were grown at 37 °C, 250 rpm, until OD_{600} reached 0.4–0.6. Appropriate concentrations of IPTG were added and the cultures were grown at 30 °C, 250 rpm, for 24 h. The produced D-psicose was confirmed by high resolution electrospray ionization mass spectrometry using a Thermo Electron LTQ-Orbitrap Hybrid MS at the Mass Spectrometry Facility in the University of California, Davis.

3.4.5 HPLC Analysis

To measure D-glucose and D-psicose, cell culture supernatant was analyzed using HPLC (Shimadzu) equipped with a refractive index detector (RID) 10 A and a Hi-Plex Ca^{2+} column (Agilent). The mobile phase consisted of 100% HPLC-grade water and was run at a flow rate of 0.6 mL/min for 13 min, with the column oven at 83 °C and RID cell temperature at 40 °C.

3.4.6 Fluorescence Assays

Overnight cultures were inoculated at 1% in 3 mL of M9P media and grown at 37 °C, 250 rpm, until OD_{600} reached 0.4–0.6. Cultures were respectively induced with IPTG (1.0 mM) and grown at 37 °C, 250 rpm, for 24 h. Fluorescence emission was measured at 510 nm with a Synergy H1 hybrid plate reader (BioTek Instruments, Inc.) at 4 h and 24 h.

3.5 Supplementary Information

Table S1. Strain list

Strain no.	*E. coli* strain	Plasmid	Key Genotype
1	AL3601	pET-16b	*ss9*::P_{lacUV5}:*T7rnap, P_{T7}:EV*
2	AL3601	pAL1946	*ss9*::P_{lacUV5}:*T7rnap, P_{T7}:alsE-yniC*
3	AL3601	pAL1947	*ss9*::P_{lacUV5}:*T7rnap, P_{T7}:alsE-ybiV*
4	AL3689	-	*ss9*::P_{lacUV5}:*T7rnap ΔpfkB*
5	AL3694	-	*ss9*::P_{lacUV5}:*T7rnap ΔpfkA*
6	AL3689	pAL1946	*ss9*::P_{lacUV5}:*T7rnap ΔpfkB, P_{T7}:alsE-yniC*
7	AL3694	pAL1946	*ss9*::P_{lacUV5}:*T7rnap ΔpfkA, P_{T7}:alsE-yniC*
8	AL3730	-	*ss9*::P_{lacUV5}:*T7rnap ΔpfkB Δzwf ΔrpiB*
9	AL3729	-	*ss9*::P_{lacUV5}:*T7rnap ΔpfkA Δzwf ΔrpiB*
10	AL3730	pAL1946	*ss9*::P_{lacUV5}:*T7rnap ΔpfkB Δzwf ΔrpiB, P_{T7}:alsE-yniC*
11	AL3729	pAL1946	*ss9*::P_{lacUV5}:*T7rnap ΔpfkA Δzwf ΔrpiB, P_{T7}:alsE-yniC*
12	AL3729	pAL2056	*ss9*::P_{lacUV5}:*T7rnap ΔpfkA Δzwf ΔrpiB, P_{T7}:alsE*
13	AL3729	pAL2055	*ss9*::P_{lacUV5}:*T7rnap ΔpfkA Δzwf ΔrpiB, P_{T7}:yniC*
14	AL3729	pET-16b	*ss9*::P_{lacUV5}:*T7rnap ΔpfkA Δzwf ΔrpiB, P_{T7}:EV*
15	AL3729	pAL2126	*ss9*::P_{lacUV5}:*T7rnap ΔpfkB Δzwf ΔrpiB, P_{T7}:A3DC21*
16	AL3729	pAL2127	*ss9*::P_{lacUV5}:*T7rnap ΔpfkB Δzwf ΔrpiB, P_{T7}:Q5LGR4*
17	AL3729	pAL2096	*ss9*::P_{lacUV5}:*T7rnap ΔpfkB Δzwf ΔrpiB, P_{T7}:alsE-A3DC21*
18	AL3729	pAL2097	*ss9*::P_{lacUV5}:*T7rnap ΔpfkB Δzwf ΔrpiB, P_{T7}:alsE-Q5LGR4*
19	AL3756	**-**	*ΔpfkA Δzwf ΔrpiB ΔP*$_{lacUV5}$:*T7rnap*
20	AL3756	pAL1946	*ΔpfkA Δzwf ΔrpiB ΔP*$_{lacUV5}$:*T7rnap,*

			P_{LlacO1}:alsE-yniC
21	AL1050	pAL1952 + pAL2063	*P_{LtetO1}:dCas9 p15A kan^r sgRNA-empty guide, P_{LlacO1}:sfgfp colE1 amp^r*
22	AL1050	pdCas9 + pAL2066	*P_{LtetO1}:dCas9 p15A kan^r sgRNA (ttaaagaggagaaaggtacc), P_{LlacO1}:sfgfp colE1 amp^r*
23	AL1050	pdCas9 + pAL2173	*P_{LtetO1}:dCas9 p15A kan^r sgRNA (aagatactgagcacatcagc), P_{LlacO1}:sfgfp colE1 amp^r*
24	AL1050	pdCas9 + pAL2174	*P_{LtetO1}:dCas9 p15A kan^r sgRNA (tagatctattaaattgtgag), P_{LlacO1}:sfgfp colE1 amp^r*

Table S2. Strains used in this study

Strain	Genotype	Source
XL-1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [*F´ proAB lacIq ZΔM15 Tn10* (*tet^r*)]	Agilent (Santa Clara, CA)
AL1050	*F- lambda- ilvG- rfb-50 rph-1 attB::lacI^q tetR spec^r*	Rodriguez et al. 2015[58]
AL3601	*F- lambda- ilvG- rfb-50 rph-1 attB::lacI^q tetR spec^r ss9::P_{lacUV5}:T7rnap*	Zhang et al. 2021[59]
AL3689	As AL3601, but Δ*pfkB*	This study
AL3694	As AL3601, but Δ*pfkA*	This study
AL3725	As AL3601, but Δ*zwf*	This study
AL3727	As AL3694, but Δ*zwf*	This study
AL3728	As AL3689, but Δ*zwf*	This study
AL3729	As AL3727, but Δ*rpiB*	This study
AL3730	As AL3728, but Δ*rpiB*	This study
AL3756	As AL3729, but ΔP_{lacUV5}:*T7rnap*	This study

Table S3. Plasmids used in this study

Plasmid	Genotype
pCas	*P_{cas}:cas9 P_{araB}:Red lacI^q P_{trc}:sgRNA pMB1 repA101(Ts) kan^r*
pET-16b	*P_{T7}:10xHis pBR322 amp^r*
pdCas9	*P_{LtetO1}:dCas9 p15A cm^r*
pAL421	*P_{LlacO1}:sfgfp ColE1 amp^r*
pAL1023	*P_{LtetO1} ColA kan^r*
pAL1354	*P_{LlacO1} ColE1 amp^r*
pAL1851	*sgRNA-lacZ pMB1 amp^r*
pAL1946	*P_{T7}:alsE-yniC pBR322 amp^r*
pAL1947	*P_{T7}:alsE-ybiV pBR322 amp^r*
pAL1950	*sgRNA-pfkA pMB1 amp^r*
pAL1951	*sgRNA-pfkB pMB1 amp^r*
pAL1952	*P_{LtetO1}:dCas9 p15A kan^r*
pAL1957	*sgRNA-rpiB pMB1 amp^r*

pAL1958	*sgRNA-zwf pMB1 amp^r*
pAL2001	*P_{LlacO1}:alsE-yniC ColE1 amp^r*
pAL2045	*sgRNA-pgm pMB1 amp^r*
pAL2055	*P_{T7}:alsE pBR322 amp^r*
pAL2056	*P_{T7}:yniC pBR322 amp^r*
pAL2062	*sgRNA-pfkA, P_{LlacO1}:sfgfp pMB1 amp^r*
pAL2063	*sgRNA-empty, P_{LlacO1}:sfgfp pMB1 amp^r*
pAL2064	*sgRNA-atactgagcacatcagcagg, P_{LlacO1}:sfgfp pMB1 amp^r*
pAL2065	*sgRNA- atcatgacattaacctataa, P_{LlacO1}:sfgfp pMB1 amp^r*
pAL2066	*sgRNA- ttaaagaggagaaaggtacc, P_{LlacO1}:sfgfp pMB1 amp^r*
pAL2096	*P_{T7}:alsE-A3DC21 pBR322 amp^r*
pAL2097	*P_{T7}:alsE-Q5LGR4 pBR322 amp^r*
pAL2103	*P_{T7}:A3DC21 pMB1 kan^r*
pAL2104	*P_{T7}:Q5LGR4 pMB1 kan^r*
pAL2126	*P_{T7}:A3DC21 pBR322 amp^r*
pAL2127	*P_{T7}:Q5LGR4 pBR322 amp^r*
pAL2063	*sgRNA-empty guide, P_{LlacO1}:sfgfp colE1 amp^r*
pAL2066	*sgRNA (ttaaagaggagaaaggtacc), P_{LlacO1}:sfgfp colE1 amp^r*
pAL2173	*sgRNA (aagatactgagcacatcagc), P_{LlacO1}:sfgfp colE1 amp^r*
pAL2174	*sgRNA (tagatctattaaattgtgag), P_{LlacO1}:sfgfp colE1 amp^r*

Table S4. Oligonucleotides used in this study

Name	Sequence 5' → 3'
AZ56	CTAGGTCTAGGGCGGCGGATTTG
AZ195	GCACATCAGCGTTTTAGAGCTAGAAATAGC
AZ196	TCAGTATCTTACTAGTATTATACCTAGGAC
AZ403	ACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG
AZ466	GCCATATCGAAGGTCGTCATATGAAAATCTCCCCCTCGTTAATG
AZ467	ATCTCCTTTTGTTAGCAGCCGGATCCTTATGCTGTTTTTGCATGAG GCTG
AZ468	GGCTGCTAACAAAAGGAGATATACATATGTCAACCCCGCGTCAGA TTCTTGC
AZ469	GCTTTGTTAGCAGCCGGATCCTCAACCGAGAAGGTCTTTTGCGGT G
AZ470	GGATCCGGCTGCTAACAAAGCC
AZ471	ATGACGACCTTCGATATGGCCG
AZ472	GGCTGCTAACAAAAGGAGATATACATATGAGCGTAAAAGTTATCG TCAC
AZ473	GCTTTGTTAGCAGCCGGATCCTCAGCTGTTAAAAGGGGATGTG
AZ477	GCGTTCCGCCTGATTGATGAAATCC
AZ478	CATTTCACTGGTTGAAGAGACACGTCC
AZ482	CTGACCTGAATCAATTCAGCAGGAAGTGATTGTTATACTATTTGCA CATTCGTTGGAT
AZ483	TCTGTTGCCGGAAGTCTTCTTGCACATCGAAGTGATCCAACGAAT GTGCAAATAGTAT

AZ484	AGACTTCCGGCAACAGATTTCATTTTGCATTCCAAAGTTCAGAGG TAGTCTGATTTCG
AZ485	TGTCATCGGTTTCAGGGTAAAGGAATCTGCCTTTTTCCGAAATCA GACTACCTCTGAAC
AZ486	ACCCTGAAACCGATGACAGAAGCAAAAATGCCTGATGCGCTTCG CTTATCAGGCCTACAT
AZ487	CCTACAAAAGTTTGCAAATTCAATAAATTGCAGAATTCATGTAGGC CTGATAAGCGA
AZ490	TAGCGCTGGCAGGATCATCCATGAC
AZ491	CTGTTGCTATTCCATTCCTCCCAGGTCG
AZ492	ACAGATTTTTATTTATATATATTTATCTGCAAAATTTTAAATAAAGCT CCAA
AZ493	AGCGGAAAGTGAAGAAATTAACAATATGATTTATTGGAGCTTTATT TAAAATTTTGCAGA
AZ494	TGTTAATTTCTTCACTTTCCGCTGATTCGGTGCCAGACTGAAATCA GCCT
AZ495	CCCAATGCTGGGGGAATGTTTTTGCATTTCCTCCTATAGGCTGAT TTCAGTCTGGCAC
AZ496	TCCCCCAGCATTGGGGGAATCATCACCAACCTGTCGGC
AZ497	TGACTTTTGAGCATAGTCGGAGAAACGCGTTGCCGACAGGTTGG TGATG
AZ498	CCGACTATGCTCAAAAGTCATGTGATAACAAAGGGGTGAACTATG GCCAGTGGCGAT
AZ499	AACGGACAAGATCGCCACTGGCCAT
AZ500	GGGGCCATCCGTTTTAGAGCTAGAAATAGC
AZ501	GGATTCAATTACTAGTATTATACCTAGGAC
AZ502	CAAGCGAGCTCGATATCAAATCAGAAGAACTCGTCAAGAAGGC
AZ503	CAGGAGCTAAGGAAGCTAAAATGATTGAACAAGATGGATTGCACG CAGG
AZ504	TTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTC
AZ505	TTTGATATCGAGCTCGCTTGGACTCC
AZ520	TCGGTCTGCCGTTTTAGAGCTAGAAATAGC
AZ521	TGCACGGGAAACTAGTATTATACCTAGGAC
AZ524	TTTATTGTGAACAATGGCGAGTGGC
AZ525	GGCGGTCGTCATATTGTCTGCG
AZ526	CGCTGTCAGATGTAACTCTGTAAAACAGATCAGGAAGGCGTA
AZ527	TGTGATGTTAATGAATTAAAAACACCCAAAATCGATGAATTACGCC TTCCTGATCTGTT
AZ528	GGGTGTTTTTAATTCATTAACATCACAAATGTTTTTTGATTGTGAA GTTTTGCACGGACG
AZ529	CTCATCCATGCAAGTAGTGGATGAATCTCATCTTCCCCGTCCGTG CAAAACTTCACAAT
AZ530	CCACTACTTGCATGGATGAGTAATGATTAATGTGGATAGAGTTTCT TTTTGAGGTTGGCT
AZ531	AGCGGAAAGCGTTTCATTAGCCAACCTCAAAAAGAAACTCTA

AZ532	ATGAAACGCTTTCCGCTATTTCTTTTATTCACCCTGCTCACGCTGTCCACCGTTC
AZ533	CGGAACGGTGGACAGCGTGAGC
AZ536	TTACGCCTGTGTGCCGTGTTAATGAC
AZ537	CGCGTAACAATTGTGGATTCATAAAGGCTCCTG
AZ538	TAATCGCACGGGTGGATAAGCGTTTACAGTTTTCGCAAGCTCGTAAAAGCAGTACAGTGC
AZ539	CGGTACTTAAGCCAGGGTATACTTGTAATTTTCTTACGGTGCACTGTACTGCTTTTACGA
AZ540	ACCCTGGCTTAAGTACCGGGTTAGTTAACTTAAGGAGAATGACTATCTGCGCTTATCCT
AZ541	GCGCAAGATCATGTTACCGGTAAAATAACCATAAAGGATAAGCGCAGATAGTCATT
AZ542	CGGTAACATGATCTTGCGCAGATTGTAGAACAATTTTTACACTTTCAGGCCTCGTGCGGA
AZ543	CAGTCAGTGTAATAAAAAAAGCCTCGTGGGTGAATCCGCACGAGGCCTGAA
AZ666	AAGAGGAGAAAAGATATACCATGAAAATCTCCCCCTCGTTAATGTG
AZ667	GGTACCTTAGCAGCCGGATCTCAACCGAGAAGGTCTTTT
AZ668	GATCCGGCTGCTAAGGTACCTAATCTAGAGGCATC
AZ669	GGTATATCTTTTCTCCTCTTTAATGAATTCGGTCAGTGCG
AZ670	AAAGTCCTAGGTTTTAGAGCTAGAAATAGC
AZ671	CTTGTTGATTACTAGTATTATACCTAGGAC
AZ823	ATGTCAACCCCGCGTCAGATTCTTG
AZ824	ATGACGACCTTCGATATGGCCGCTG
AZ887	ACGCGTGCTAGAGGCATC
AZ888	GGTACCTTTCTCCTCTTTAATGAATTCGG
AZ924	TAACAAAAGGAGATATACATATGATCAAGTACAAGGCGGTGTTC
AZ925	ACTCAGCTTCCTTTCGGGCTTCACAGCATAAACATATCCAAGAGGCC
AZ926	TAACAAAAGGAGATATACATATGAAGTATACTGTCTACCTGTTTG
AZ927	TGTTAGCAGCCGGATCCTCACAAAGGGCAGCCGCTCTTATCTTC
AZ963	TTAAAGAGGAGAAAGGTACCATGTCAACCCCGCGTCAGATTCTTG
AZ964	TTGATGCCTCTAGCACGCGTTCAACCGAGAAGGTCTTTTGCGGTG
AZ1025	AAATTGTGAGGTTTTAGAGCTAGAAATAGC
AZ1026	AATAGATCTAACTAGTATTATACCTAGGAC
JG184	GTTTTAGAGCTAGAAATAGC
JG185	ACTAGTATTATACCTAGGAC
JG190	GAAAGGTACCGTTTTAGAGCTAGAAATAGC
JG191	TCCTCTTTAAACTAGTATTATACCTAGGAC
JG205	GTCGACCTGCAGAAGCTTAGATCTATTAAATTGTGAGCGGATAACAATTG
JG208	TAATAGATCTAAGCTTCTGCAGGTCGACTCTAGAGAATTC
JG268	CTCGAGTAGGGATAACAGGGTTATTTGTACAGTTCGTCCA

JG267	CCCTGTTATCCCTACTCGAGTTCATGTGCA
MC124	GCTAGTTATTGCTCAGCGGTGGC
MMM40	GAGTCAGTGAGCGAGGAAGC
SD62	GGCCCTTTCGTCTTCACCTCGAG
YT092	CTACTCAGGAGAGCGTTCAC
YT430	CCAGTAGTAGGTTGAGGCCGTTGAG
YT680	CTTCAGACTTCCGAGTCATCCATGC
YT682	CGAGAGCGTATGAAACGAATCGAAG
YT683	GTAAACTATCGCCTTGTCCAGACAC
YT684	ATGCGTATCTAAATGCCGTCGTTGG
YT685	AAGGAAATGAGCTGGCTCTGCCAAG
YT695	GTTCACTCAAACCTCCACGTTCAGC
YT724	GGGTCAAACCCAATGACAAAGCAATG

Table S5. Guide for CRISPR-Cas9-mediate gene deletions and insertions

	pTargetF		PCR Linear Repair Fragment	
Modification	Plasmid	20 bp sgRNA sequence 5'→ 3'	Primers	Template
Δ*pfkA*	pAL1950	TTCCCGTGCATCGGTCTGCC	AZ482-487	-
Δ*pfkB*	pAL1951	AATTGAATCCGGGGCCATCC	AZ492-499	-
Δ*rpiB*	pAL1957	GTCGCACTGGCTGTTGCTGG	AZ526-533	-
Δ*zwf*	pAL1958	GCAGAAGAAGTGGGGATCGA	AZ538-543	-
ΔP_{lacUV5}-*T7rnap*	pAL2002	AATCAACAAGAAAGTCCTAG	MM131, MM132	MG1655

Table S6. Plasmid construction guide

	PCR for Vector			PCR for Insert(s)			
Plasmid	Primer (F)	Primer (R)	Template	Primer (F)	Primer (R)	Template	Sequence of Interest
pAL1946	AZ470	AZ471	pET-16b	AZ466	AZ467	MG1655	*alsE*
				AZ468	AZ469	MG1655	*yniC*
pAL1947	AZ470	AZ471	pET-16b	AZ466	AZ467	MG1655	*alsE*
				AZ472	AZ473	MG1655	*ybiV*
pAL1950*	AZ520	AZ521	pAL1851				
pAL1951*	AZ500	AZ501	pAL1851				
pAL1952	AZ504	AZ505	pdCas9	AZ502	AZ503	pAL1023	*kanR*
pAL1957*	AZ522	AZ523	pAL1851				
pAL1958*	AZ534	AZ535	pAL1851				
pAL2001	AZ668	AZ669	pAL1354	AZ666	AZ667	pAL1946	*alsE-yniC*
pAL2002*	AZ670	AZ671	pAL1851				
pAL2045*	AZ796	AZ797	pAL1851				

pAL2055*	AZ823	AZ824	pAL1946				
pAL2056*	AZ470	AZ467	pAL1946				
pAL2062	JG267	JG208	pAL1950	JG205	JG268	pAL421	P_{LlacO1}:*sfgfp*
pAL2063*	JG184	JG185	pAL2062				
pAL2066*	JG190	JG191	pAL2062				
pAL2096	AZ470	AZ471	pAL1946	AZ466	AZ467	MG1655	*alsE*
				AZ924	AZ925	pAL2103	*A3DC21*
pAL2097	AZ470	AZ471	pAL1946	AZ466	AZ467	MG1655	*alsE*
				AZ926	AZ927	pAL2104	*Q5LGR4*
pAL2126*	AZ897	AZ824	pAL2096				
pAL2127*	AZ897	AZ824	pAL2097				
pAL2173*	AZ195	AZ196	pAL2062				
pAL2174*	AZ1025	AZ1026	pAL2062				

*Q5-site directed mutagenesis (NEB).

3.6 References

1. Ahuja, Kunal; Mamtani, K. *Rare Sugar Market Size By Product (D-Mannose, Allulose, Tagatose, D-Xylose, L-Arabinose, L-Fucose), By Application (Dietary Supplements, Cosmetics & Personal Care, Pharmaceuticals, Food & Beverages), Industry Analysis Report, Regional Outlook, Application*. https://www.gminsights.com/industry-analysis/rare-sugar-market (2020).
2. *Rare Sugar Market (By Product: D-Mannose, Allulose, Tagatose, D-Xylose, L-Arabinose, L-Fucose, D-Psicose; By Application: Dietary Supplements, Cosmetics, Personal Care, Pharmaceuticals, Food and Beverages, Sweetner, Others) - Global Industry Analysis, Mar*. https://www.acumenresearchandconsulting.com/rare-sugar-market (2020).
3. Bilal, M., Iqbal, H. M. N., Hu, H., Wang, W. & Zhang, X. Metabolic engineering pathways for rare sugars biosynthesis, physiological functionalities, and applications-a review. *Crit. Rev. Food Sci. Nutr.* **58**, 2768–2778 (2018).
4. Jiang, S. *et al.* Review on D-Allulose: In vivo Metabolism, Catalytic Mechanism, Engineering Strain Construction, Bio-Production Technology . *Frontiers in Bioengineering and Biotechnology* vol. 8 26 (2020).
5. J, J., P, G., Guruchandran, S. & Chakravarthy, M. Biocatalytic Production of D-Tagatose: A Potential Rare Sugar with Versatile Applications. *Crit. Rev. Food Sci. Nutr.* **57**, 0 (2016).
6. Zhang, W., Yu, S., Zhang, T., Jiang, B. & Mu, W. Recent advances in d-allulose: Physiological functionalities, applications, and biological production. *Trends Food Sci. Technol.* **54**, 127–137 (2016).
7. Mu, W., Zhang, W., Feng, Y., Jiang, B. & Zhou, L. Recent advances on applications and biotechnological production of d-psicose. *Appl. Microbiol. Biotechnol.* **94**, 1461–1467 (2012).
8. Hossain, A. *et al.* Rare sugar d-allulose: Potential role and therapeutic monitoring in maintaining obesity and type 2 diabetes mellitus. *Pharmacol. Ther.* **155**, 49–59 (2015).
9. Matsuo, T. & Izumori, K. Retraction: $_{D}$-Psicose Inhibits Intestinal α-

Glucosidase and Suppresses the Glycemic Response after Ingestion of Carbohydrates in Rats. *J. Clin. Biochem. Nutr.* **45**, 202–206 (2009).
10. IIDA, T. *et al.* Acute <small>D</small>-Psicose Administration Decreases the Glycemic Responses to an Oral Maltodextrin Tolerance Test in Normal Adults. *J. Nutr. Sci. Vitaminol. (Tokyo).* **54**, 511–514 (2008).
11. Eggleston, G. & Lima, I. Sustainability Issues and Opportunities in the Sugar and Sugar-Bioproduct Industries. *Sustainability* vol. 7 (2015).
12. Men, Y. *et al.* Co-expression of d-glucose isomerase and d-psicose 3-epimerase: Development of an efficient one-step production of d-psicose. *Enzyme Microb. Technol.* **64**–**65**, 1–5 (2014).
13. Chen, X. *et al.* Production of d-psicose from d-glucose by co-expression of d-psicose 3-epimerase and xylose isomerase. *Enzyme Microb. Technol.* **105**, 18–23 (2017).
14. Li, Z., Cai, L., Qi, Q. & Wang, P. G. Enzymatic synthesis of D-sorbose and D-psicose with aldolase RhaD: effect of acceptor configuration on enzyme stereoselectivity. *Bioorg. Med. Chem. Lett.* **21**, 7081–7084 (2011).
15. Li, Z. *et al.* Synthesis of rare sugars with L-fuculose-1-phosphate aldolase (FucA) from Thermus thermophilus HB8. *Bioorganic & Med. Chem. Lett.* **21**, 5084–5087 (2011).
16. Choi, J.-G., Ju, Y.-H., Yeom, S.-J. & Oh, D.-K. Improvement in the thermostability of D-psicose 3-epimerase from Agrobacterium tumefaciens by random and site-directed mutagenesis. *Appl. Environ. Microbiol.* **77**, 7316–7320 (2011).
17. Dedania, S. R., Patel, M. J., Patel, D. M., Akhani, R. C. & Patel, D. H. Immobilization on graphene oxide improves the thermal stability and bioconversion efficiency of D-psicose 3-epimerase for rare sugar production. *Enzyme Microb. Technol.* **107**, 49–56 (2017).
18. Patel, S. N. *et al.* Improved operational stability of d-psicose 3-epimerase by a novel protein engineering strategy, and d-psicose production from fruit and vegetable residues. *Bioresour. Technol.* **216**, 121–127 (2016).
19. Zhang, W. *et al.* Characterization of a Novel Metal-Dependent D-Psicose 3-Epimerase from Clostridium scindens 35704. *PLoS One* **8**, e62987 (2013).
20. Li, A. *et al.* Recent advances in the synthesis of rare sugars using DHAP-dependent aldolases. *Carbohydr. Res.* **452**, 108–115 (2017).
21. Schümperli, M., Pellaux, R. & Panke, S. Chemical and enzymatic routes to dihydroxyacetone phosphate. *Appl. Microbiol. Biotechnol.* **75**, 33–45 (2007).
22. Tchieu, J. H., Norris, V., Edwards, J. S. & Saier, M. H. J. The complete phosphotransferase system in Escherichia coli. *J. Mol. Microbiol. Biotechnol.* **3**, 329–346 (2001).
23. Jeckelmann, J.-M. *et al.* Structure and function of the glucose PTS transporter from Escherichia coli. *J. Struct. Biol.* **176**, 395–403 (2011).
24. Shiue, E., Brockman, I. M. & Prather, K. L. J. Improving product yields on D-glucose in Escherichia coli via knockout of pgi and zwf and feeding of supplemental carbon sources. *Biotechnol. Bioeng.* **112**, 579–587 (2015).
25. Hollinshead, W. D. *et al.* Examining Escherichia coli glycolytic pathways, catabolite repression, and metabolite channeling using Δpfk mutants. *Biotechnol. Biofuels* **9**, 212 (2016).

26. Saier Jr, M. H. Families of transmembrane sugar transport proteins. *Mol. Microbiol.* **35**, 699–710 (2000).
27. Liu, J. Y., Miller, P. F., Willard, J. & Olson, E. R. Functional and Biochemical Characterization of Escherichia coli Sugar Efflux Transporters . *J. Biol. Chem.* **274**, 22977–22984 (1999).
28. Chan, K. K., Fedorov, A. A., Fedorov, E. V, Almo, S. C. & Gerlt, J. A. Structural Basis for Substrate Specificity in Phosphate Binding (β/α)8-Barrels: d-Allulose 6-Phosphate 3-Epimerase from Escherichia coli K-12. *Biochemistry* **47**, 9608–9617 (2008).
29. Kim, C., Song, S. & Park, C. The D-allose operon of Escherichia coli K-12. *J. Bacteriol.* **179**, 7631–7637 (1997).
30. Poulsen, T. S., Chang, Y. Y. & Hove-Jensen, B. D-Allose catabolism of Escherichia coli: involvement of alsI and regulation of als regulon expression by allose and ribose. *J. Bacteriol.* **181**, 7126–7130 (1999).
31. Kuznetsova, E. *et al.* Genome-wide Analysis of Substrate Specificities of the Escherichia coli Haloacid Dehalogenase-like Phosphatase Family * . *J. Biol. Chem.* **281**, 36149–36161 (2006).
32. Dubendorf, J. W. & Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **219**, 45–59 (1991).
33. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).
34. Jiang, Y. *et al.* Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **81**, 2506 LP – 2514 (2015).
35. Kotlarz, D. & Buc, H. Two Escherichia coli fructose-6-phosphate kinases. Preparative purification, oligomeric structure and immunological studies. *Biochim. Biophys. Acta* **484**, 35–48 (1977).
36. Kotlarz, D., Garreau, H. & Buc, H. Regulation of the amount and of the activity of phosphofructokinases and pyruvate kinases in Escherichia coli. *Biochim. Biophys. Acta* **381**, 257–268 (1975).
37. Hellinga, H. W. & Evans, P. R. Nucleotide sequence and high-level expression of the major Escherichia coli phosphofructokinase. *Eur. J. Biochem.* **149**, 363–373 (1985).
38. Westwood, A. W. & Doelle, H. W. Glucose 6-phosphate and 6-phosphogluconate dehydrogenases and their control mechanisms in Escherichia coli K-12. *Microbios* **9**, 143–165 (1974).
39. Banerjee, S. & Fraenkel, D. G. Glucose-6-phosphate dehydrogenase from Escherichia coli and from a ‘high-level’ mutant. *J. Bacteriol.* **110**, 155–160 (1972).
40. Roos, A. K., Mariano, S., Kowalinski, E., Salmon, L. & Mowbray, S. L. D-ribose-5-phosphate isomerase B from Escherichia coli is also a functional D-allose-6-phosphate isomerase, while the Mycobacterium tuberculosis enzyme is not. *J. Mol. Biol.* **382**, 667–679 (2008).
41. Zhang, R.-G. *et al.* The 2.2 A resolution structure of RpiB/AlsB from Escherichia

coli illustrates a new approach to the ribose-5-phosphate isomerase reaction. *J. Mol. Biol.* **332**, 1083–1094 (2003).
42. Essenberg, M. K. & Cooper, R. A. Two ribose-5-phosphate isomerases from Escherichia coli K12: partial characterisation of the enzymes and consideration of their possible physiological roles. *Eur. J. Biochem.* **55**, 323–332 (1975).
43. Dietz, G. W. & Heppel, L. A. Studies on the uptake of hexose phosphates. I. 2-Deoxyglucose and 2-deoxyglucose 6-phosphate. *J. Biol. Chem.* **246**, 2881–2884 (1971).
44. Studier, F. W. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219**, 37–44 (1991).
45. Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D. & Voigt, C. A. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Mol. Syst. Biol.* **10**, 742 (2014).
46. Lee, T. S. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J. Biol. Eng.* **5**, 12 (2011).
47. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
48. Brockman, I. M. & Prather, K. L. J. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol. J.* **10**, 1360–1369 (2015).
49. Hawkins, J. S., Wong, S., Peters, J. M., Almeida, R. & Qi, L. S. Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi). *Methods Mol. Biol.* **1311**, 349–362 (2015).
50. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
51. Brückner, R. & Titgemeyer, F. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol. Lett.* **209**, 141–148 (2002).
52. Auzat, I., Le Bras, G. & Garel, J. R. The cooperativity and allosteric inhibition of Escherichia coli phosphofructokinase depend on the interaction between threonine-125 and ATP. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5242–5246 (1994).
53. Chan, K. K., Fedorov, A. A., Fedorov, E. V, Almo, S. C. & Gerlt, J. A. Structural basis for substrate specificity in phosphate binding (beta/alpha)8-barrels: D-allulose 6-phosphate 3-epimerase from Escherichia coli K-12. *Biochemistry* **47**, 9608–9617 (2008).
54. Hernández-Montalvo, V. *et al.* Expression of galP and glk in a Escherichia coli PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. *Biotechnol. Bioeng.* **83**, 687–694 (2003).
55. Brockman, I. M. & Prather, K. L. J. Dynamic knockdown of E. coli central metabolism for redirecting fluxes of primary metabolites. *Metab. Eng.* **28**, 104–113 (2015).
56. Jiang, Y. *et al.* Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
57. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to

generate recombinant DNA via SLIC. *Nat. Methods* **4**, 251–256 (2007).
58. Rodriguez, G. M., Tashiro, Y. & Atsumi, S. Expanding ester biosynthesis in Escherichia coli. *Nat. Chem. Biol.* **10**, 259–265 (2014).
59. Zhang, A. *et al.* Microbial production of human milk oligosaccharide lactodifucotetraose. *Metab. Eng.* **66**, 12–20 (2021).

Chapter 4: Final Thoughts

The work described in this dissertation demonstrates the versatile application of microbial cell platforms to lower the cost and increase the efficiency to produce carbohydrate targets[1]. For the formation of complex human milk oligosaccharides, biorefineries can synthesize and regenerate energy cofactors required for creating nucleotide-activated sugar substrates, thus reducing the cost and complexity of production experienced in traditional chemical synthesis and *in-vitro* enzymatic methods. Native phosphorylation and dephosphorylation mechanisms in microorganisms can create thermodynamic sinks to drive the regio- and stereospecific conversion of common monosaccharides into rare sugars. However, to bring these two technologies up to the industrial scale in a sustainable, low cost, and safe manner for the mass use of milk oligosaccharides and D-psicose in the pharmaceuticals and functional food sector, the following challenges remain to be addressed in the metabolic engineering field.

Currently, the *E. coli* strains established in these two platforms are limited to using highly purified sugar substrates from edible food crop or animal milk sources for chemical synthesis, which can threaten our global food security that is already impacted by extreme weather conditions bought on by greenhouse-gas mediate climate change[2–4]. The development of microbial strains with degradation pathways of renewable materials containing suitable carbohydrate precursors would avoid the dilemma between food consumption and carbohydrate commodities production. Over the last few years, brown algae has emerged as a promising inexpensive and renewable carbon source that can serve the dual purpose of sequestering atmospheric CO_2 and turning it

into biomass rich in polyoses that can be broken down into various monosaccharide substrates[5–7]. It has already been demonstrated that ethanol can be fermented from unpretreated kelp powder at a yield of 0.25 g/g kelp in thermophilic bacterium *Defluviitalea phaphyphila*[8] and from treated kelp at 0.28 g/g kelp in *E. coli*[9] and at 0.12 g/g in *Saccharomyces cerevisiae*[10]. Fucoidan, found in the cell walls of seaweed, would be an excellent substrate for HMOs and rare sugar synthesis, for it can be broken down into L-fucose, D-galactose, D-glucose, D-xylose, and D-mannose[11] that are all be able to be catabolized by *E. coli*. Advancements in recent years have been made in establishing more environmentally friendly enzyme-assisted fucoidan extraction methods[12,13] to replace traditional mild acidic, neutral hot water, and organic solvent extraction strategies[14,15]. Due to macroalage's important roles as key habitat-structuring agents in marine ecology, great emphasis is also being placed in the sustainable management of seaweed aquaculture to minimize environmental risks brought on by large-scale farming of this renewable resource[16,17].

The transition into using mixed sugar feedstocks from renewable carbon sources should also be accompanied with increased omics and ^{13}C flux analyses[18–21] to observe the effects of mixed sugar metabolism on biomass formation, cofactor energy balance, and activity of key metabolic pathways. Glucose metabolism has been well studied over the years, but analysis of other monosaccharides' metabolism has only been recently initiated by the interest in using lignocellulose and macroalgae biomass as a renewable carbon source[22–26]. Although co-sugar substrate conditions have been applied to produce LNT from lactose and galactose in *E. coli*[27] and 2'-FL from lactose and xylose in *S. cerevisiae*[28], only a few studies on the mechanism of regulating sugar co-utilization

are available for reference[29–31]. Information gathered from these future studies will be essential in rationalizing new strategies in strain design to improve product titer, yield and productivity.

While *E. coli* is an excellent host for proof-of-principle chemical production in the lab-space, it may not be the best platform for the industrial scale-up production of pharmaceutical- and food-grade products for human consumption. *E. coli* natively produces lipopolysaccharides (LPS), a major component attached to its outer membrane leaflet, and its lipid A moiety acts as an endotoxin agonist of the Toll-like receptor 4/myeloid differentiation factor in mammals that can trigger septic shock[32]. Contamination of endotoxins in *in vitro* tissue culture studies that elucidate the biological effects of carbohydrate products may lead to false positives in toxicological response. Technologies have been developed to detect and remove endotoxin for biopharmaceutical purification of recombinant proteins[33–35] but adds significant manufacturing costs on the industrial scale. Therefore, using endotoxin-free *E. coli* strains would be a proactive upstream preventative option for mitigating safety concerns. In 2015, LPS-free derivatives of *E. coli* K-12 and BL21 (DE3) were constructed and shown to effectively produce recombinant proteins with negligible endotoxin contamination but have not yet caught mainstream attention in bulk and fine chemical production[36]. Additionally, it would also be important to examine the effects of outer membrane modifications in the endotoxin-free *E. coli* strains on long-term cell stability and robustness against physical, mechanical, and chemical stresses in commercial bioreactors. Alternatively, we can translate proof-of-concept technologies into naturally endotoxin-free organisms such as the Gram-positive *Bacillus subtilis*,

lactic acid bacteria*, Corynebacterium* species and *Saccharomyces* cerevisiae. Advancements in adapting CRISPR gene editing technologies[37–39] and standardization of synthetic biology toolboxes for non-traditional hosts[40–42] in the last few years have helped to secure alternative microorganisms a place in industrial bioproduction.

Overall, great strides have been made in altering the biology of microorganisms to turn simple sugars into functional commodities, but we have only begun to scratch the surface with the types of carbohydrate nutraceuticals we can produce in these microbial cell factories. Designing novel pathways for new-to-nature carbohydrates is currently limited by the rate of enzyme discovery and de novo construction of enzymes still needs time for improvement[43,44]. Therefore, more focus should be placed into engineering existing enzymes to repurpose their functions. The recent public release of AlphaFold[45], a program created by Google's DeepMind that combines of bioinformatics with machine learning to predict protein structures from gene sequences, will expand our understanding of biological structures and assist in rational protein design for new enzymatic reactions. The gene sequences, structures and mechanisms of glycosyltransferases[46,47] and sugar epimerases[48] that have been elucidated can be provided as learning data sets for artificial intelligence to mine for carbohydrate enzymes from genomic sequencing data of novel microorganisms and predict their structure-activity relationships.

References

1. Zhang, A. *et al.* Microbial production of human milk oligosaccharide lactodifucotetraose. *Metab. Eng.* **66**, 12–20 (2021).
2. Tenenbaum, D. J. Food vs. fuel: diversion of crops could cause more hunger. *Environ. Health Perspect.* **116**, A254–A257 (2008).
3. Tomei, J. & Helliwell, R. Food versus fuel? Going beyond biofuels. *Land use policy* **56**, 320–326 (2016).

4. Graham-Rowe, D. Agriculture: Beyond food versus fuel. *Nature* **474**, S6–S8 (2011).
5. Wang, D., Kim, D. H. & Kim, K. H. Effective production of fermentable sugars from brown macroalgae biomass. *Appl. Microbiol. Biotechnol.* **100**, 9439–9450 (2016).
6. Gao, G., Burgess, J. G., Wu, M., Wang, S. & Gao, K. Using macroalgae as biofuel: current opportunities and challenges. *Bot. Mar.* **63**, 355–370 (2020).
7. Milledge, J. J., Nielsen, B. V & Bailey, D. High-value products from macroalgae: the potential uses of the invasive brown seaweed, Sargassum muticum. *Rev. Environ. Sci. Bio/Technology* **15**, 67–88 (2016).
8. Ji, S.-Q., Wang, B., Lu, M. & Li, F.-L. Direct bioconversion of brown algae into ethanol by thermophilic bacterium Defluviitalea phaphyphila. *Biotechnol. Biofuels* **9**, 81 (2016).
9. Wargacki, A. J. *et al.* An Engineered Microbial Platform for Direct Biofuel Production from Brown Macroalgae. *Science (80-.).* **335**, 308 LP – 313 (2012).
10. Ruangrit, K. *et al.* Environmental-friendly pretreatment and process optimization of macroalgal biomass for effective ethanol production as an alternative fuel using Saccharomyces cerevisiae. *Biocatal. Agric. Biotechnol.* **31**, 101919 (2021).
11. Li, B., Lu, F., Wei, X. & Zhao, R. Fucoidan: structure and bioactivity. *Molecules* **13**, 1671–1695 (2008).
12. Nguyen, T. T. *et al.* Enzyme-Assisted Fucoidan Extraction from Brown Macroalgae Fucus distichus subsp. evanescens and Saccharina latissima. *Marine Drugs* vol. 18 (2020).
13. Lakmal, H. H. C., Lee, J.-H. & Jeon, Y.-J. Enzyme-Assisted Extraction of a Marine Algal Polysaccharide, Fucoidan and Bioactivities BT - Polysaccharides: Bioactivity and Biotechnology. in (eds. Ramawat, K. G. & Mérillon, J.-M.) 1065–1077 (Springer International Publishing, 2015). doi:10.1007/978-3-319-16298-0_46.
14. January, G. G., Naidoo, R. K., Kirby-McCullough, B. & Bauer, R. Assessing methodologies for fucoidan extraction from South African brown algae. *Algal Res.* **40**, 101517 (2019).
15. Lim, S. J. & Wan Aida, W. M. Chapter 3 - Extraction of Sulfated Polysaccharides (Fucoidan) From Brown Seaweed. in (eds. Venkatesan, J., Anil, S. & Kim, S.-K. B. T.-S. P.) 27–46 (Elsevier, 2017). doi:https://doi.org/10.1016/B978-0-12-809816-5.00003-7.
16. Campbell, I. *et al.* The Environmental Risks Associated With the Development of Seaweed Farming in Europe - Prioritizing Key Knowledge Gaps . *Frontiers in Marine Science* vol. 6 107 (2019).
17. Charrier, B. *et al.* Furthering knowledge of seaweed growth and development to facilitate sustainable aquaculture. *New Phytol.* **216**, 967–975 (2017).
18. Ishii, N. & Tomita, M. Multi-Omics Data-Driven Systems Biology of E. coli BT - Systems Biology and Biotechnology of Escherichia coli. in (ed. Lee, S. Y.) 41–57 (Springer Netherlands, 2009). doi:10.1007/978-1-4020-9394-4_3.
19. Amer, B. & Baidoo, E. E. K. Omics-Driven Biotechnology for Industrial Applications . *Frontiers in Bioengineering and Biotechnology* vol. 9 30 (2021).
20. Guo, W., Sheng, J. & Feng, X. (13)C-Metabolic Flux Analysis: An Accurate Approach to Demystify Microbial Metabolism for Biochemical Production. *Bioeng.*

(Basel, Switzerland) **3**, 3 (2015).
21. Basler, G., Fernie, A. R. & Nikoloski, Z. Advances in metabolic flux analysis toward genome-scale profiling of higher organisms. *Biosci. Rep.* **38**, (2018).
22. Kim, J., Cheong, Y. E., Jung, I. & Kim, K. H. Metabolomic and Transcriptomic Analyses of Escherichia coli for Efficient Fermentation of L-Fucose. *Mar. Drugs* **17**, (2019).
23. Veras, H. C. T. *et al.* Metabolic flux analysis for metabolome data validation of naturally xylose-fermenting yeasts. *BMC Biotechnol.* **19**, 58 (2019).
24. Wasylenko, T. M. & Stephanopoulos, G. Metabolomic and (13)C-metabolic flux analysis of a xylose-consuming Saccharomyces cerevisiae strain expressing xylose isomerase. *Biotechnol. Bioeng.* **112**, 470–483 (2015).
25. Bergdahl, B., Heer, D., Sauer, U., Hahn-Hägerdal, B. & van Niel, E. W. J. Dynamic metabolomics differentiates between carbon and energy starvation in recombinant Saccharomyces cerevisiae fermenting xylose. *Biotechnol. Biofuels* **5**, 34 (2012).
26. Lixia, L. *et al.* Phosphoketolase Pathway for Xylose Catabolism in Clostridium acetobutylicum Revealed by 13C Metabolic Flux Analysis. *J. Bacteriol.* **194**, 5413–5422 (2012).
27. Baumgärtner, F., Sprenger, G. A. & Albermann, C. Galactose-limited fed-batch cultivation of Escherichia coli for the production of lacto-N-tetraose. *Enzyme Microb. Technol.* **75**–**76**, 37–43 (2015).
28. Lee, J. W. *et al.* Enhanced 2′-Fucosyllactose production by engineered Saccharomyces cerevisiae using xylose as a co-substrate. *Metab. Eng.* **62**, 322–329 (2020).
29. Salvy, P. & Hatzimanikatis, V. Emergence of diauxie as an optimal growth strategy under resource allocation constraints in cellular metabolism. *Proc. Natl. Acad. Sci.* **118**, e2013836118 (2021).
30. Lin, Z., Satarupa, D. & A., B. R. Utilization of Lactose and Galactose by Streptococcus mutans: Transport, Toxicity, and Carbon Catabolite Repression. *J. Bacteriol.* **192**, 2434–2444 (2010).
31. Ammar, E. M., Wang, X. & Rao, C. V. Regulation of metabolism in Escherichia coli during growth on mixtures of the non-glucose sugars: arabinose, lactose, and xylose. *Sci. Rep.* **8**, 609 (2018).
32. Yang, K. H. & Lee, M. G. Effects of endotoxin derived from Escherichia coli lipopolysaccharide on the pharmacokinetics of drugs. *Arch. Pharm. Res.* **31**, 1073–1086 (2008).
33. Lopes, A. M. *et al.* LPS removal from an E. coli fermentation broth using aqueous two-phase micellar system. *Biotechnol. Prog.* **26**, 1644–1653 (2010).
34. Schneier, M., Razdan, S., Miller, A. M., Briceno, M. E. & Barua, S. Current technologies to endotoxin detection and removal for biopharmaceutical purification. *Biotechnol. Bioeng.* **117**, 2588–2609 (2020).
35. Fung, F. M., Su, M., Feng, H. & Li, S. F. Y. Extraction, separation and characterization of endotoxins in water samples using solid phase extraction and capillary electrophoresis-laser induced fluorescence. *Sci. Rep.* **7**, 10774 (2017).
36. Mamat, U. *et al.* Detoxifying Escherichia coli for endotoxin-free production of recombinant proteins. *Microb. Cell Fact.* **14**, 57 (2015).

37. Wang, Q. *et al.* Advances and Perspectives for Genome Editing Tools of Corynebacterium glutamicum . *Frontiers in Microbiology* vol. 12 610 (2021).
38. Cai, P., Gao, J. & Zhou, Y. CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications. *Microb. Cell Fact.* **18**, 63 (2019).
39. Hao, W. *et al.* Design and Construction of Portable CRISPR-Cpf1-Mediated Genome Editing in Bacillus subtilis 168 Oriented Toward Multiple Utilities . *Frontiers in Bioengineering and Biotechnology* vol. 8 1043 (2020).
40. Popp, P. F., Dotzler, M., Radeck, J., Bartels, J. & Mascher, T. The Bacillus BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with Bacillus subtilis. *Sci. Rep.* **7**, 15058 (2017).
41. Walker, R. S. K. & Pretorius, I. S. Applications of Yeast Synthetic Biology Geared towards the Production of Biopharmaceuticals. *Genes (Basel).* **9**, 340 (2018).
42. Woo, H. M. Synthetic Biology for Corynebacterium glutamicum: An Industrial Host for White Biotechnology. *Emerging Areas in Bioengineering* 321–329 (2018) doi:https://doi.org/10.1002/9783527803293.ch18.
43. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
44. Pan, X. & Kortemme, T. Recent advances in de novo protein design: Principles, methods, and applications. *J. Biol. Chem.* **296**, 100558 (2021).
45. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* (2021) doi:10.1038/s41586-021-03819-2.
46. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
47. Breton, C., Fournel-Gigleux, S. & Palcic, M. M. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.* **22**, 540–549 (2012).
48. Allard, S. T. M., Giraud, M.-F. & Naismith, J. H. Epimerases: structure, function and mechanism. *Cell. Mol. Life Sci. C.* **58**, 1650–1665 (2001).

Appendix: Characterizing Mutations Acquired Through Adaptive Laboratory Evolution of *Escherichia coli* For Enhanced Tolerance Towards Isobutyl Acetate

A.1 Introduction

Over 6000 products, such as gasoline, plastics, food flavorings, fragrances, cosmetics, and pharmaceuticals, are derived from petroleum fossil fuels. In 2020, the United States alone consumed 6.63 billion barrels of petroleum to supply energy for transportation and heating and to synthesize precursor chemicals for manufacturing essential commodities. However, there is a finite quantity to this naturally occurring raw material, and at the current rate of global petroleum consumption this resource will be depleted within 47 years. Having released over 177 million metric tons of carbon dioxide (CO_2) into the atmosphere in 2019, petroleum refineries are also major contributors to greenhouse gas emissions that are driving global climate change. Therefore, it is critical within this century for our society to establish production methods of petrochemicals in a sustainable manner with a lower carbon footprint on the environment.

Microbial cell factories have emerged as compatible production platforms that can be translated to current petroleum refinery infrastructures. *Escherichia coli*, *Saccharomyces cerevisiae*, and *Bacillus* sub-strains are a few host microorganisms that have been successfully engineered to produce short-chained drop-in biofuels and solvents such as ethanol[1], butanol[2], isobutanol[3], and their respective esters ethyl acetate[4], butyl acetate[5], and isobutyl acetate[6] (IBA) in gram-scale titers. While great strides have been made in optimizing the metabolism of the host chassis and in engineering more efficient and robust enzyme catalysts for biofuel production[7–12], the product's toxic effect on the host organism is still a limiting factor in establishing

microbial chemical production at the multi-gram scale to be economically competitive against current petroleum-based chemical syntheses[13].

In our lab's previous work, an *E. coli* strain was established to produce IBA[6,14] and we determined that the strain has an IBA toxicity limit of 2 g/L, which makes it immensely difficult to maintain cell growth and sustain cell viability during fermentation. To overcome this growth burden, we implemented a hexadecane bilayer culturing strategy to extract IBA from the aqueous culture supernatant into the organic solvent layer so that the produced IBA can be sequestered from the microbial host[6,14]. However, this method increases production costs, complicates culturing conditions, and is not environmentally friendly. As an alternative to this liquid-liquid extraction method, we are interested in harnessing IBA's innate low solubility of 6 g/L in aqueous conditions to facilitate the formation of its own bilayer. To use this culturing strategy, we first need to increase the toxicity limit of E. coli towards IBA.

In this project, we used adaptive laboratory evolution (ALE) and P1-phage mediated genome shuffling techniques on our lab's established IBA production strain AL17 to generate mutants with increased tolerance towards IBA. For this section of the dissertation, we will only be characterizing mutations found in strains collected from the first 500 generations grown under the selective pressure of increasing concentrations of IBA. Whole genome sequencing revealed that mutants M1 to M7 collected from this ALE process have three shared mutations: an SNP within each of the genes of *rho* and *metH* and an IS30 insertion within the gene *yjjY*. We used CRISPR-Cas9 mediated genome editing to construct single, double, and triple mutants containing the above mutations to test their effects on cell growth and discovered that only the triple mutant

exhibited similar tolerance levels towards IBA as the M1 mutant, indicating that the epistatic effects of all three mutations are essential in alleviating the toxicity response in E. coli. We also tested M1 against AL17 for IBA production and discovered that the strain can produce up to 3 g/L of IBA in 24 h, a 2-fold increase to AL17's 1.5 g/L of IBA.

A.2 Results

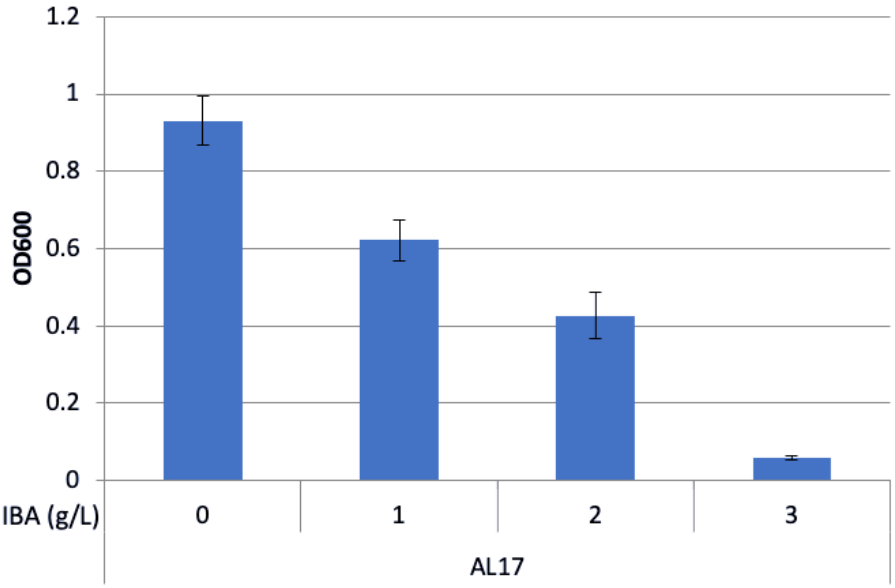


Figure A1. IBA tolerance of parent strain AL17. AL17 was inoculated to an OD_{600} 0.1 in M9 production media with increasing concentrations of IBA (1, 2, 3, and 4 g/L). Cultures were grown at 37 ºC for 24 h and culture density was measured at OD_{600}. Error bars indicate s.d. (n = 4 biological replicates).

Parent strain AL17 (Table A1) exhibits decreased cell growth in the presence of IBA starting at 1 g/L and cell growth is completely inhibited at 3 g/L (Figure A1). To increase AL17's tolerance towards IBA, ALE was performed on this strain by sequential serial passages approximately 500 generations. Due to the volatility of IBA, screw cap shake flasks were used. Cultures were initially grown in M9 production media containing 1.5 g/L of IBA as a selective pressure and cultures with increased fitness that grew above an OD_{600} of 1.0 after 24 h were passaged under the selective pressure of

increasing concentrations of IBA at the rate of 0.05 g/L/24 h. Seven mutants M1 through M7 (Table A1) were collected across the 500 generations and their genomic DNA were sent in for whole genome sequencing. A total of seven mutations were identified on the genome: two missense SNPs within the genes of *rho* and *metH*, two silent SNPs within the genes of *nanK* and *nanE*, one SNP in the intergenic region between genes of *ykgR* and *ykgP*, one 1 bp indel in the gene of *lhr*, and one IS30 insertion within the gene of *yjjY* (Figure A2). Of these seven mutations, only the *rho* and *metH* SNPs and the yjjY IS30 insertion are conserved between mutant strains M1 through M7. In an IBA tolerance assay against AL17, M1 exhibits cell growth in the presence of 3 g/L, suggesting that one or more of the three mutations are responsible for decreasing *E. coli*'s toxicity response towards this ester.

Gene	Description	Mutation	Position	M1	M2	M3	M4	M5	M6	M7
rho	transcription termination factor	G→T	G63C (GGT→TGT)	■	■	■	■	■	■	■
metH	homocysteine-N5-methyltetrahydrofolate transmethylase, B12-dependent	C→T	A387V (GCG→GTG)	■	■	■	■	■	■	■
yjjY	uncharacterized protein	IS30 (+) +2 bp	coding (17-18/141 nt)	■	■	■	■	■	■	■
ykgR / *ykgP*	uncharacterized protein/pseudogene, oxidoreductase family	G→A	intergenic (-301/+173)		■	■				
nanK	N-acetylmannosamine kinase	T→C	M1M (ATG→GTG)				■			
nanE	putative N-acetylmannosamine-6-P epimerase		L229L (CTA→CTG)				■			
lhr	putative ATP-dependent helicase	+C	coding (2399/4617 nt)				■			

Figure A2. Mutations identified from genome sequencing analysis from mutant strains M1 through M7. Shaded grey block indicates presence of the mutation in the respective strains.

To test the individual effects of the three mutations in AL17, we constructed three single mutants using CRISPR-Cas9 mediate genome editing. For the *rho* SNP, a 250 bp linear donor DNA was designed with the G to T missense mutation along with three

silent mutations to the sgRNA recognition sequence region to prevent Cas9 from recutting the rho locus. This CRISPR reaction generated Strain 1 (Table A1). Similarly with the *metH* SNP, a 250 bp linear donor DNA was designed with the C to T missense mutation along with one silent mutation to the PAM recognition sequence to prevent Cas9 from recutting the *metH* locus. This CRISPR reaction generated Strain 2 (Table A1). For the *yjjY* IS30 insertion, we initially attempted to replicate the mutation by amplifying the 1221 bp IS30 insertion sequence from the *yjjY* locus in M1 with additional 50bp homology arms, which contains a 1 bp mutation to the PAM recognition sequence to prevent recutting by Cas9 at the *yjjY* locus, as the linear donor DNA. However, all transformants screened after CRISPR editing contained wild-type *yjjY*. We examined the position of the *yjjY* gene in the E. coli genome and saw that *yjjY* is encoded between the gene *arcA* and a series of seven *arcA* promoters. We hypothesized that the yjjY IS30 insertion may act as a transcriptional disrupter to the *arcA* operon, therefore as a

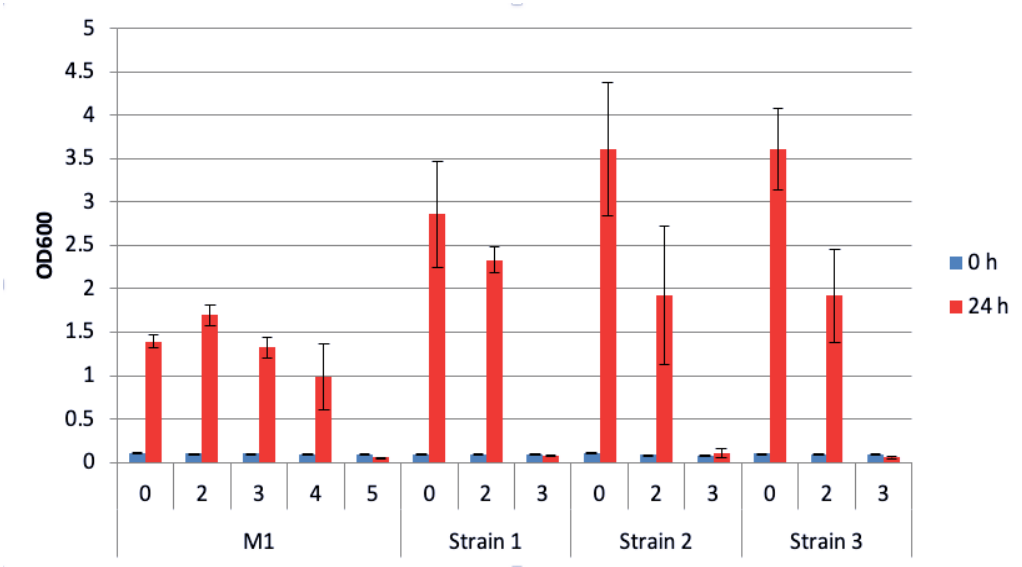


Figure A3. IBA tolerance of single mutants. M1, Strains 1, 2, and 3 were inoculated to an OD_{600} 0.1 in M9 production media with increasing concentrations of IBA (0, 2, 3, 4, and 5 g/L, respectively). Cultures were grown at 37 ºC for 24 h and culture density was measured at OD_{600}. Error bars indicate s.d. (n = 3 biological replicates).

substitute for the IS30 insertion, we deleted the genomic region containing *arcA-yjjY-arcA*$_{P1\text{-}7}$ from AL17. This CRISPR reaction generated Strain 3 (Table A1).

In an IBA tolerance experiment against M1, Strains 1, 2, and 3 did not exhibit growth at 3 g/L of IBA (Figure A3). This indicates that the primary effect of each individual mutation does not confer higher tolerance towards IBA and that perhaps an epistatic effect of combining two of the three mutations together may be necessary for the tolerant phenotype. To test the synergistic effect of paired mutations, we constructed double mutants containing the following mutations: *rho* + *metH* (Strain 4, Table A1), *rho* + Δ*arcA-yjjY-arcA*$_{p1\text{-}7}$ (Strain 5, Table A1), and *metH* + Δ*arcA-yjjY-arcA*$_{p1\text{-}7}$ (Strain 6, Table A1). Similar Strains 1, 2, and 3, Strains 4, 5, and 6 did not exhibit the same level of tolerance as M1 in the presence of 3 g/L IBA (Figure A4). Strain 4 grew to an OD_{600} of approximately 0.5 whereas Strains 5 and 6 remained at an OD_{600} of 0.1.

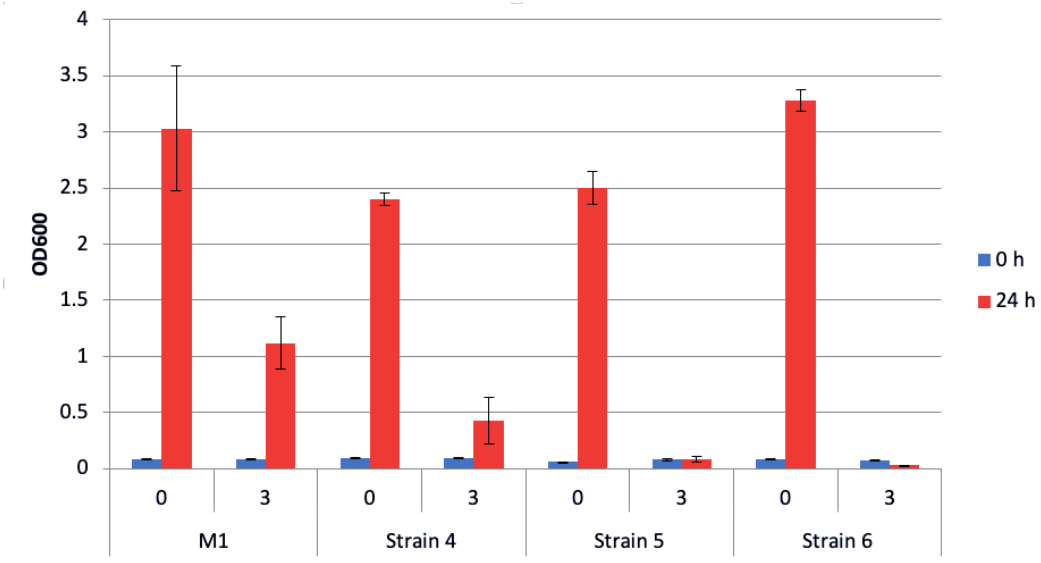


Figure A4. IBA tolerance of double mutants. M1, Strains 4, 5, and 6 were inoculated to an OD_{600} 0.1 in M9 production media with increasing concentrations of IBA (0 and 3 g/L, respectively). Cultures were grown at 37 ºC for 24 h and culture density was measured at OD_{600}. Error bars indicate s.d. (n = 3 biological replicates).

To test the synergistic effect of all three combined mutations, we constructed a triple mutant of *rho* + *metH* + Δ*arcA-yjjY-arcA*$_{p1-7}$ in AL17, forming Strain 7 (Table A1). Strain 7 displayed growth at 3 g/L of IBA with an approximate OD_{600} of 1.5, which is comparable to M1's growth at 3 g/L of IBA (Figure A5). The combination of all three mutations enabled higher tolerance towards IBA, indicating that all three mutations are

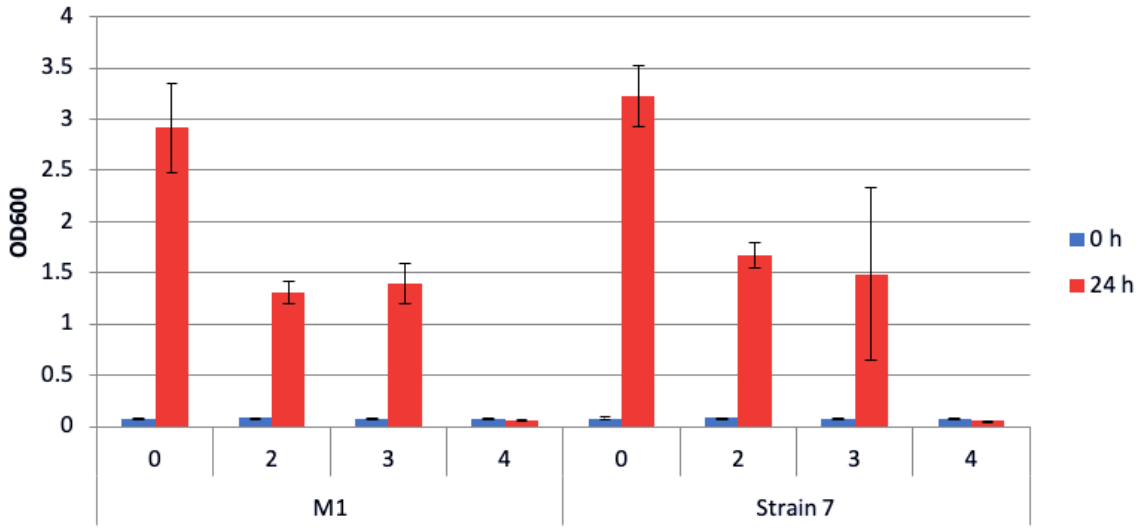


Figure A5. IBA tolerance of single mutants. M1 and Strain 7 were inoculated to an OD_{600} 0.1 in M9 production media with increasing concentrations of IBA (0, 2, 3, and 4 g/L, respectively). Cultures were grown at 37 ºC for 24 h and culture density was measured at OD_{600}. Error bars indicate s.d. (n = 3 biological replicates).

necessary for this improved fitness.

From the ALE process, we isolated strains with an improved IBA toxicity limit raised from 2 g/L to 3 g/L. With this improved fitness, mutant strain M1 may be able to produce higher levels of IBA than parent strain AL17. To test the production of IBA from glucose in AL17 and M1, a plasmid encoding an isobutanol (ISO) production pathway under a P_{LlacO1} promoter (pAL603) and a plasmid encoding an alcohol acetyltransferase *ATF1*, which converts ISO into IBA, under a P_{LlacO1} promoter (pAL1114) were introduced into the strains to form Strains 8 and 9, respectively (Table A1). From 50 g/L glucose,

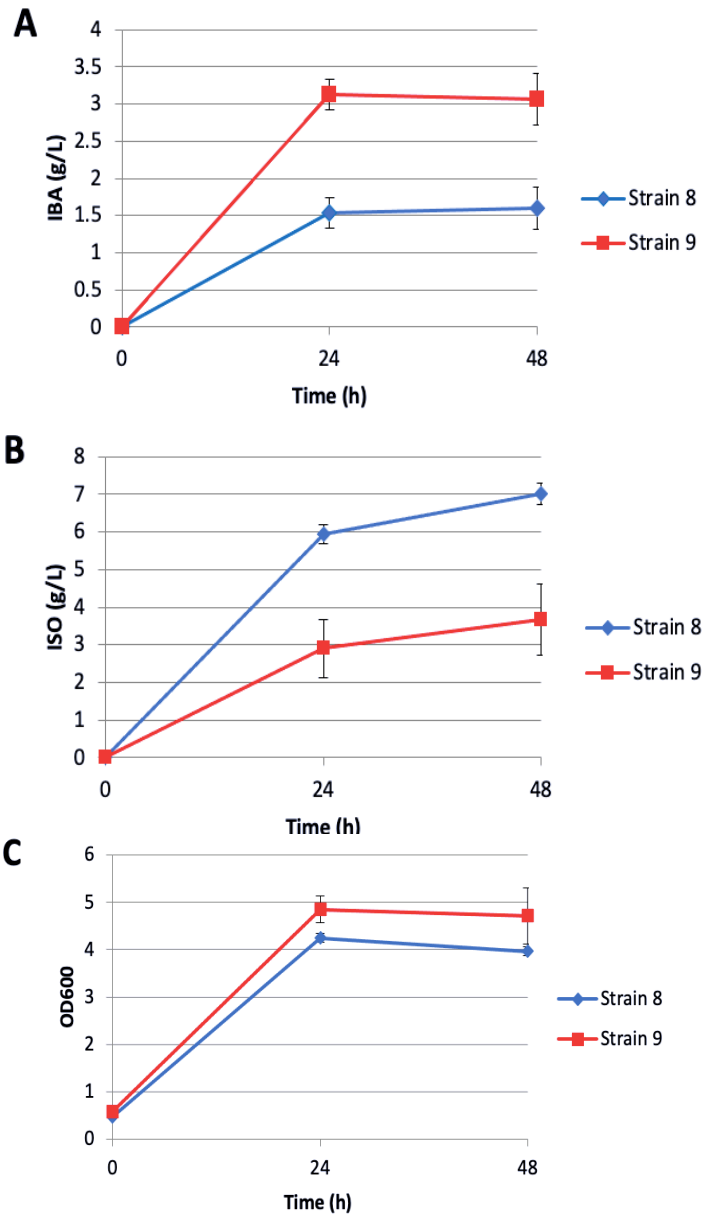


Figure A6. IBA production and ISO accumulation in Strains 8 and 9. Strains 8 and 9 were induced with 1 mM IPTG at OD_{600} 0.4 and grow at 30 ºC for 48 h. 1 mL of culture was collected at 0, 24, and 48 h for optical density measurement and GC analysis. (A) IBA production reported in g/L. (B) ISO production reported in g/L. (C) Optical density reported in OD_{600}. Error bars indicate s.d. ($n \geq 3$ biological replicates).

Strain 8 produced 1.5 g/L IBA with 6.0 g/L ISO accumulated in 24 h (Figure A6). Strain

9 produced 3.1 g/L IBA with a lower accumulated level of ISO at 2.9 g/L in 24 h (Figure

A6). The mutations acquired through the ALE process enabled Strain 9 to produce IBA up to its toxicity limit of approximately 3 g/L.

To characterize the individual mutations' effect in M1 on IBA production, ISO accumulation and culture growth, we introduced production plasmids pAL603 and pAL1114 into Strains 1, 2, and 3 to form Strains 10, 11, and 12, respectively. From the effects of the single mutations, Strain 10 containing the *rho* SNP surprisingly produced the most IBA at 4.0 g/L in 24 h, which exceeds the toxicity limit of M1, whereas Strain 12 containing the Δ*arcA-yjjY-arcA*$_{P1\text{-}7}$ had the poorest performance, producing only 0.7 g/L of IBA in 24 h (Figure A7). Strain 11 containing the *metH* SNP produced similar quantities of IBA as Strain 8 at approximately 3.0 g/L in 24 h. Strains 8, 10, and 11 accumulated between 4 to 6 g/L ISO in 24 h while Strain 12 accumulated 9.5 g/L ISO in 24 h and increased to 13 g/L by 48 h (Figure A7). All strains exhibited a similar density of OD_{600} 4.0–5.0 at 24 h with no increase in growth at 48 h.

We further characterized the effects of paired mutations and the group effect of the three combined mutations on IBA production, ISO accumulation and culture growth by introducing pAL603 and pAL1114 into Strains 4, 5, 6 and 7 to form Strains 13, 14, 15, and 16, respectively. Strain 13 containing the *rho* and *metH* SNP produced an even greater amount of IBA than Strain 10 at 6 g/L in 24 h (Figure A8). Strain 14 containing the *rho* SNP and Δ*arcA-yjjY-arcA*$_{P1\text{-}7}$ produced 1.2 g/L of IBA and Strain 15 containing the *metH* SNP and Δ*arcA-yjjY-arcA*$_{P1\text{-}7}$ produced 1.7 g/L of IBA (Figure A8). For ISO accumulation, Strain 13 had the lowest accumulation at 4 g/L while Strains 14 and 15 had 9 g/L in 24 h (Figure A8). With the three combined mutations, Strain 16 produced only 1.7 g/L IBA and accumulated 8 g/L ISO in 24 h.

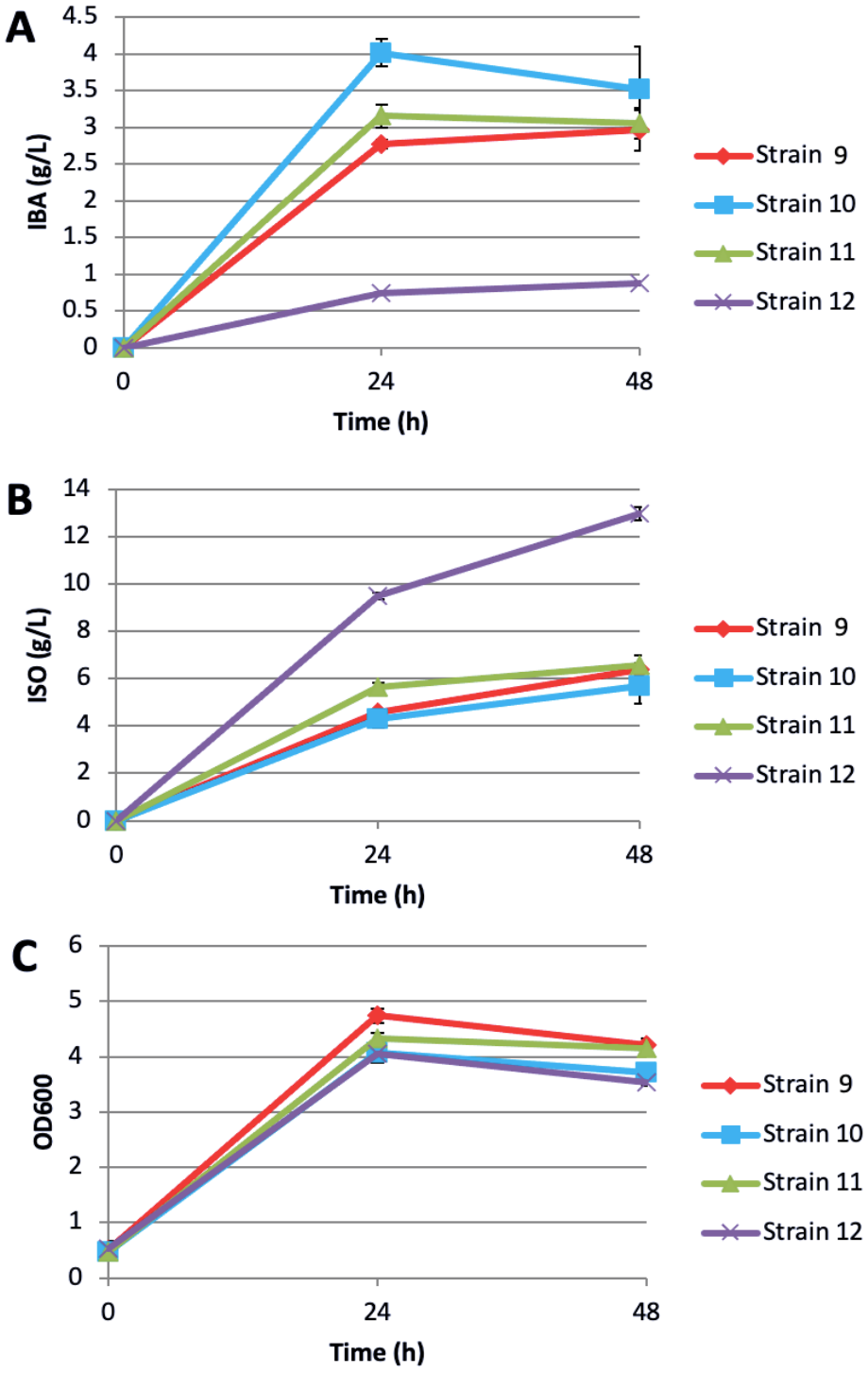


Figure A7. Effect of single mutations on IBA production and ISO accumulation. Strains 9, 10, 11, and 12 were induced with 1 mM IPTG at OD_{600} 0.4 and grow at 30 ºC for 48 h. 1 mL of culture was collected at 0, 24, and 48 h for optical density measurement and GC analysis. (A) IBA production reported in g/L. (B) ISO production reported in g/L. (C) Optical density reported in OD_{600}. Error bars indicate s.d. ($n \geq 3$ biological replicates).

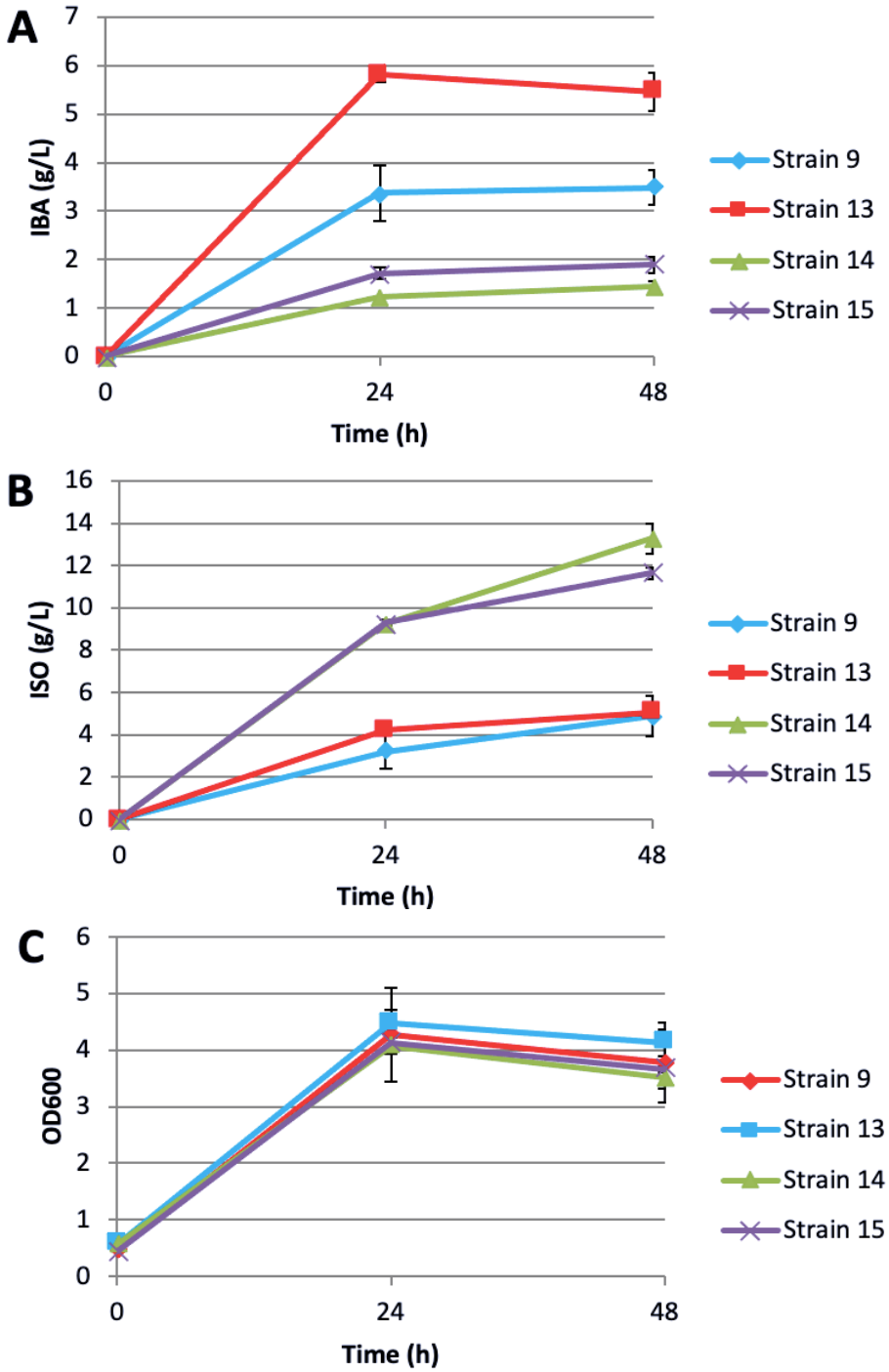


Figure A8. Effect of paired mutations on IBA production and ISO accumulation. Strains 9, 13, 14, and 15 were induced with 1 mM IPTG at OD_{600} 0.4 and grown at 30 ºC for 48 h. 1 mL of culture was collected at 0, 24, and 48 h for optical density measurement and GC analysis. (A) IBA production reported in g/L. (B) ISO production reported in g/L. (C) Optical density reported in OD_{600}. Error bars indicate s.d. ($n \geq 3$ biological replicates).

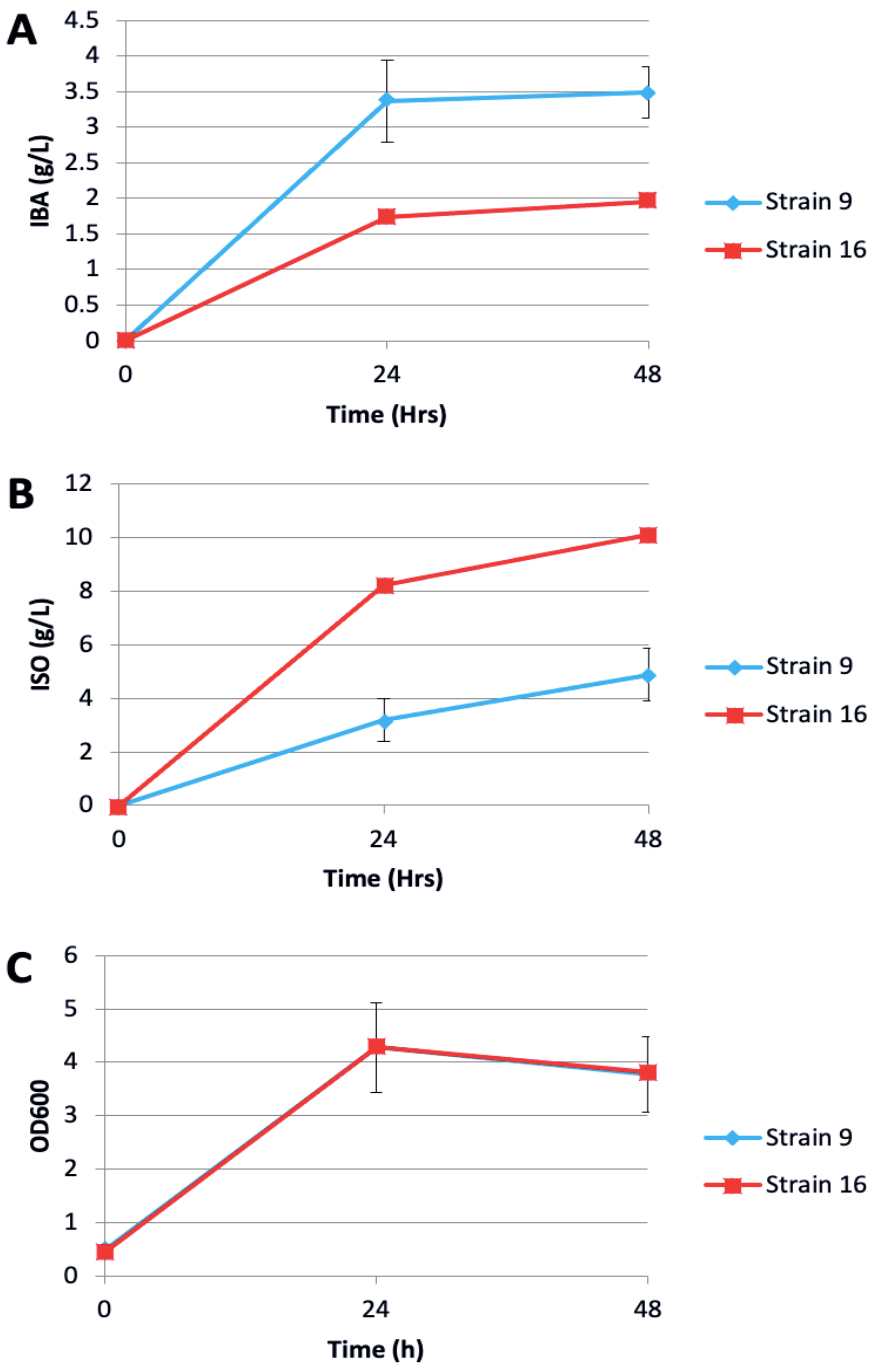


Figure A9. Effect of all three combined mutations on IBA production, ISO accumulation, and strain growth. Strains 9 and 16 were induced with 1 mM IPTG at OD_{600} 0.4 and grown at 30 ºC for 48 h. 1 mL of culture was collected at 0, 24, and 48 h for optical density measurement and GC analysis. (A) IBA production reported in g/L. (B) ISO production reported in g/L. (C) Optical density reported in OD_{600}. Error bars indicate s.d. ($n \geq 3$ biological replicates).

A.3 Discussion

In this study, we demonstrated that the adaptive laboratory evolution of *E. coli* to increasing levels of IBA can facilitate changes in its genome to confer higher resistance against this toxic compound. Using genome sequencing analysis, we found seven mutations among the seven isolated strains from the serial dilution evolution process, with three of the mutations present in all the strains. To discern the effects of each of the three mutations, we used CRISPR-Cas9 mediate genome editing to reconstruct them in our original parent strain of the ALE process. From the reconstructed strains, we were able to observe how the three mutations contributed to our desired phenotype towards IBA.

During the evolution process, we isolated strain M1 that has an increased tolerance against IBA up to 3 g/L. In our sequence analysis, we used the parent strain AL17's genome as a reference and identified three of the following mutations: a G → T SNP in the gene *rho*, a C → T SNP in the gene *metH*, and an IS30 insertion element within the gene *yjjY*. The IBA tolerance phenotype may be attributed to one or more of these three mutations, so we systematically reconstructed each single, double and triple mutation combination to demonstrate their effects.

The gene *rho* encodes for Rho, a transcriptional terminator regulator protein, and it has been shown that mutations to the *rho* gene are effective in increasing the tolerance of *E. coli* strains towards ethanol[15]. Exposure to aliphatic alcohols can disrupt *E. coli*'s membrane peptidoglycan and lipid composition and can inhibit in-vitro transcriptional elongation by RNA polymerase[15,16]. Catalytic inactivation of Rho enhances mRNA transcription, which counteracts transcriptional interference by

alcohols. Due to IBA's hydrophobicity, it may impose similar effects on *E. coli* as ethanol to disrupt these cellular functions. While Strain 1, the reconstructed single SNP mutant of *rho,* did not display tolerance towards the sudden exposure towards 3 g/L of IBA in our toxicity experiments (Figure A4), this individual mutation enabled Strain 10 to produce up to 4 g/L of IBA in a 24 h period (Figure A7). This missense point mutation in *rho* that converts amino acid residue 63 from a glycine to a cysteine may catalytically deactivate the protein and globally enhance the production of mRNA transcripts for enzymes that slowly acclimates *E. coli* to increasing concentrations of this toxic stressor.

Integration of the IS30 insertion element into the gene yjjY disrupts the transcriptional operon of *arcA*, which encodes for ArcA, a quinone-dependent DNA transcriptional regulator. Its expression is enhanced as a response to respiratory distress caused by isobutanol disrupting quinone-membrane interactions[17]. It has also been shown that ArcA is a redox regulator in micro-anaerobic conditions[18]. Δ*arcA* mutants have enhanced expression of NADH-producing enzymes in the Krebs cycle, which encourages cellular respiration in oxygen-deprived conditions. Removing the arcA operon was not effective in increasing the fitness of Strain 3 against IBA but was successful in improving ISO accumulation in Strain 12 up to 9.5 g/L (Figure A7). Our previous work in testing the ISO tolerance of parent strain AL17 indicates that at 8 g/L of ISO cell growth is severely hindered (Figure A.S1), which suggests removing *arcA* expression improves ISO tolerance. Although this individual mutation does not seem to directly contribute to the improved IBA fitness of strain M1, the acquisition and retention

of this mutation in throughout the ALE process suggests that the removal of *arcA* serves to acclimate E. coli to micro-anaerobic ALE culturing conditions.

Lastly, the gene *metH* encodes for MetH, a B12-dependent homocysteine-N5-methyltetrahydrofolate transmethylase responsible for the final step of L-methionine synthesis in *E. coli* in oxygen-deprived conditions[19]. Under normal aerobic conditions, this reaction is catalyzed by MetE, a B12-independent homocysteine-N5-methyltetrahydropteroyl-tri-l-glutamate transmethylase. It has been shown that MetE is an aggregation-prone enzyme under high temperature, acid, and oxidative stress conditions[20]. Termed as a "metabolic fuse," MetE's aggregation under stress conditions limits methionine synthesis, which leads to a domino effect in halting DNA, RNA, and protein synthesis that ultimately arrests growth. The C →T SNP mutation of *metH* in our ALE M1 mutant creates a missense mutation that converts amino acid residue 387, which is in the pferin binding domain of MetH, from an alanine to a valine. MetH may be similar to MetE in its aggregation potential and this SNP mutation may stabilize the binding of the N5-methyltetrahydrofolate cofactor for the methyl-group transfer to L-homocysteine in forming L-methionine during the micro-anaerobic ALE process, thus maintaining cellular functions while under IBA stress. Similar to single mutant Strains 1 and 3, the independent effect of the *metH* SNP in Strain 2 is not sufficient to confer resistance against sudden exposure to 3 g/L IBA. However, this mutation enabled production of 5.8 g/L IBA in Strain 13 when paired with the rho SNP, suggesting that these two mutations work synergistically to deregulate and stabilize mRNA transcription of cellular functions and the IBA plasmid-based production modules.

It is most perplexing that the combined effect of all three mutations helps Strain 7 confer comparable tolerance towards 3 g/L of IBA as the ALE M1 strain but does not help Strain 7 produce the same quantity of IBA as M1. Strain 7 performs much poorly in IBA production, making only 1.7 g/L of IBA in 24 h. The accumulation of 8 g/L of ISO at 24 and up to 10.1 g/L of ISO by 48 h may be disrupting cellular functions and membrane stability and is supported by the decrease in culture cell density from 24 to 48 h. There may be additional underlying mutations not identified from our current genome sequencing analysis that affects the balance amongst ISO production, accumulation, and conversion to IBA. Our parent strain AL17 also contains a 242,042 bp F' episome that was not included in our initial alignment reference sequence, so we will further determine if mutations are in this DNA region.

A.4 Methods

A.4.1 Reagents

All enzymes involved in the molecular cloning experiments were purchased from New England Biolabs (NEB). All synthetic oligonucleotides were synthesized by Integrated DNA Technologies. Sanger sequencing was provided by Genewiz. D-glucose, 99% isobutanol, 99% isobutyl acetate, and IPTG were purchased from Sigma-Aldrich.

A.4.2 Strains and plasmids

All strains used in this study are listed in Tables A.S1 and A.S2. All plasmids and primers are listed on Tables A.S3 and A.S4. Gene deletions and integrations were constructed using CRISPR-Cas9-mediated homologous recombination[21]. Linear DNA repair fragments for gene deletions were constructed by PCR assembly or amplification

from genomic DNA using primers listed in Tables A.S4 and A.S5. All genomic modifications were PCR and sequence verified.

Plasmids encoding sgRNAs for CRISPR-Cas9-mediated homologous recombination were constructed with Q5 site-directed mutagenesis using a modified template pTargetF (Addgene plasmid # 62226). Templates used for DNA amplification and cloning are listed in Table A.S6. All plasmids were verified by PCR and Sanger sequencing.

A.4.3 Culture conditions

Overnight cultures were grown at 37 °C, 250 rpm, in 3 mL of Luria-Bertani (LB) media with appropriate antibiotics. Antibiotic concentrations were as follows: tetracycline (25 μg/mL), ampicillin (200 μg/mL), and kanamycin (50 μg/mL). IBA tolerance assays and IBA production were carried out in M9 minimal medium (33.7 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 9.4 mM NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$) including 1000 × A5 trace metal mix (2.86 g H_3BO_3, 1.81 g $MnCl_2{\cdot}4H_2O$, 0.079 g $CuSO_4{\cdot}5H_2O$, 49.4 mg $Co(NO_3)_2{\cdot}6H_2O$ per liter water) supplemented with 5 g/L yeast extract.

A.4.4 IBA and ISO tolerance assays

Overnight cultures were inoculated to an OD_{600} ~0.1 in 5 mL of M9P supplemented with 5 g/L yeast extract, 10 g/L glucose, and appropriate concentrations of IBA in 10 mL screw cap tubes. To prevent evaporation of IBA or ISO, the tube caps were wrapped with parafilm. Cultures were grown at 37 °C 250 rpm for 24 h. Optical densities were measured at 600 nm (OD_{600}) with a Synergy H1 hybrid plate reader (BioTek Instruments, Inc.).

A.4.5 IBA production

Overnight cultures were inoculated to 1% in 25 mL of M9P supplemented with 5 g/L yeast extract and 50 g/L glucose in 250 mL screw cap flasks wrapped with parafilm. Cultures were grown at 37 °C at 250 rpm to an OD_{600} of 0.4 to 0.6 and ~ 5 mL of culture was removed to leave 20 mL for induction with 1mM IPTG. To prevent evaporation of produced ISO and IBA, the flask caps were wrapped with parafilm. Cultures were grown at 30 °C at 250 rpm for 48 h. Culture optical density was measured and supernatant was collected at 0, 24 and 48 h. ISO and IBA concentrations were measured using GC analysis.

A.5 Supplementary Information

Table A.S1. Strain List

Strain no.	*E. coli* strain	Plasmid	Key Genotype
-	AL17	-	*BW25113/F'[traD36, proAB$^+$, lacI^q ZΔM15] ΔadhE, ΔfrdBC, Δfnr, ΔldhA, Δpta, ΔpflB*
1	AL17	-	As AL17, with *rho* G → T
2	AL17	-	As AL17, with *metH* C → T
3	AL17	-	As AL17, with Δ*arcA-yjjY-arcA$_{P1-7}$*
4	AL17	**-**	As Strain 1, with *metH* C → T
5	AL17	-	As Strain 1, with Δ*arcA-yjjY-arcA$_{P1-7}$*
6	AL17	-	As Strain 2, with Δ*arcA-yjjY-arcA$_{P1-7}$*
7	AL17	-	As Strain 4 with Δ*arcA-yjjY-arcA$_{P1-7}$*
8	AL17	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*
9	M1	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*
10	**1**	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*
11	**2**	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*
12	**3**	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*
13	**4**	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*
14	**5**	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*
15	**6**	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*
16	**7**	pAL603, pAL1114	*P$_{LlacO1}$: alsS–ilvCD, P$_{LlacO1}$: kivd–adhA P$_{LlacO1}$: ATF1*

Table A.S2. Strains used in this study

Strain	Genotype	Source
AL17	*BW25113/F'[traD36, proAB$^+$, lacI^q ZΔM15] ΔadhE, ΔfrdBC, Δfnr, ΔldhA, Δpta, ΔpflB*	Atsumi et al. 2008
M1	As AL17 with the following identified mutations: *rho* G → T SNP, *metH* C → T SNP, *yjjY*[IS30]	This study

Table A.S3. Plasmids used in this study

Plasmid	Genotype
pCas	*P$_{cas}$:cas9 P$_{araB}$:Red lacI^q P$_{trc}$:sgRNA pMB1 repA101(Ts) kan^r*

pAL603	*P_{LlacO1}: alsS–ilvCD, P_{LlacO1}: kivd–adhA*
pAL1114	*P_{LlacO1}: ATF1*
pLA1851	*sgRNA-lacZ pMB1 amp^r*
pAL1859	*sgRNA-metH pMB1 amp^r*
pAL1860	*sgRNA-rho pMB1 amp^r*
pAL2091	*sgRNA-yjjY pMB1 amp^r*

Table A.S4. Oligonucleotides used in this study

Name	Sequence 5' → 3'
AZ349	GCGCGTCAACgttttagagctagaaatagc
AZ350	GACATCCAGCactagtattatacctaggac
AZ351	CCTACCTCGCgttttagagctagaaatagc
AZ352	AGCTGTCTGCactagtattatacctaggac
AZ918	CGGGTCCTGAgttttagagctagaaatagc
AZ919	ACAGTGTCAAactagtattatacctaggac
AZ932	ATTCATGGTACGGGACAGTAGGTTGC
AZ949	TCACTGCCGAAAATGAAAGCCAGTAAAGAAGTTACAACGGACGATG AGTTACGTATCT
AZ950	CGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTATTTCCAGATACGTA ACTCATCGTCCG
AZ951	ACGGCGCTAAAAAGCGCCGTTTTTTTTGACGGTGGTAAAGCCGACA GAAGGATATGT
AZ952	TTCCTGACTGTACTAACGGTTGAGTTGTTAAAAAATGCTACATATCC TTCTGTCGGCT
AZ953	ACCGTTAGTACAGTCAGGAAATAGTTTAGCCTTTTTTAAGCTAAGTA AAGGGCTTTTTCT
AZ954	GGTGCGAATTTACAAATTCTTAACGTAAGTCGCAGAAAAAGCCCTT TACTTAGCTTAAA
AZ1001	TGCAAGCGGTATTGAAAGGTTGGTGC
MMM211	GGAGCCGCTGAACATTGGCGAAGATAGCCTGTTTGTGAACGT
MMM212	TGAACTTAGCGGAACCGGTGACGTTGGTGCGTTCACCCACGTTCA CAAACAGGCTATCTT
MMM213	ACCGGTTCCGCTAAGTTCAAGCGCCTGATCAAAGAAGAGAAATACA GCGAGGCGCTGGAT
MMM214	ATATCGATAATCTGCGCGCCGTTTTCCACTTGTTGACGCACGACAT CCAGCGCCTCGCTG
MMM215	GGCGCGCAGATTATCGATATCAACATGGATGAAGGGATGCTCGAT GCCGAAGCGGCGATG
MMM216	GCGAGCGATATCCGGTTCACCGGCAATCAGATTGAGAAAACGCAC CATCGCCGCTTCGGC
MMM217	TGGGGCTGGAAAACCTGGCTCGTATGCGTAAGCAGGACATTATTTT TGCCATCCTG
MMM218	CACCAAAGATATCTTCGCCACTCTTTGCGTGCTGCTTCAGGATGGC AAAAATAATGTCCT

MMM219	GTGGCGAAGATATCTTTGGTGATGGCGTACTGGAGATATTGCAGG ATGGATTTTGTTTCC
MMM220	CATCAGGACCGGCCAGATAAGAGCTGTCTGCGGAACGGAGGAAAC AAAATCCATCCTGCA
MMM221	TCTGGCCGGTCCTGATGACATCTACGTTTCCCCTAGCCAAATCCGC CGTTTCAACCTCCG
MMM222	GGCGGGCGAATCTTACCAGAGATGGTATCACCAGTGCGGAGGTTG AAACGGCG
MMM241	CTGCGCCAGTACGTGCAGGAG
MMM242	GAAGATGATATCTTCTGGCG
MMM243	GCGAAGTGAACAGATTTCTG
MMM244	CCACGAAGACCTTTATTCAG

Table A.S5. Guide for CRISPR-Cas9-mediate gene deletions and insertions

	pTargetF		PCR Linear Repair Fragment
Modification	Plasmid	20 bp sgRNA sequence 5'→ 3'	Primers
rho G → T	pAL1860	GCAGACAGCTCCTACCTCGC	MMM217-222
metH C → T	pAL1859	GCTGGATGTCGCGCGTCAAC	MMM211-216
Δ*arcA-yjjY-arcA*$_{P1-7}$	pAL2091	TTGACACTGTCGGGTCCTGA	AZ949-954

Table A.S6. Plasmid construction guide

	PCR for Vector		
Plasmid	Primer (F)	Primer (R)	Template
pAL1859*	AZ349	AZ350	pAL1851
pAL1860*	AZ351	AZ352	pAL1851
pAL2091*	AZ918	AZ919	pAL1851

*Q5-site directed mutagenesis (NEB).

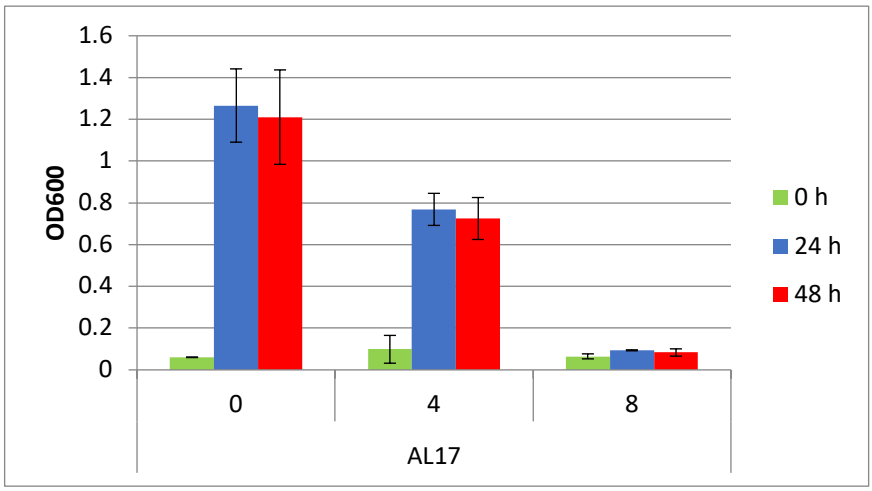


Figure A.S1. Isobutanol tolerance of AL17. AL17 was inoculated to an OD_{600} 0.1 in M9 production media with increasing concentrations of ISO (0, 4, and 8 g/L, respectively). Cultures were grown at 37 ºC for 24 h and culture density was measured at OD_{600}. Error bars indicate s.d. (n = 3 biological replicates).

A.6 References

1. Förster, A. H. & Gescher, J. Metabolic Engineering of Escherichia coli for Production of Mixed-Acid Fermentation End Products . *Frontiers in Bioengineering and Biotechnology* vol. 2 16 (2014).
2. Dong, H. *et al.* Engineering Escherichia coli Cell Factories for n-Butanol Production. *Adv. Biochem. Eng. Biotechnol.* **155**, 141–163 (2016).
3. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).
4. Bohnenkamp, A. C. *et al.* Multilevel optimisation of anaerobic ethyl acetate production in engineered Escherichia coli. *Biotechnol. Biofuels* **13**, 65 (2020).
5. Horton, C. E. & Bennett, G. N. Ester production in E. coli and C. acetobutylicum. *Enzyme Microb. Technol.* **38**, 937–943 (2006).
6. Rodriguez, G. M., Tashiro, Y. & Atsumi, S. Expanding ester biosynthesis in Escherichia coli. *Nat. Chem. Biol.* **10**, 259–265 (2014).
7. Fischer, C. R., Klein-Marcuschamer, D. & Stephanopoulos, G. Selection and optimization of microbial hosts for biofuels production. *Metab. Eng.* **10**, 295–304 (2008).
8. Adegboye, M. F., Ojuederie, O. B., Talia, P. M. & Babalola, O. O. Bioprospecting of microbial strains for biofuel production: metabolic engineering, applications, and challenges. *Biotechnol. Biofuels* **14**, 5 (2021).
9. Lin, H., Wang, Q., Shen, Q., Zhan, J. & Zhao, Y. Genetic engineering of microorganisms for biodiesel production. *Bioengineered* **4**, 292–304 (2013).
10. Seo, H., Lee, J.-W., Garcia, S. & Trinh, C. T. Single mutation at a highly conserved region of chloramphenicol acetyltransferase enables isobutyl acetate

production directly from cellulose by Clostridium thermocellum at elevated temperatures. *Biotechnol. Biofuels* **12**, 245 (2019).
11. Peralta-Yahya, P. P. & Keasling, J. D. Advanced biofuel production in microbes. *Biotechnol. J.* **5**, 147–162 (2010).
12. Madhavan, A. *et al.* Synthetic Biology and Metabolic Engineering Approaches and Its Impact on Non-Conventional Yeast and Biofuel Production . *Frontiers in Energy Research* vol. 5 8 (2017).
13. Mukhopadhyay, A. Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. *Trends Microbiol.* **23**, 498–508 (2015).
14. Tashiro, Y., Desai, S. H. & Atsumi, S. Two-dimensional isobutyl acetate production pathways to improve carbon yield. *Nat. Commun.* **6**, 7488 (2015).
15. Haft, R. J. F. *et al.* Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. *Proc. Natl. Acad. Sci.* **111**, E2576 LP-E2585 (2014).
16. Cao, H. *et al.* Systems-level understanding of ethanol-induced stresses and adaptation in E. coli. *Sci. Rep.* **7**, 44150 (2017).
17. Brynildsen, M. P. & Liao, J. C. An integrated network approach identifies the isobutanol response network of Escherichia coli. *Mol. Syst. Biol.* **5**, 277 (2009).
18. Alexeeva, S., Hellingwerf, K. J. & Teixeira de Mattos, M. J. Requirement of ArcA for redox regulation in Escherichia coli under microaerobic but not anaerobic or aerobic conditions. *J. Bacteriol.* **185**, 204–209 (2003).
19. Cai, X. Y. *et al.* Methionine synthesis in Escherichia coli: effect of the MetR protein on metE and metH expression. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 4407–4411 (1989).
20. Mordukhova, E. A. & Pan, J.-G. Evolved cobalamin-independent methionine synthase (MetE) improves the acetate and thermal tolerance of Escherichia coli. *Appl. Environ. Microbiol.* **79**, 7905–7915 (2013).
21. Jiang, Y. *et al.* Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).