UNIVERSIDADE FEDERAL DE SÃO CARLOS
DEPARTAMENTO DE ENGENHARIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

GABRIEL LUZ CHAVES

RATIONAL DESIGN AND ENGINEERING OF A MICROBIAL CELL FACTORY FOR 3-HYDROXYPROPIONIC ACID PRODUCTION

SÃO CARLOS - SP
2021

GABRIEL LUZ CHAVES

RATIONAL DESIGN AND ENGINEERING OF A MICROBIAL CELL FACTORY FOR 3-HYDROXYPROPIONIC ACID PRODUCTION

"*Desenho e construção racional de uma fábrica celular para a produção de ácido 3-hidroxipropiônico*"

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de São Carlos, como parte dos requisitos necessários para obtenção do título de Mestre em Engenharia Química.

Orientador: Prof. Adilson José da Silva (DEQ-UFSCar)

SÃO CARLOS - SP

2021

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Engenharia Química

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Gabriel Luz Chaves, realizada em 20/04/2021.

Comissão Julgadora:

Prof. Dr. Adilson Jose da Silva (UFSCar)

Profa. Dra. Danielle Biscaro Pedrolli (UNESP)

Prof. Dr. Flavio Henrique da Silva (UFSCar)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Engenharia Química.

Para as minhas sobrinhas, Maria Antônia e Helena, dedico este trabalho.
Que a Educação seja transformadora na vida de vocês, como foi e ainda é na minha.
Também dedico à minha mãe, Rosiani, e à memória de meu pai, Geraldo,
que sempre deram tudo de si pra que eu corresse atrás das melhores oportunidades.
Tudo que eu faço é por vocês e por causa de vocês.

Tudo o que fizerem,
seja em palavra seja em ação,
façam-no em nome do Senhor Jesus,
dando por meio dele graças a Deus Pai.
(Colossensses, 3:17)

AGRADECIMENTOS

Agradeço a Deus pela graça e misericórdia da vida.

À minha família, Rosiani, Thais, Lucas, Vinícius, Helena e Maria Antônia, por todo apoio. Foram fundamentais nos momentos mais difíceis e mais felizes deste período, e de toda a vida. Amo tanto vocês. Mãe, sinto muito sempre ir pra longe da senhora e as vezes não poder estar presente quando precisa, mas preciso correr atrás dos meus sonhos. Helena e Maria, vocês me inspiram. Quero que vocês cresçam com as melhores oportunidades que puderem e quero ser parte disso.

A todos meus amigos, que também me deram suporte. Àqueles que carrego desde Muriaé, os que trago de Viçosa e os que ganhei aqui, Ana Carolina, Bruna, Ana Luísa, Letícia e Maria Carolina. E em especial, Raíssa, que topou dividir comigo a experiência da convivência e a construção de um lar em São Carlos, e Lorenna (mais que amiga, hóspede), que sempre esteve com a gente com o eterno mantra "vai dar certo". Paula, Débora, Raquel, Fernanda, Jéssica, Daniela e Nathan (mais que amigo, revisor), me sinto tão feliz e abençoado por sentir que posso contar com vocês, mesmo de longe.

Aos colegas do LaFaC, Daniel, Davi, Rafael, Bruno, por toda ajuda. Além de Raquel, que tanto me ajudou no fim dos experimentos e me acompanhou em tanta correria e fins de semana de trabalho pesado, e Josivan, que compartilhou comigo e também me ajudou com tanta coisa, muito obrigado. Esse trabalho também é obra de vocês. Ninguém faz nada sozinho.

Ao grupo de cultivos, por todo suporte quando necessário.

Ao meu orientador, Prof. Adilson, pela paciência e pela disponibilidade em me atender sempre que precisava, além de todo o ensinamento e apoio em minhas ideias. Obrigado pela confiança depositada.

À FAPESP e ao CNPq, pelo apoio financeiro (Processo FAPESP 2019/07902-6 e CNPq 132794/2019-5).

Ao Prof. Uelinton Manoel Pinto, da Faculdade de Ciências Farmacêuticas da USP pela doação da cepa de *Pseudomonas aeruginosa* PA01 e ao Prof. Anderson Ferreira da Cunha, do Departamento de Genética e Evolução da UFSCar pela doação do DNA genômico de *Saccharomyces cerevisiae* CAT-1.

Sou muito grato a todos que me ajudaram desenvolver este trabalho e concluir essa etapa: técnicos de laboratório, funcionários e servidores da UFSCar.

ABSTRACT

The US Department of Energy ranked the 3-hydroxypropionic acid (3-HP) among the Top 10 most promising value-added chemicals that can be derived from biomass in a biorefinery. This acid has great potential to serve as a building block for the industry, serving as raw material for paints, coatings, and polymers. Nevertheless, its production from chemical routes encompasses processes that are highly toxic and environmentally harmful. In this context, fermentative bioprocesses are a promising alternative for 3-HP production. Here, we proposed the obtention of an *Escherichia coli* strain genetically modified to produce 3-HP through the β-alanine pathway, yet poorly studied. To build an *E. coli* strain able to produce 3-HP, the genes that encode the last three reactions of the pathway were cloned and combined in the same plasmid (pEbtyGpD): *pa0132* from *Pseudomonas aeruginosa*, *ydfG* from *E. coli*, and *panD* from *Corynebacterium glutamicum*. This first engineered strain, named PS100, produced up to 0.338 ± 0.044 g/L of 3-HP after 24 h of induction with IPTG using glucose as a carbon source. Surprisingly, cultivations on a mixture of glucose and xylose (1:1 on C-mol basis) yielded a final titer of 1.040 ± 0.050 g/L by this strain, from the same substrate amount. To optimize the production obtained from PS100, new genetic modifications were investigated through *in silico* optimization of a genome-scale metabolic model of *E. coli* K-12 MG1655. The model *i*ML1515 was modified to include the heterologous reaction of β-alanine conversion to malonic semialdehyde, and three reactions were identified as potential metabolic targets for enhancing 3-HP production by *E. coli* cells: the reactions of alanine racemase (ALAR), L-alanine aminotransferase (ALAT), and L-valine transaminase (VALTA). These target reactions were modified in the strain PS100, generating the strain PSO107 that was able to improve nearly 2-fold the final titer of the acid when compared to PS100 in cultivations with glucose as carbon source, reaching 0.743 ± 0.016 g/L of 3-HP. For cultivations with the glucose:xylose mixture, a 10% increment was observed for the PSO107 strain compared to PS100, with the former reaching 1.147 ± 0.015 g/L of 3-HP. These results confirm that the targets predicted by the evolutionary optimizations of the genome-scale metabolic model were assertive and enabled an increment in the final production of 3-HP by the engineered *E. coli* strain.

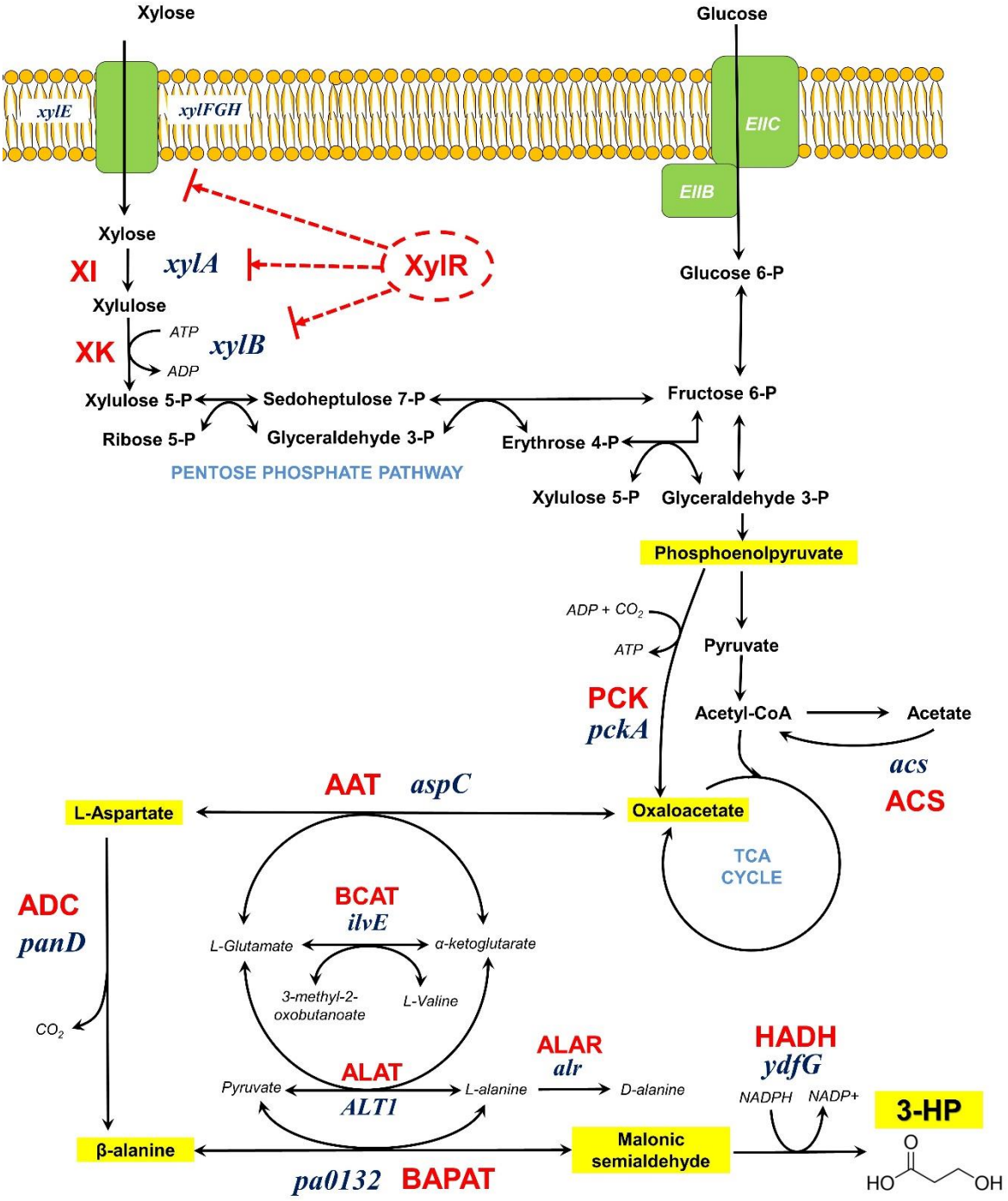
Key-words: metabolic engineering, genome-scale metabolic model, 3-hydroxypropionic acid, β-alanine, biorefinery, microbial cell factory.

RESUMO

O Departamento de Energia dos Estados Unidos (DOE, US) listou, em 2004, o ácido 3-hidroxipropiônico (3-HP) entre os 10 mais promissores químicos de alto valor-agregado que podem ser obtidos a partir da biomassa em biorrefinarias. Este ácido tem potencial de servir como bloco construtor para a indústria, como a de tintas, de revestimentos e polímeros. Entretanto, sua produção a partir de rotas químicas envolve processos altamente tóxicos e prejudiciais ao meio ambiente. Assim, bioprocessos fermentativos são uma alternativa promissora para a produção do 3-HP. Neste trabalho, propôs-se a obtenção de uma linhagem de *Escherichia coli* geneticamente modificada para produzir 3-HP pela via da β-alanina, ainda pouco explorada na literatura. Para produzir uma linhagem capaz de produzir o ácido, os genes que codificam as três últimas reações da via foram clonados e combinados num mesmo plasmídeo (pEbtyGpD): *pa0132* de *Pseudomonas aeruginosa*, *ydfG* de *Escherichia coli* e *panD* de *Corynebacterium glutamicum*. A primeira linhagem construída, PS100, foi capaz de produzir até 0,338 ± 0.044 g/L de 3-HP após 24 h de indução com IPTG, usando glicose como fonte de carbono. Surpreendentemente, em cultivos com uma mistura glicose : xilose (1:1 em base C-mol) a produção final de 3-HP por essa linhagem aumentou até 1,040 ± 0,050 g/L partindo-se da mesma concentração inicial de substrato. Para otimizar a produção obtida pela PS100, buscou-se pela predição *in silico* de novos alvos não-intuitivos através de otimização evolucionária de um modelo metabólico em escala genômica de *E. coli* K-12 MG1655. O modelo *i*ML1515 foi modificado para passar a conter a reação heteróloga de conversão da β-alanina a semialdeído malônico, e três reações foram identificadas como alvos metabólicos potenciais para incrementar a produção de 3-HP por células de *E. coli*: a reação da alanina racemase (ALAR), da L-alanina aminotransferase (ALAT) e da L-valina transaminase (VALTA). Essas modificações foram introduzidas na linhagem PS100, gerando a linhagem PSO107 que foi capaz de aproximadamente dobrar a concentração final do ácido, em relação à PS100, em cultivos com glicose como fonte de carbono, produzindo 0,743 ± 0,016 g/L de 3-HP. Para cultivos com a mistura glicose : xilose (1:1 em base C-mol), um aumento de 10% n concentração do ácido foi observado para a PSO107, comparado à PS100, chegando a 1,147 ± 0,015 g/L de 3-HP. Estes resultados confirmam que os alvos metabólicos preditos pelas otimizações do modelo metabólico em escala genômica foram assertivos e possibilitaram um incremento final na produção de 3-HP pelas linhagens produtoras construídas de *E. coli*.

Palavras-chave: engenharia metabólica, modelo metabólico em escala genômica, ácido 3-hidroxipropiônico, β-alanina, biorrefinarias, fábricas celulares.

GRAPHICAL ABSTRACT



Enzymes and the XylR regulator are shown in red and their respective genes in blue. Key metabolites from the β-alanine pathway for 3-hydroxypropionic acid (3-HP) production are highlighted in yellow.

LIST OF ENZYMES/GENE ABBREVIATIONS AND ASSOCIATIONS[1]

XylR	DNA-binding transcriptional activator of *xyl* operon	*xylR*
-	D-xylose proton symporter	*xylE*
-	D-xylose transporter	*xylFGH*
XI	Xylose isomerase	*xylA*
XK	Xylulokinase	*xylB*
PCK	Phosphoenolpyruvate carboxykinase	*pckA*
AAT	L-Aspartate aminotransferase	*aspC*
ADC	L-Aspartate 1-decarboxylase	*panD*
BAPAT	β-Alanine:pyruvate transaminase	*pa0132*
HADH	3-Hydroxyacid dehydrogenase	*ydfG*
BCAT	Branched-chain amino acid aminotransferase	*ilvE*
ALAT	Alanine aminotransferase	*ALT1*
ALAR	Alanine racemase	*alr*

[1] Relative to the most relevant reactions depicted in the Graphical Abstract.

SUMMARY

TABLES LIST

FIGURES LIST

1 INTRODUCTION

Biotechnological routes for chemical compounds production have gained more attention due to environmental problems associated with the petrochemical industry and the increasing demand for more sustainable production processes. The biorefinery concept, in which biomass is transformed into valuable products and energy minimizing residues and greenhouse gas emissions, has emerged in this context. Nowadays, biorefineries are presented as one of the best alternatives to settle an industrial unit capable of producing a range of compounds from renewable raw materials like sugars present in lignocellulosic biomass and glycerol from the biofuel industry (ADOM *et al.*, 2014).

The US Department of Energy reported the 3-hydroxypropionic acid (3-HP) in 2004 and 2010 among the top 10 compounds that could be obtained from biomass to serve as a building block for the chemical industry (WERPY AND PETERSEN, 2004; KUMAR, ASHOK AND PARK, 2013). However, for 3-HP and many other compounds, some hurdles remain to establish bioprocesses that are economically feasible and competitive to traditional ones, such as product titer, yield, and productivity. These parameters directly impact the product minimum selling price and the installed capacity of biorefineries and, often, hinder the establishment of these bioprocesses on a larger scale. Thus, it is necessary to develop an optimized bioprocess that overcomes these issues (JIANG; MENG; XIAN, 2009).

Regarding the fermentative production of 3-hydroxypropionic acid, several pathways from glucose and glycerol have been reported, and BASF and Cargill have demonstrated an industrial process at pilot-scale to produce bio-based 3-HP and acrylic acid. The most studied pathways are the malonyl-CoA and glycerol pathways, with the latter reaching higher titers so far (KUMAR, ASHOK AND PARK, 2013; ZHAO *et al.*, 2019). However, the glycerol pathway depends on continuous supplementation of vitamin B12, which can be an economic bottleneck for its implementation. Malonyl-CoA pathway also has great potential of being implemented in bioprocesses for 3-HP industrial production, and two patents have already been deposited regarding an *E. coli* strain capable of producing large amounts of 3-HP using this route (LYNCH, LIPSOCOMB AND GILL, 2013). Nonetheless, another pathway that produces 3-HP from β-alanine, yet poorly studied in the literature, has been predicted by *in silico* simulations to have an even greater potential than the other two aforementioned (BORODINA *et al.*, 2015). Therefore, this pathway is the study object in this master thesis that intends to construct and optimize an engineered *Escherichia coli* strain for 3-HP production.

The β-alanine pathway was already assessed in *Escherichia coli* (SONG *et al.*, 2016) and *Saccharomyces cerevisiae* (BORODINA *et al.*, 2015) but still needs to be optimized to fulfill industrial requirements. This optimization process depends on protein engineering, enzyme overexpression, and carbon flux redirection to form the desired product, besides product tolerance improvement.

Recent advances in biotechnology, especially in systems biology and metabolic engineering, have boosted the design and construction of new engineered microorganisms for industrial fermentations by traditional hosts (e. g., *Escherichia coli, Saccharomyces cerevisiae,* and *Corynebacterium glutamicum*). Often, these so-called cell factories present several advantages compared to chemical and enzymatic processes in which the capacity of a single-step process can be highlighted (CHEN AND NIELSEN, 2016). Moreover, bioinformatics tools provide a powerful approach to understanding the microorganism's metabolic network and rapidly identifying non-intuitive targets for improving product formation (KLAMT; HADICKE; VON KAMP, 2014). Thus, these computational analyses of cell biocatalysts have been successful in developing processes that will be ready to scale up in the upcoming years.

In this work, we aim to continue a previous work performed by Lorrany Rocha in our research group in the Postgraduate Program of Chemical Engineering – PPGEQ, UFSCar – that began the *in silico* assessment of 3-HP production in *E. coli*. Here, we planned to rationally construct an efficient *E. coli* strain for 3-hydroxypropionic acid production through the β-alanine pathway, based on *in silico* optimization combined with the most recent developments reported for this end.

The specific goals of this work are: *i.* assess the 3-HP pathway *in silico* by using a genome-scale metabolic model of *E. coli* and identify new metabolic targets for pathway optimization, *ii.* construct the heterologous route for 3-hydroxypropionic acid production in *E. coli* by overexpression of pathway enzymes and, finally, *iii.* optimize the pathway by overexpressing or deleting target reactions predicted by *in silico* analysis and literature review.

This work is part of the FAPESP Thematic Project "*Da fábrica celular à biorrefinaria integrada biodiesel-bioetanol: uma abordagem sistêmica aplicada a problemas complexos em micro e macroescalas*" coordinated by Prof. Roberto Giordano (DEQ/UFSCar).

2 LITERATURE REVIEW

2.1 Bioengineering applied to the context of biorefineries

Since the major petroleum crisis and the growing concerns about environmental problems, society has increasingly searched for alternative energy sources to minimize environmental damage. In this context, the bio-based economy arises as one of the most promising models for sustainable development (ADOM *et al.*, 2014).

Energy production is mainly derived from non-renewable sources, such as petroleum, natural gas, and charcoal, which are the main responsible for greenhouse gas emissions. In order to minimize our dependency on these sources, alternatives have been developed. Among them, biofuels are now a consolidated renewable source of energy.

Biofuels are produced from agricultural products such as soybean, corn, and sugarcane. In this case, they are known as first-generation biofuels. However, their production competes for farmlands meant for food production. An alternative is the employment of agricultural wastes, forestry residues, and energy crops, composed of lignocellulosic biomass, as feedstocks to produce second-generation biofuels. However, biofuels obtained from lignocellulosic biomass are not economically feasible yet. Integrated generation of other products with high added value can help circumvent this problem in the biorefinery context (YAO AND SHIMIZU, 2013).

In 2004, the US Department of Energy published a list of top 12 platform chemical products that could be obtained from biomass, either by microbial fermentation or chemical processes, and then converted to more relevant products to the industry. These compounds were chosen based on the involved process, economic and industrial feasibility, market size, and their capacity as a platform to further conversion into other products (WERPY AND PETERSEN, 2004).

Standing at third place on the US DOE list is 3-hydroxypropionic acid, the object of this study. Organic acids can be treated like essential commodities to the chemical industry since they can be used to synthesize many other chemical products. Therefore, they serve as building blocks for the chemical industry. Organic acids with hydroxyl or keto groups are even more useful due to their bifunctionality, making them more attractive to chemical transformations, especially for the polymer industry (BECKER *et al.*, 2015).

Since DOE's publication, many efforts have been made to develop bio-based processes to produce these compounds in high titer, yield, and productivity, required by the industry (VAN DIEN, 2013). Gunukula *et al.* (2017) studied the economic feasibility of some processes.

They stated that among these three parameters, the one with the major effect on the minimum selling price of bio-based chemical commodities is the yield. This parameter is closely related to the upstream processes' costs, like production and preparation of the fermentable feedstock and fermentation operational costs. Hence, special attention should be given again to second-generation feedstocks, considering that it represents a byproduct of human activity and might be cheaper than others. Moreover, downstream processes (recovery and purification) are also key aspects in making a bioprocess economically feasible. Therefore, developing microbial strains capable of high performance to titer, productivity, and yield is crucial, as much as the usage of non-edible feedstocks like lignocellulosic biomass (CHEN AND NIELSEN, 2016). Table 1 summarizes how these three parameters impact process costs and the challenges related to each of them.

Metabolic engineering integrated with systems biology, synthetic biology, and evolutionary engineering lead to an increase in developing strains highly efficient in the production of a given compound and bioprocess optimization in an integrated manner, considering upstream, fermentation, and downstream stages (CHOI *et al.*, 2015; LEE *et al.*, 2019), thus making bio-based economy increasingly advantageous (WESTBROOK *et al.*, 2019).

However, some hurdles remain for highly efficient production performance, such as inhibitors present in the lignocellulosic biomass hydrolysate, efficient utilization of pentose fraction (hemicelluloses), pH decrease due to organic acid build-up, water use/reuse strategies, thermodynamics viability, enzyme availability for a pathway, and intermediates toxicity (CHEN AND NIELSEN, 2016; CHOI *et al.*, 2015; LEE *et al.*, 2019; VALDEHUESA *et al.*, 2013).

2.2 3-Hydroxypropionic acid (3-HP)

3-hydroxypropionic acid (3-HP, pKa = 4.5) is a structural isomer from lactic acid and, due to its bifunctionality (carboxyl and hydroxyl groups), can be used for multiple chemical reactions, having a great potential for transformations into useful chemicals. The acid group can be reduced to alcohol or converted to various amides, esters, and derivatives. The alcohol group can, for example, be dehydrated into unsaturated compounds. Moreover, it can also be transformed into polyesters, oligomers, and other polymers (CORMA CANOS; IBORRA; VELTY, 2007).

Table 1 - Summary of process parameters impacts and challenges. Adapted from Van Dien (2013).

Parameter	Impact	Challenges	Tools and Methods
Titer (> 50 g/L)	Downstream separation costs	Product toxicity	Adaptive evolution with sequencing and transcriptomics Genome-wide methods Efflux pumps
		Enzyme inhibition	Protein engineering
Yield (g/g) (> 80 % of the theoretical yield)	Substrate cost	Low pathway flux	Fermentation optimization All approaches for productivity (below)
		Byproducts, energetics, and NAD(P)H balance	Metabolic modeling, fluxomics, metabolomics, transcriptomics Fermentation optimization Protein engineering
Productivity (> 2 g/L-h)	Bioreactor size	Bottleneck enzymes	Protein engineering Protein scaffolding
		Expression imbalance	Combinatorial expression testing Promoter and RBS libraries Transcriptomics and proteomics

The polymer obtained from 3-HP - poly-(3-HP) - and others in which 3-HP is employed as comonomer present up-and-coming features, and they might be used for further substitution of petroleum-based polymers (DE FOUCHÉCOUR *et al.*, 2018). Especially, poly-(3-HP) has an excellent glass transition temperature and a melting point over 170 °C, besides being biodegradable and biocompatible, which make it suitable for new applications like drug release materials and biocomposites for surgery (JIANG; MENG; XIAN, 2009).

The production of bio-based chemical products that might be promising for substituting petroleum-based derivatives is essential for sustainability and reducing fossil energy dependence. A variety of chemical products, including acrylamide, ethyl 3-HP, 1,3-PDO, malonic acid, and propiolactone, can be obtained from 3-HP (Figure 1). Hydrogenation of 3-HP results in 1,3-propanediol (1,3-PDO), one of the monomers used for the poly(trimethylene terephthalate) (PTT) production, a valuable polymer for the manufacturing of fibers and resins. Dehydration of 3-HP results in acrylic acid, which demand was expected to reach 8 million tons worldwide in 2020 with a market size estimated at more than 18 billion US dollars. 3-HP can also be employed as a platform for 3-carbon intermediates production, which are nowadays produced from petrochemicals (CHEN AND NIELSEN, 2016; CORMA CANOS; IBORRA; VELTY, 2007).

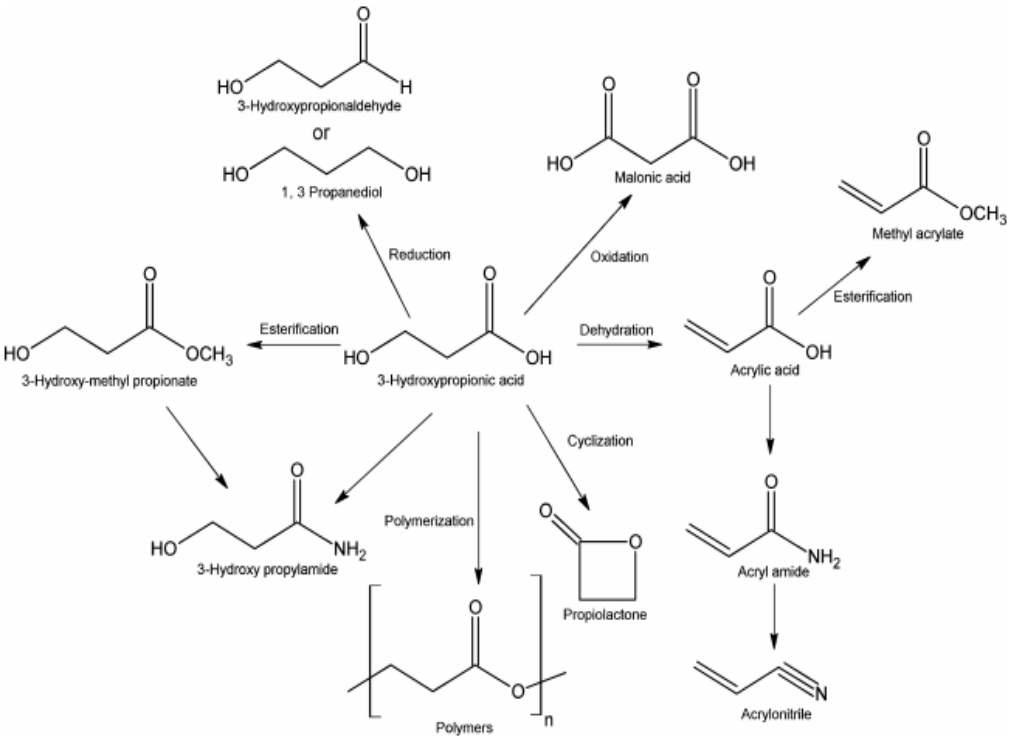


Figure 1 - Potential routes for converting 3-hydroxypropionic acid into valuable chemicals for industry (PINA; FALLETTA; ROSSI, 2011).

Adom *et al.* (2014) developed process models to simulate 1,3-PDO and acrylic acid production from 3-HP obtained from glycerol and achieved a 39 % and 55 % reduction of greenhouse gas emissions, respectively. Compared to its fossil homologous, there was a 24 % and 58 % reduction in fossil fuel consumption, respectively.

The chemical routes proposed for the production of 3-HP, such as from β-propiolactone, acrylic acid, and β-hydroxypropionitrile, are not commercially viable for presenting expensive starting materials and/or for having environmental incompatibility (e.g., carcinogenic, toxic, and noxious products involved). Biotechnological production of 3-HP has the advantage of using renewable substrates like sugars and glycerol. However, it depends on developing genetically modified microorganism strains to be used in fermentation processes and identifying favorable metabolic pathways (DELLA PINA; FALLETTA; ROSSI, 2011; JIANG; MENG; XIAN, 2009).

In conclusion, 3-HP has a potential market, and chemical companies like BASF (Germany), Cargill (United States), Dow Chemical (United States), Novozymes (Denmark), and Perstorp (Sweden) are already developing bioprocesses for its production, aiming to produce acrylic acid from different renewable raw materials such as sugars and glycerol (CHEN AND NIELSEN, 2016; CHOI *et al.*, 2015).

2.3 Metabolic pathways for 3-HP production

Some microorganisms such as *Lactobacillus* spp., *Chloroflexus aurantiacus*, *Metallosphaera sedula*, *Saccharomyces kluyveri*, etc., are natural producers of 3-HP as an intermediary metabolite or final product, through pathways like glycerol to 3-HP, acrylic acid to 3-HP, uracil catabolism, and 3-hydroxyisobutyrate/4-hydroxybutyrate cycle. Nevertheless, none of them is known to achieve industrial requirements; thus, genetically modified organisms are necessary to establish a process for 3-HP bioproduction (KUMAR; ASHOK; PARK, 2013).

Compared with chemical routes, biochemical pathways from renewable feedstocks present some advantages, such as dependency relief from petroleum, higher energetic safety, less greenhouse gas emission and environmental pollution, and mild operation conditions (JIANG; MENG; XIAN, 2009). Therefore, it is desirable to engineer a biochemical pathway for 3-HP production from renewable feedstocks, such as lignocellulosic biomass, and insert it into well-known cell factories, like *E. coli* and *S. cerevisiae*, in order to achieve industry and market demand and minimize environmental harm.

Such a pathway must be redox balanced and ideally generate ATP for cell growth, maintenance, and product export. Moreover, the pathway must also be thermodynamically favorable. After a pathway is proposed, some aspects must be considered regarding its efficiency, like enzyme kinetics, non-native enzyme activities in recombinant hosts,

competitive pathways, and the recombinant host's genetic stability under fermentation conditions (JIANG; MENG; XIAN, 2009).

Kumar, Ashok, and Park (2013) presented several biological pathways for 3-HP production from glucose and glycerol in their review. Among them, there are seven pathways patented by US agricultural-based company Cargill, in which only 3 of them were thermodynamically favorable: pathway I (pyruvate → acetyl-CoA → malonyl-CoA → 3-oxopropanoate → 3-HP), pathway II (pyruvate/phosphoenolpyruvate → oxaloacetate → aspartate → β-alanine → 3-oxopropanoate → 3-HP), and pathway III (pyruvate → α-alanine → β-alanine → 3-oxopropanoate → 3-HP) (Figure 2).

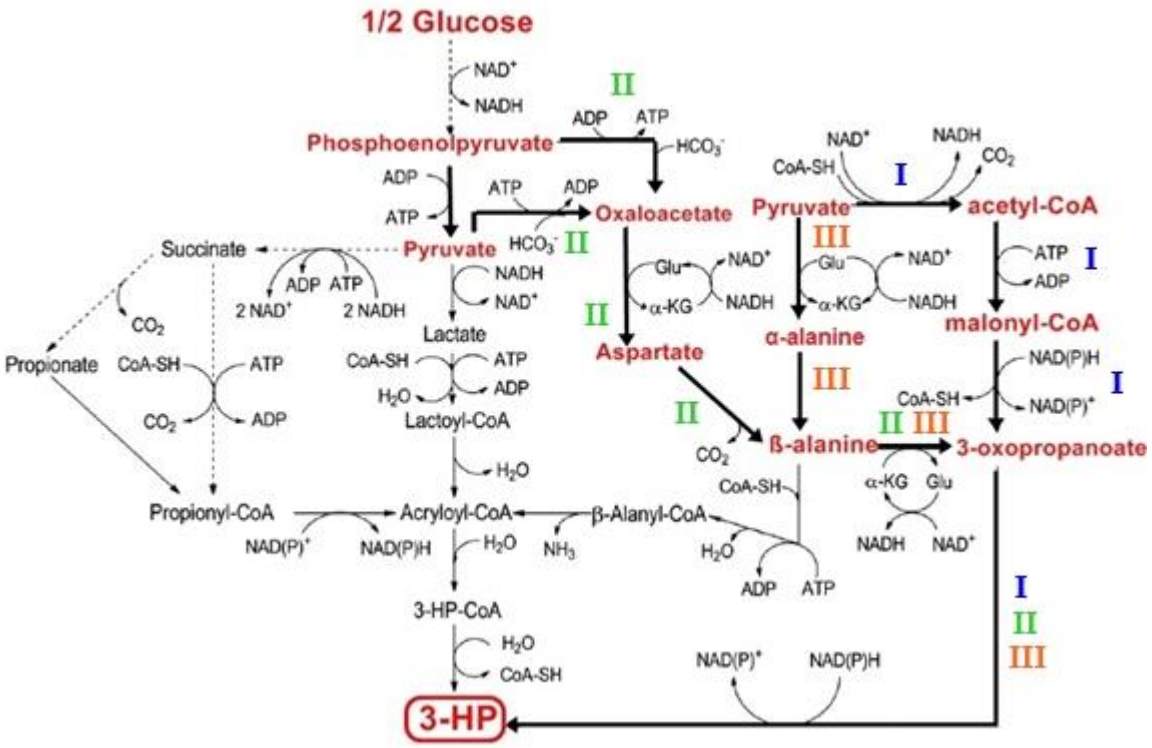


Figure 2 - Metabolic pathways for 3-HP production from sugars. Adapted from Kumar, Ashok, and Park (2013). Pathway I (blue), pathway II (green) and pathway III (orange).

There is no net production of ATP in pathway I, but the main drawback is the inability of malonyl-CoA reductase, responsible for the last two reactions, in utilizing NADH as a cofactor. However, to overcome this issue, a mutation in the enzyme may be done, or a NAD(P) transhydrogenase may be introduced into the host to keep the balance between NADPH and NADH (JIANG; MENG; XIAN, 2009).

In pathway II, pyruvate or phosphoenolpyruvate (PEP) is converted into oxaloacetate (OAA) by carboxylation. PEP carboxylation by PEP carboxykinase produces one mole of ATP for every mole of OAA, whereas the same reaction does not produce ATP when catalyzed by PEP carboxylase. The subsequent reactions do not produce or consume ATP. Consequently, ATP net production in this pathway depends on the intermediate (PEP or pyruvate) and the employed enzyme. This pathway's redox balance depends on the enzyme that catalyzes 3-oxopropanoate reduction to 3-HP since both NADPH and NADH can be used as a cofactor.

In pathway III, energy-consuming steps are circumvented using the enzyme alanine 2,3-aminomutase, which catalyzes amino group transference from alpha to beta carbon. Therefore, ATP net production in this pathway is one mole for every half mole of glucose consumed. The only reported work for this pathway is a patent that utilized a mutated version of lysine 2,3-aminomutase to convert α-alanine into β-alanine (LIAO *et al.*, 2007; LIAO *et al.*, 2012). However, this enzyme presented low catalytic efficiency in promoting this conversion.

Another intensively studied route to produce 3-HP is the glycerol pathway (Figure 3), carried out in only two steps. This pathway presents some problems when expressed in *E. coli* and *S. cerevisiae*, such as the necessity of continuous medium supplementation with vitamin B_{12}, oxygen sensitivity by pathway enzymes, and others. However, this pathway holds the best 3-HP titer reported so far, to the best of our knowledge: 102.61 g/L 3-HP in 98 h in fed-batch cultivation. This result was obtained from Zhao *et al*. (2019) with a *Klebsiella pneumoniae* strain (a natural producer of vitamin B_{12}) overexpressing its aldehyde dehydrogenase (*puuC* gene) and using a three tandem repetitive *tac* promoter in order to trap RNAP and enhance gene expression.

Pathways I (malonyl-CoA pathway) and II (β-alanine pathway) are the most promising pathways for 3-HP production from sugars, and then, in the next section, efforts to establish and improve these pathways will be reviewed. They will be named malonyl-CoA and β-alanine pathways, respectively. One will notice that the malonyl-CoA pathway was more intensively studied than β-alanine so far, possibly because it can be more easily engineered by expressing just one or two enzymes. Table 2 summarizes some of the most recent works of 3-HP production in different hosts and using Pathways I and II, besides the glycerol pathway.

Table 2 – Production of 3-HP in different hosts and using different pathways.

Pathway	Microorganism	Carbon source	Final titer [g/L] – Fed-batch	Final titer [g/L] – Batch (shake flasks)	Reference
malonyl-CoA	*E. coli*	glucose	-	0.2	Rathnasingh *et al.*, 2012
	S. cerevisiae	glucose	-	0.463	Chen *et al.*, 2014
	S. cerevisiae	glucose	9.83	-	Kildegaard *et al.*, 2016
	Schizosaccharomyces pombe	glucose glucose	-	3.5 7.6 (high cell density)	Suyama *et al.*, 2017a
	E. coli	glucose	40.6	3.72	Liu *et al.*, 2016
	S. pombe	cellobiose	11.2	3.5	Takayama *et al.*, 2018
	E. coli	glucose	30	-	Liu *et al.*, 2018a
β-alanine	*S. cerevisiae*	glucose	13.7	1.27	Borodina *et al.*, 2015
	E. coli	glucose	31.1	3.69	Song *et al.*, 2016
	S. cerevisiae	xylose	7.37	1.84	Kildegaard *et al.*, 2015
glycerol	*C. glutamicum*	glucose glucose and xylose	62 54.8	37.4 36.2	Chen *et al.*, 2017
	E. coli	glucose and xylose	29.7	0.97	Jung *et al.*, 2015
	E. coli	glucose and xylose	37.6	1.44	Heo *et al.*, 2019
	Klebsiella pneumoniae	glycerol	102.61	-	Zhao *et al.*, 2019
	Pseudomonas denitrificans	glucose and glycerol	-	4.77	Park *et al.*, 2019
	Lactobacillus reuteri	glycerol	-	8.59	Zabed *et al.*, 2019
	E. coli	glycerol	56.4	5.22	Sankaranarayanan *et al.*, 2017

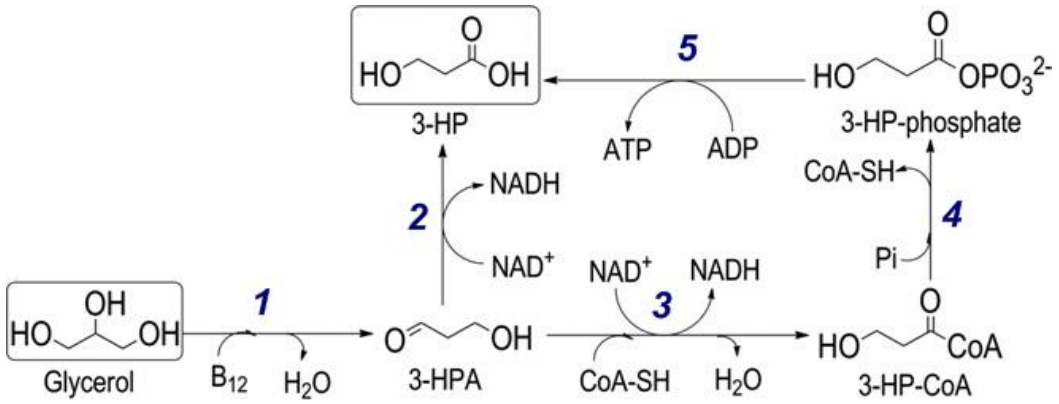


Figure 3 - Glycerol pathways for 3-HP production (adapted from Kumar, Ashok, and Park (2013)): 1 - glycerol dehydratase, 2 - aldehyde dehydrogenase, 3 - propionaldehyde dehydrogenase, 4 - phosphate propanoyltransferase, and 5 - propionate kinase.

2.4 3-HP production from sugar pathways by engineered microorganisms

2.4.1 Malonyl-CoA pathway

Malonyl-CoA is an intracellular metabolite used in many important biochemical pathways, such as fatty acids, polyketides, and flavonoids biosynthesis, and its conversion into 3-HP can be carried out with only one or two enzymatic reactions (Figure 4). Although there is no net ATP production in this pathway, the process can be performed in fully aerobic conditions where a small amount of the substrate is utilized in respiration to produce energy (CHEN *et al.*, 2014).

Rathnasingh *et al.* (2012) engineered, for the first time, an *E. coli* strain expressing the malonyl-CoA reductase from *Chloroflexus aurantiacus* and overexpressing acetyl-CoA carboxylase (*accADBCb*) and nicotinamide nucleotide transhydrogenase (*pntAB*) endogenous genes. The *pntAB* overexpression addressed the inability of malonyl-CoA reductase to utilize NADH as a cofactor, thus maintaining the redox balance. However, the final 3-HP titer was low, only 0.20 g/L in 24 h of shake flasks cultivation. The authors knocked out genes involved in byproducts pathways, but no increase in 3-HP titer was observed, differently from further works discussed in this section.

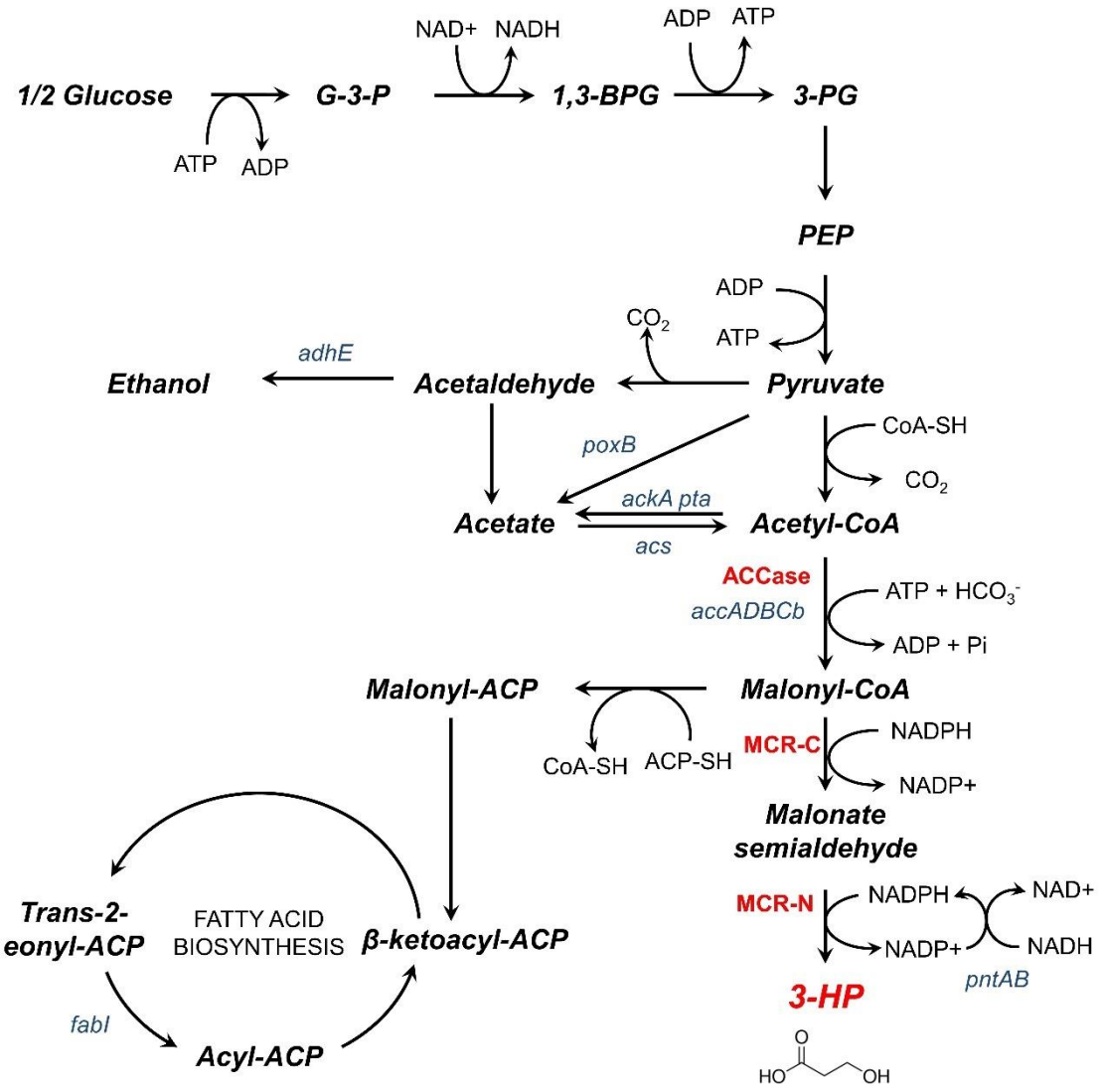


Figure 4 - Malonyl-CoA pathway for 3-HP production. Abbreviations: G-3-P, glyceraldehyde 3-phosphate; 1,3-BPG, 1,3-biphosphoglycerate; 3-PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; ACCase, acetyl-CoA carboxylase; MCR-C and MCR-N, C-terminal and N-terminal fragments from *C. aurantiacus* malonyl-CoA reductase, respectively.

Malonyl-CoA reductase from *C. aurantiacus* is a natural fused enzyme with two dehydrogenase/reductive domains: an alcohol dehydrogenase domain (aa 1-549) and an aldehyde dehydrogenase (CoA acylating) domain (aa 550 - 1219). This enzyme catalyzes the two-step conversion of malonyl-CoA into 3-HP with malonate semialdehyde as an intermediary. Liu *et al.* (2013) reported the expression of these two fragments separately as an N-terminal fragment (MCR-N, which catalyzes the malonate semialdehyde conversion into 3-

HP) and a C-terminal fragment (MCR-C, which catalyzes malonyl-CoA conversion into malonate semialdehyde). The results showed that MCR-C fragment is more active than the C-domain of the fused enzyme, MCR, both *in vitro* and *in vivo*, and the reaction catalyzed by the MCR-C fragment is likely the rate-limiting step of the malonyl-CoA conversion into 3-HP. By expressing the fragments separately in an *E. coli* strain overexpressing acetyl-CoA carboxylase, the 3-HP production reached 0.15 g/L in shake flasks, whereas the strain expressing MCR was able to produce only 0.107 g/L.

Due to the enzimatic activity imbalance between MCR-C and MCR-N and the fact that the reaction catalyzed by MCR-C is the rate-limiting step, Liu *et al.* (2016) aimed to improve MCR-C enzymatic activity by direct evolution while reducing MCR-N expression level. Their final strain harboring a plasmid containing the mutated MCR-C, with MCR-N integrated into the chromosome and overexpressing acetyl-CoA carboxylase, was able to produce 40.6 g/L of 3-HP in 72 h of a previously optimized fermentation (30 °C and fully aerobic conditions) with a yield of 0.19 g/g glucose.

In a patent assigned to OPX Biotechnologies, Lynch, Gill, and Lipsocomb (2013) achieved 49 g/L of 3-HP in 69 h of fermentation, increasing acetyl-CoA supply by deleting byproducts formation (lactate, ethanol, acetate, and methylglyoxal) and reducing the expression level of *fabI*, which encodes for enoyl-ACP reductase from fatty acids biosynthesis pathway in *E. coli*.

Using a genome-scale editing technique to create a pathway library with different gene expression levels in a combinatorial way, Tarasava *et al.* (2018) aimed to improve metabolic flux towards malonyl-CoA. In an iterative cloning process composed of 5 rounds, six genes were combinatorially repressed by CRISPRi: *ldhA, ackA/pta, adhE, pflB, poxB,* and *msgA*, and combinations of 2-4 repressed genes showed better results. The same group also designed an *E. coli* strain to efficiently produce 3-HP using a technique called iCREATE (iterative CRISPR EnaAble Trackable Genome Engineering) (LIU *et al.*, 2018a). The genes involved in fermentation byproducts pathways (acetate - *ackA, poxB* and *pta*, lactate - *ldhA*, formate -*pflB*, and ethanol - *adhE*) were inactivated rather than repressed and combined with other genetic modifications. The final strain was able to produce 30 g/L of 3-HP in a fed-batch fermentation in 72 h.

2.4.2 β-alanine pathway

Using a genome-scale metabolic model of *S. cerevisiae*, Borodina *et al.* (2015) predicted a higher theoretical yield of 3-HP in the β-alanine pathway compared to the malonyl-CoA pathway since the latter is strongly ATP-dependent for acetyl-CoA synthesis. The β-alanine is the precursor of pantothenic acid (vitamin B5), a component of coenzyme A and acyl carrier protein, which are cofactors required for many enzymes. In bacteria, β-alanine is produced by aspartate decarboxylation catalyzed by aspartate-1-decarboxylase codified by the *panD* gene (Borodina *et al.*, 2015).

In a study presenting the production of 3-HP from CO_2 by cyanobacteria *Synechococcus elongatus* PCC7942 expressing both malonyl-CoA and β-alanine dependent pathways, Lan *et al.* (2015) reported that replacing ADC from *C. glutamicum* by another amino acid decarboxylase can improve carbon flux to MSA. ADC is an enzyme that is first translated into a pro-protein and then is proteolyzed by a maturation enzyme into α and β subunits that form the active enzyme. The maturation step may be challenging to control. Therefore, the authors cloned a glutamate decarboxylase homolog, since glutamate is structurally similar to aspartate, from *Aedes aegypti*, which uses pyridoxal phosphate (PLP) as a cofactor to facilitate decarboxylation. As a result, they obtained a higher 3-HP titer from the strains expressing the glutamate decarboxylase than the corresponding strains expressing *panD*.

Borodina *et al.* (2015) engineered a *Saccharomyces cerevisiae* strain capable of producing 13.7 ± 0.3 g/L of 3-HP in fed-batch fermentation, with a yield of 14.0 % C-mol/C-mol of glucose and 0.17 g/L-h of productivity after 80 h. To achieve these results, they first identified an enzyme capable of converting β-alanine into malonic semialdehyde and presenting activity when expressed in *S. cerevisiae*. The β-alanine pyruvate aminotransferase from *Bacillus cereus* (BcBAPAT) was chosen. Several enzymes were screened to catalyze malonic semialdehyde conversion into 3-HP, and 3-hydroxypropionate dehydrogenase from *E. coli* K-12 (EcHADH) presented the highest production of 3-HP. Next, aspartate-1-decarboxylase from *Tribolium castaneum* (TcCSADC) was inserted in the recombinant yeast to establish a synthetic pathway to produce β-alanine from aspartate. Finally, the overproduction of native aspartate aminotransferase and pyruvate carboxylase, as well as BcBAPAT, EcHADH, and multiple copies of TcCSADC, enhanced heterologous gene expression and metabolic flux towards aspartate (Figure 5-A). For this engineered strain, the yield from glucose was further improved to 25.6 % by cultivation under P-limitation in a 1-L fed-batch reactor at a dilution rate of 0.05 h^{-1} (LIS *et al.*, 2019).

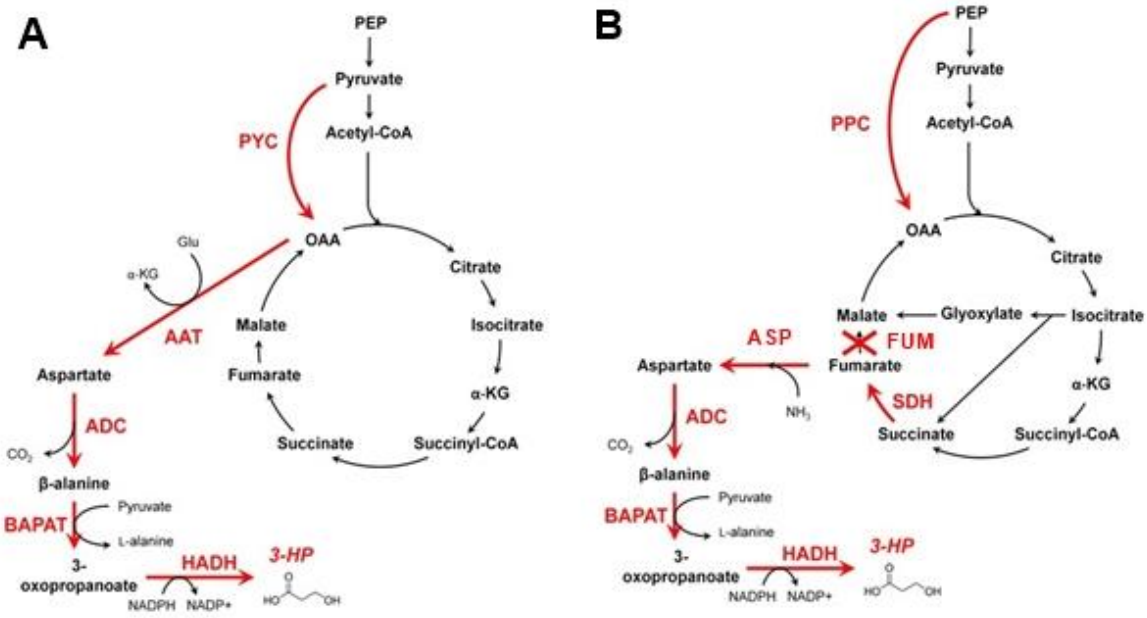


Figure 5 - The β-alanine pathway to produce 3-HP from (A) Borodina *et al.* (2015) and (B) Song *et al.* (2016). Enzymes (in red): PPC, PEP carboxylase; PYC, pyruvate carboxylase; AAT, aspartate aminotransferase; ASP, aspartase; SDH, succinate dehydrogenase; FUM, fumarase; ADC, L-aspartate-α-decarboxylase, BAPAT, β-alanine-pyruvate aminotransferase; HADH, 3-hydroxyacid dehydrogenase. Red arrows represent overexpressed reactions. The red X represent reaction knock-out.

Song *et al.* (2016) reported a study of 3-HP production in *E. coli* cells through the β-alanine using a different pathway in which aspartate is produced by fumaric acid rather than OAA (Figure 5-B). Starting from an *E. coli* strain able to efficiently produce fumaric acid (SONG *et al.*, 2013), the authors overproduced the enzymes aspartase (*aspA*, from *E. coli* W3110) and L-aspartate-α-decarboxylase (*panD*, from *C. glutamicum* ATCC12032) to create an efficient β-alanine producer (SONG *et al.*, 2015). In their next work, the authors expressed the genes for two enzymes: β-alanine-pyruvate aminotransferase (*pa0132* gene from *P. aeruginosa* PAO1) and 3-hydroxyacid dehydrogenase (*ydfG* from *E. coli* W3110) to convert β-alanine into 3-HP. The strain used in the study, CWF4NAS, had six knocked-out genes to increase the flux towards fumaric acid: *iclR*, which is a repressor of *aceBAK* operon (responsible for glyoxylate shunt); *fumC, fumA, and fumB*, which prevent fumarate conversion into malate; *ptsG*, to increase PEP pool in the cell and thus the formation of PEP-derived products; and *lacI*, to prevent the IPTG inducible system from being used since it increases process cost. Also, in the CWF4NAS strain, the native promoter of aspartase and succinate

dehydrogenase were replaced for the *trc* promoter to increase the metabolic flux for β-alanine and fumaric acid, respectively. The enzymes responsible for converting fumaric acid into β-alanine had its genes expressed in the plasmid pTac15kPTAP in which *ppc* gene codifying for PEP-carboxylase was also overexpressed. The fed-batch fermentation of the final strain produced 31.1 g of 3-HP in 49 h, with a yield of 0.423 g 3-HP/g glucose and less than 3 g/L of acetic acid.

2.5 Utilization of xylose as a substrate for 3-HP production

Thus far, we have considered microbial fermentations to 3-HP production utilizing glucose as the only carbon source. However, since xylose is the predominant sugar in the hemicellulosic fraction, which composes 30-45 % of lignocellulosic biomass, efficient conversion of this sugar is a prerequisite for the implementation of economically feasible projects for lignocellulosic bioconversion (JUNG *et al.*, 2015; KWAK *et al.*, 2019).

Microbial cells, such as *E. coli*, growing in a sugar mixed medium exhibit a mechanism of catabolite repression of sugars (or carbon catabolite repression - CCR), where glucose inhibits the consumption of other sugars, like xylose, before glucose depletion (KIM; BLOCK; MILLS, 2010). Hence, simultaneous glucose and xylose consumption has been treated as a hurdle to be overcome to convert lignocellulosic hydrolysates efficiently into bio-based chemical products for industry.

Xylose can enter glycolysis by two different pathways. Some yeasts can convert xylose into xylulose through an oxidative pathway composed of xylose reductase (XR), which has NADH or NADPH as cofactors, and xylitol dehydrogenase (XDH), which has NAD+ as cofactor. Another possibility to convert xylose into xylulose is through the isomerization pathway, by xylose isomerase, which does not require a cofactor, and it is mainly found in bacteria (BRUINENBERG *et al.*, 1984; HÖFER; BETZ; KOTYK, 1971). Xylulose is then converted into xylulose 5-phosphate by xylulose kinase (XK), which enters glycolysis through the pentose phosphate pathway (PPP) (Figure 6).

Jung *et al.* (2015) engineered an *E. coli* strain to produce 3-HP through the glycerol pathway but using glucose and xylose as substrates. These sugars were initially converted into glycerol and then to 3-HP. In order to reduce CCR, the authors deleted the *ptsG* gene, which codifies EIIBC, a specific enzyme of PTS, and overexpressed the *xylR* gene with a synthetic promoter, P_{mut}. The *xylR* gene codifies a transcriptional activator, XylR, from *xyl* operon in *E. coli* composed by six genes, including *xylE*, *xylF*, *xylG* and *xylH*, xylose transporter

components, apart from *xylA* (XI) and *xylB* (XK). Deletion of *ptsG* gene (*ΔptsG*) and *xylR* overexpression promoted a synergic effect on CCR reduction and increased transcriptional and translational levels of the *xyl* operon. For 3-HP production, the authors inserted the gene clusters *dhaB*$_1$*B*$_2$*B*$_3$ (glycerol dehydratase) and *dhaR*$_1$*R*$_2$ (glycerol dehydratase reactivase) from *L. brevis* and *aldH* (aldehyde dehydrogenase) from *P. aeruginosa*. Then, in sugar limiting fed-batch fermentation (since inhibitory effects of glucose and/or xylose over glycerol conversion into 3-HP were verified), the engineered strain produced 29.7 g/L of 3-HP with a yield of 0.36 g/g of sugars and a productivity of 0.54 g/L-h. Further modifications led to a strain producing 37.6 g/L of 3-HP, with yield and productivity of 0.63 g/L-h and 0.17 g/g of sugars in fed-batch cultures (HEO *et al.*, 2019**).** However, as discussed before, the main challenge of 3-HP production through the glycerol pathway is the requirement of vitamin B_{12}, which is a coenzyme for glycerol dehydratase, or the use of an oxygen-sensitive vitamin B_{12}-independent enzyme. The medium supplementation with vitamin B_{12} is impractical since it increases process costs, and the process needs to be conducted under aerobic conditions for NAD+ regeneration in ETC.

Kildegaard *et al.* (2015) utilized a strain of *S. cerevisiae* efficient in consuming xylose, engineered with the insertion of the genes of XDH, XR, and XK (from *Pichia stipitis*) and evolved in mineral medium with xylose to produce 3-HP through malonyl-CoA pathway and β-alanine pathway. The last reaction of β-alanine pathway, malonic semialdehyde to 3-HP, was carried out utilizing both NADH-dependent enzyme 3-hydroxyisobutyrate dehydrogenase (HIBADH, from *Pseudomonas putida*) and NADPH-dependent enzyme 3-hydroxypropionate dehydrogenase (HADH, from *E. coli*). The engineered strains showed higher 3-HP production with xylose as carbon source by β-alanine pathway, achieving a titer of 7.37 ± 0.17 g/L in 120 h of fed-batch fermentation and a yield of 29 ± 1 % C-mol of 3-HP/C-mol of xylose.

By using the genome-scale engineering technique cited before (iCREATE technique), Liu *et al.* (2018b) developed an *E. coli* strain (BGgxk4P4) capable of xylose and glucose simultaneous consumption and with a higher biomass growth than the parental strain (BW25113 *ΔpoxBΔldhAΔpflBΔadhEΔptaΔackA fabI*$_{S241F}$, or BG - LIU *et al.*, 2018a), by deletion of *ptsHI* genes and mutating *glk* and *galP* genes (glucokinase and galactose permease, respectively). The authors also identified mutations in the *soxR* gene, which codifies a transcriptional factor, as favorable for improving *E. coli* tolerance to furfural and acetate, inhibitors present in lignocellulosic hydrolysate. 3-HP production, in a medium containing corn stover hydrolysate with furfural and acetate addition, proceeded in fed-batch mode. The results obtained from BG strain capable of producing 3-HP were compared to the tolerant strain

(BGHPT). In 72 h of growth, BGHPT strain produced 8 g/L of 3-HP, 8-fold more than BG strain.

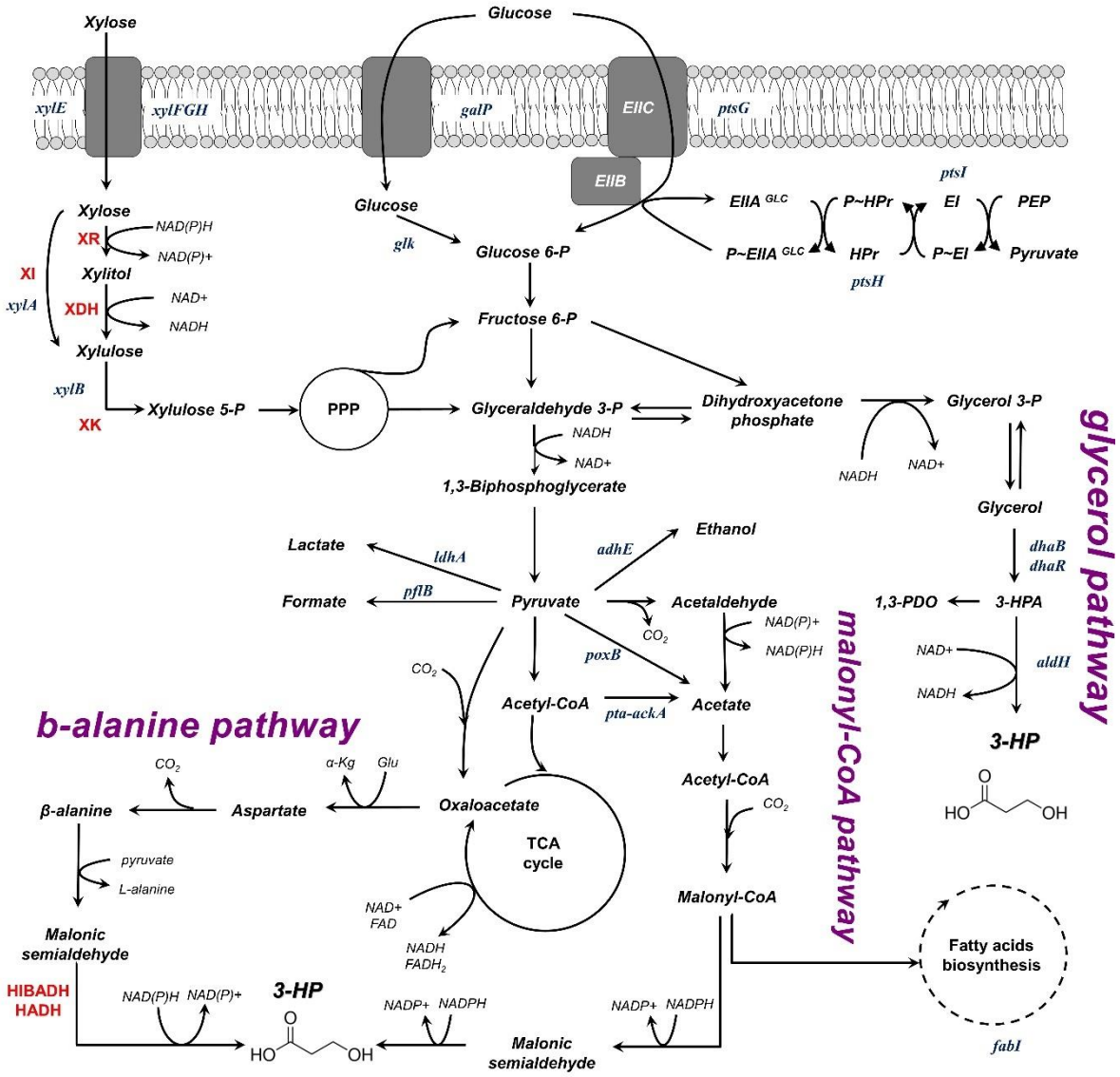


Figure 6 - Sugars uptake system and xylose metabolism for the production of 3-HP. Genes (blue) and enzymes (red): *galP*, galactose permease; *ptsG*, PTS glucose specific EIIBC component; *ptsH*, phosphohistidine carrier protein; *ptsI*, phosphoenolpyruvate protein phosphotransferase; *glk*, glucokinase; XR, xylose reductase; XDH, xylose dehydrogenase; *xylA*, xylose isomerase (XI); *xylB*, xylulose kinase (XK); *dhaB*, glycerol dehydratase; *dhaR*, glycerol dehydratase reactivase; *aldH*, aldehyde dehydrogenase; *ldhA*, lactate dehydrogenase; *pflB*, pyruvate-formate lyase; *adhE*, aldehyde-alcohol dehydrogenase; *poxB*, pyruvate:quinone oxidase; *pta-ackA*, phosphate-acetyltransferase and acetate kinase; *fabI*, enoyl-ACP reductase; PYC, pyruvate carboxylase; HIBADH, 3-hydroxyisobutyrate dehydrogenase; HADH, 3-

hydroxypropionate dehydrogenase. Abbreviations: ACP, acyl carrier protein; PPP, pentose phosphate pathway; PEP, phosphoenolpyruvate; 3-HPA, 3-hydroxypropionaldehyde; 1,3-PDO, 1,3-propanediol; 3-HP, 3-hydroxypropionic acid; TCA, tricarboxylic acid; PTS, phosphotransferase system.

2.6 Metabolic engineering of microbial cell factories using metabolic models

Industrial biotechnology has been increasing its presence in the economics scenario thanks to the advanced genetic tools and computational methods for analyzing complex problems. In metabolic engineering, the traditional approaches for designing the desired phenotype in a microbial cell have been replaced, to some extent, by mathematical modeling of complex metabolic networks, which can cover up to a significant extension of the microbial genome in the case of genome-scale metabolic models (GEMs). Such models allow predicting cell features based on *in silico* simulations, leading to a rational redesign of the microbial cells that is less time consuming and, thus, ultimately relies on computational algorithms analysis to further optimize the production of a given compound. This topic is under rapid expansion, but only the fundamental aspects will be presented here, providing enough information to understand and discuss the *in silico* analysis performed in this work. Further details can be found in the reviews of Maia, Rocha, and Rocha (2016) and the work of Klamt, Hädicke, and von Kamp (2014).

A metabolic network from a microbial strain can be described by combining some information as *i.* the set of internal metabolites, transport systems, and enzymatic biochemical reactions *ii.* reversibility, *iii.* gene-enzyme-reactions associations and *iv.* reaction kinetics (KLAMT; HADICKE; VON KAMP, 2014). These networks, such as the one in Figure 7-A, can be written in the form of mass balance equations that, along with the boundary conditions defined by the constraints imposed to the phenotype by thermodynamics, genetic and physicochemical laws (LEWIS; NAGARAJAN; PALSSON, 2012), compose the stoichiometric model of the network.

These stoichiometric models can be combined with constraint-based modeling methods to predict microbial strain features and be used in metabolic engineering tasks for strain optimization. Some assumptions can be adopted to simplify network modeling to avoid the need for kinetics laws and parameters data. Since microbial metabolism often involves fast reactions and small turnover times of metabolites, their concentration can be set in a quasi-steady state (KLAMT; HADICKE; VON KAMP, 2014). In other words, changes in the

metabolism are faster than the cell growth rate and dynamic environmental changes (MAIA; ROCHA; ROCHA, 2016). This assumption implies that the dynamic term of the internal metabolites' mass balance can be set as null. Therefore, the network can be represented by Equation 1:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0} \qquad \text{(Equation 1)}$$

where **S** is the *m x n* matrix of the stoichiometric coefficients from a given network with *m* metabolites and *n* internal reactions and **v** is the vector of *n* reaction rates (or, metabolic fluxes). The boundary conditions and reversibility of each reaction are also imposed in the model as in Equation 2:

$$\begin{aligned} 0 \leq v_i \leq \beta_i, \forall\, i \in N_{irrev} \\ \alpha_i \leq v_i \leq \beta_i, \forall\, i \in N_{rev} \end{aligned} \qquad \text{(Equation 2)}$$

where v_i is the metabolic flux of the reaction *i*, N_{rev} and N_{irrev} are subsets of *N* (set of all reactions) and α_i and β_i are the lower and upper bounds of the reaction *i*, respectively (MAIA; ROCHA; ROCHA, 2016). The stoichiometric matrix (**S**) and the fluxes matrix (**v**) for the given network are represented in Figure 7-B.

As the number of biochemical reactions of a given network is often superior to the number of compounds, the system given by Equation 1 is undetermined, and the possible solutions are represented by a flux cone (or hypercone, Figure 7 - C). This system is then solved by an optimization problem with biological assumptions (KO *et al.*, 2020; TAMURA; LU; AKUTSU, 2015). With Flux Balance Analysis (FBA), a constraint-based modeling method, the system is solved with an objective function, using linear programming. The solution is then a flux vector, from inside the flux cone, with defined metabolic fluxes for each network's reaction, which satisfies the objective function (Figure 7-D). This objective function is usually defined as the biomass growth rate, but it does not narrow to it. The use of the biomass growth rate as the objective function is supported by the biological assumption that microbial cells optimize their metabolism towards biomass formation in the exponential growth phase. One alternative to the FBA method is the parsimonious FBA (pFBA) that calculates the flux distribution minimizing the sum of all fluxes in the model while fulfilling the objective function (LEWIS *et al.*, 2010; KO *et al.*, 2020). In this case, the pFBA method looks for a solution that has a biological meaning of the least amount of enzyme used for each reaction. Other methods like

Minimization of Metabolic Adjustment (MOMA), Regulatory On/Off Minimization of Metabolic flux changes (ROOM), and Metabolic Flux Analysis (MFA), the latter based on the use of previously measured fluxes as constraints, can also be applied for solving the mass balance problem.

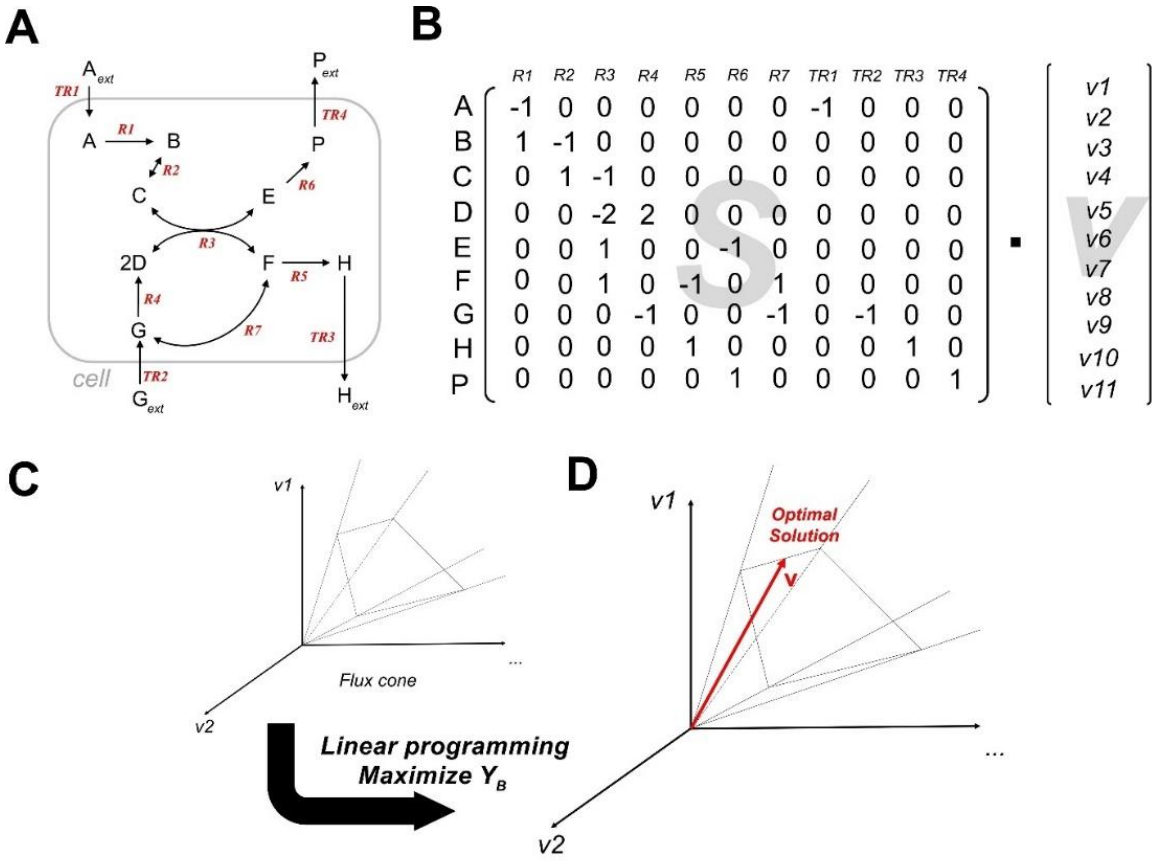


Figure 7 – A – Diagram of a given metabolic network, B – the stoichiometric matrix of the proposed network, C – the flux (hyper)cone, and D – the optimal solution vector.

The FBA and pFBA methods are used for *in silico* simulations of a given metabolic network. However, microbial cell factories' rational design often requires pathway optimization by gene/reaction knock-outs and/or under/overexpression. This computational strain optimization uses sophisticated algorithms like *OptGene* (that uses Evolutionary Algorithms – EA – and Simulated Annealing – SA) and *OptKnock* (that uses Mixed Integer Linear Programming – MILP).

In this context, *Optflux* (ROCHA *et al.*, 2010) is an open-source platform for metabolic engineering tasks that allows the simulation of metabolic models of wild-type and mutant strains (gene/reaction deletion or under/overexpression) based on different methods (FBA,

pFBA, MOMA, ROOM, MFA) and under different environmental conditions defined by the user. It also allows the optimization of models with a range of algorithms (EA, SA, and others) that look for reaction/gene deletion or under/overexpression based on a pre-defined objective function. Another remarkable feature is the possibility of visualizing the strain simulations using model layouts that the user can modify to deal with a specific pathway in the model. This software was used in this work to find metabolic targets that could enhance 3-HP production by *E. coli* cells.

3 METHODOLOGY

3.1 *In silico* optimizations

The genome-scale metabolic model of *Escherichia coli* K-12 MG1655 *i*ML1515 (MONK *et al*., 2017) was modified to include the transamination step of β-alanine and pyruvate into malonic semialdehyde and L-alanine, which is the heterologous reaction of the β-alanine pathway necessary for 3-HP production in *E. coli*. This modification was made in the SBML (Systems Biology Markup Language) file of the model by inserting the code for the R_APATr reaction (β-alanine:pyruvate aminotransferase – see Appendix A). The modified *i*ML1515 model comprises 2377 internal reactions, 1075 internal metabolites, 802 external metabolites, and their respective drain reactions, besides 1516 genes and 2266 gene rules (gene-reaction associations). See Figure A2 in Appendix A to better understand the heterologous pathway for 3-HP production through the pyrimidine degradation pathway in the original *i*ML1515 and R_APATr insertion point.

Simulation and optimization runs were carried out in *Optflux* software v.3.4.0 (ROCHA *et al.*, 2010) using a DESKTOP-NDE10UR, Intel ® Core ™ i7-4790 CPU @3.60GHz, with 8.00 GB of RAM, 64 bits operational system, x64-based processor, and operational system compilation 17134.829, from the Department of Chemical Engineering, Federal University of São Carlos – São Carlos/Brazil.

Simulations were run using the pFBA method with biomass reaction (referred to as R_BIOMASS_Ec_iML1515_core75p37M) as the objective function – OF – using two different carbon sources (defined as Environmental Conditions): glucose uptake rate of -10 mmol.gDW^{-1}.h^{-1} (by default) or xylose uptake rate of -12 mmol.gDW^{-1}.h^{-1}. The model's biomass formation is represented by a reaction that converts 94 specific metabolites required for cell growth, such as amino acids, vitamins, and ATP, into ADP, PPi, Pi, and H^+.

The reaction R_EX_3hpp_e (3-HP exchange with the extracellular medium) was also selected as the OF to evaluate the highest 3-HP drain flux [mmol.gDW^{-1}.h^{-1}] of the model in two different Environmental Conditions: *i.* in the original *i*ML1515 model and *ii.* in the modified model (with R_APATr addition). It is worth mentioning that 3-HP is represented as 3-hydroxypropanoate (3-HPP) in the model. In case of drain reactions, the signal convention in the software is (-) for metabolite uptake into the cell. Also, the signal convention in the software is (-) for reactions driven to reactants formation (for reversible reactions).

Pathway optimization runs used the Strength Pareto Evolutionary Algorithm (SPEA2; ZITZLER, LAUMANNS, THIELE; 2001) to maximize the Biomass Coupled Yield (BPCY,

Equation 3) OF. The option SPEA2 Reaction Under/Over expression (which includes knock-out) was chosen.

$$BPCY = \frac{R_{BIOMASS} \cdot R_{PRODUCT}}{R_{SUBSTRATE}} \qquad \text{(Equation 3)}$$

where $R_{BIOMASS}$, $R_{PRODUCT}$, $R_{SUBSTRATE}$ are the biomass formation flux (R_BIOMASS_Ec_iML1515_core75p37M), product exchange flux (R_EX_3hpp_e) and substrate uptake flux (R_EX_glc__D_e, in the case of glucose as substrate and R_EX_xyl__D_e, in the case of xylose as substrate), respectively. The algorithm ran 30,000 iterations (solutions evaluations) and, for each of them, utilized the pFBA method to simulate the mutant strain with biomass growth rate set as OF. The maximum number of modifications was set as ≤ 10, and drain reactions were excluded from the evaluation. This Optimization setup is shown in Figure 8.

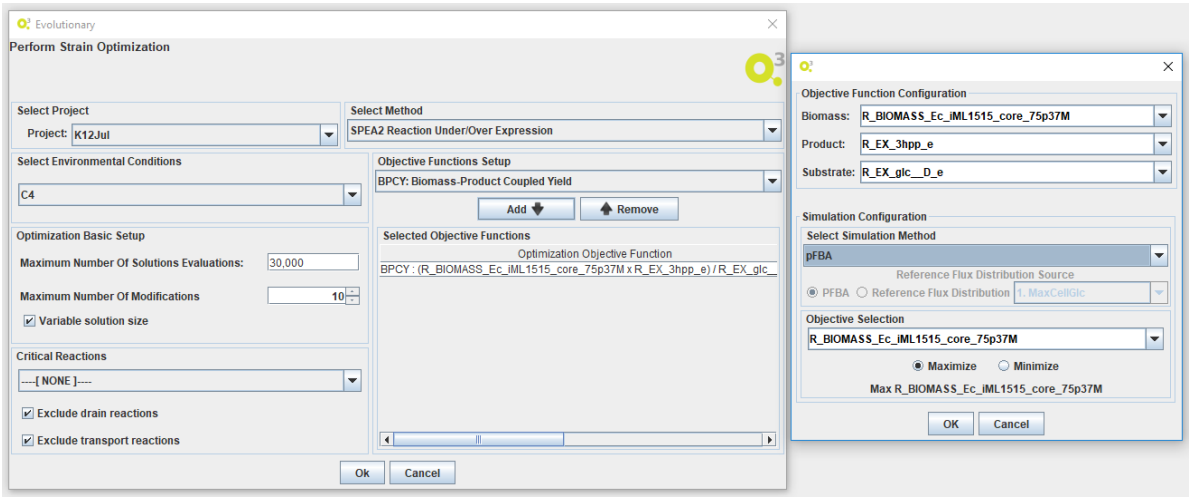


Figure 8 – Optimization setup for the *E. coli* K-12 MG1655 model *i*ML1515 to produce 3-HP: the evolutionary algorithm chosen, SPEA2, OF set as BPCY and pFBA used for the mutant strain simulations, maximizing biomass growth rate.

3.2 Bacterial strains and plasmids

Bacterial strains and plasmids used in this work are presented in Table 3. *E. coli* DH5α was used for cloning and plasmids propagation, while the other strains were used for gene amplification from genomic DNA for cloning purposes. *E. coli* K-12 MG1655 (F^- λ^- *ilvG*$^-$ *rfb*-

50 rph-1) was used as the production strain. The plasmids pETM9 and pRSM6 were used as cloning and expression vectors. These plasmids were derived from pETM6 and pRSM3 from the series *ePathBrick Vectors* (XU *et al.,* 2012), donated by Prof. Mattheos A. G. Koffas from Rensselaer Polytechnic Institute, USA. The original plasmids had their T7 promoter replaced by the lacUV5 promoter (SILVA *et al.,* 2021), and the positions of the isocaudamer enzymes *Xma*J (*Avr*II) and *Xba*I were switched. The latter modification was carried out through two rounds of site-directed mutagenesis with primers 1 – 4, from Table B1 (see Appendix B for more details about the procedure).

Genomic DNA from bacterial strains was purified using the Axygen ® AxyPrep ™ Bacterial Genomic DNA Miniprep Kit (Corning). Cells from frozen stock (50 % glycerol, - 80 °C) were incubated in 5 mL of LB medium (tryptone 10,0 g/L, yeast extract, 5,0 g/L, and NaCl 5,0 g/L) at 37 °C and 225 rpm for 16 h in a rotary shaker. Then, 1 x 10^9 bacteria (1 mL of cell culture at OD_{600} = 1 – 1.5) were harvested to proceed with genomic DNA purification. All DNA elutions were done with nuclease-free water. Genomic DNA from *S. cerevisiae* CAT-1 was kindly donated by Prof. Anderson Ferreira da Cunha, from the Genetics and Evolution Department – DGE – of UFSCar.

Table 3 - Bacterial strains and plasmids used in this work.

Strains	**Source**
Escherichia coli DH5α	LAFAC
Escherichia coli K-12 MG1655	LAFAC
Corynebacterium glutamicum ATCC 13032	CCT – FAT
Actinobacillus succinogenes ATCC 55618	CCT – FAT
Saccharomyces cerevisiae CAT-1	LAFAC
Pseudomonas aeruginosa PA01	USP[1]
Engineered strains	**Description**
PS100	*E. coli* K-12 MG1655 harboring pEbtyGpD
PS200	*E. coli* K-12 MG1655 harboring pEbtyGpD*
PS300	*E. coli* K-12 MG1655 harboring pESbtyGpD
PS400	*E. coli* K-12 MG1655 harboring pESbtyGpD*
KS	*E. coli* K-12 MG1655 harboring pES
PS001	*E. coli* K-12 MG1655 Δ*ppc*
PS051	*E. coli* K-12 MG1655 Δ*ppc* harboring pRk

PS150	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRk
PS151	*E. coli* K-12 MG1655 Δ*ppc* harboring pEbtyGpD and pRk
PS160	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRak
PS161	*E. coli* K-12 MG1655 Δ*ppc* harboring pEbtyGpD and pRak
PS003	*E. coli* K-12 MG1655 Δ*ilvE*
PS163	*E. coli* K-12 MG1655 Δ*ilvE* harboring pEbtyGpD and pRak
PS173	*E. coli* K-12 MG1655 Δ*ilvE* harboring pEbtyGpD and pRAak
PS183	*E. coli* K-12 MG1655 Δ*ilvE* harboring pEbtyGpD and pRrak
PS170	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRAak
PS180	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRrak
PS190	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRArak
PS193	*E. coli* K-12 MG1655 Δ*ilvE* harboring pEbtyGpD and pRArak
PSO101	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRA
PSO102	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRr
PSO103	*E. coli* K-12 MG1655 Δ*ilvE* harboring pEbtyGpD
PSO104	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRAr
PSO105	*E. coli* K-12 MG1655 Δ*ilvE* harboring pEbtyGpD and pRA
PSO106	*E. coli* K-12 MG1655 Δ*ilvE* harboring pEbtyGpD and pRr
PSO107	*E. coli* K-12 MG1655 Δ*ilvE* harboring pEbtyGpD and pRAr
PSX100	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRx
PSXM100	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRx*
PSR100	*E. coli* K-12 MG1655 harboring pEbtyGpD and pRSM6

Plasmids		**Source**
pETM6	*T7* promoter, Amp^R	Xu *et al.*, 2012
pRSM3	*T7* promoter, Kan^R	Xu *et al.*, 2012
pETM9	pETM6 *PT7::lacUV5* and *Xba*I / *XmaJ*I sites swapped	This work
pRSM6	pRSM3 *PT7::lacUV5* and *Xba*I / *XmaJ*I sites swapped	This work
pCas9cr4	Cm^R, Cas9 from *Streptococcus pyogenes* ,tagged with *ssrA*, under P_{TetR} control	Reisch and Prather, 2015
pKDsgRNA	$Strep^R$, λ-Red system (Bet, Exo and Gam) under P_{araBAD} control	Reisch and Prather, 2015

pKDsgRNA*ppc*	$Strep^R$, λ-Red system (Bet, Exo, and Gam) under P_{araBAD} control, harboring the sgRNA for *ppc* knock-out	This work
pKDsgRNA*ilvE*	$Strep^R$, λ-Red system (Bet, Exo, and Gam) under P_{araBAD} control, harboring the sgRNA for *ilvE* knock-out	This work
pEpD	pETM9 harboring *panD* gene	This work
pEpD*	pETM9 harboring *panD* Glu56Ser mutant gene	This work
pEyG	pETM9 harboring *ydfG* gene	This work
pEbt	pETM9 harboring *pa0132* gene	This work
pES	pETM9 harboring *acs* gene	This work
pEbtyG	pETM9 harboring *pa0132* and *ydfG*	This work
pEbtyGpD	pETM9 harboring *pa0132*, *ydfG* and *panD*	This work
pEbtyGpD*	pETM9 harboring pa0132, *ydfG*, and *panD* Glu56Ser mutant	This work
pESbtyGpD	pETM9 harboring *acs*, *pa0132*, *ydfG*, and *panD*	This work
pESbtyGpD*	pETM9 harboring *acs*, *pa0132*, *ydfG*, and *panD* Glu56Ser mutant	This work
pRk	pRSM6 harboring *pckA* gene	This work
pRa	pRSM6 harboring *aspC* gene	This work
pRA	pRSM6 harboring *ALT1* gene	This work
pRr	pRSM6 harboring *alr* gene	This work
pRak	pRSM6 harboring *aspC* and *pckA*	This work
pRAak	pRSM6 harboring *ALT1*, *aspC*, and *pckA*	This work
pRrak	pRSM6 harboring *alr*, *aspC*, and *pckA*	This work
pRArak	pRSM6 harboring *ALT1*, *alr*, *aspC* and *pckA*	This work
pRx	pRSM6 harboring *xylR* gene	This work
pRx*	pRSM6 harboring *xylR* P363S mutant gene	This work
pRAr	pRSM6 harboring *ALT1* and *alr*	This work

[1]This strain was kindly donated by prof. Uelinton Manoel Pinto, from Pharmaceutical Sciences Faculty of São Paulo University (USP).

Plasmid DNA was purified using PureYield™ Plasmid Miniprep System or Wizard® Plus SV Miniprep DNA Purification System (Promega). Before plasmid extraction, *E. coli*

DH5α cells were inoculated overnight in LB medium containing the appropriate antibiotic (Ampicillin, 80 µg/mL or Kanamycin 50 µg/mL). Genomic DNA and plasmid DNA concentration was measured using NanoDrop™ 2000 Spectrophotometer (Thermo Scientific).

3.3 Cloning procedure

3.3.1 Gene amplification by Polymerase Chain Reaction (PCR)

The target genes (*ydfG*, *aspC*, *alr*, and *xylR* from *E. coli* K-12 MG1655, *pa0132* from *P. aeruginosa* PA01, *panD* from *C. glutamicum* ATCC 13032, *ALT1* from *S. cerevisiae* CAT-1 and *pckA* from *A. succinogenes* ATCC 55618) were amplified by PCR from genomic DNA using primers listed in Table 4.

Table 4 - Primers list to amplify the correspondent target gene from genomic DNA. The restriction enzyme cut sites are underlined for each primer (*Nde*I: CATATG, *Xho*I: CTCGAG).

Gene (length)	Primer (Enz)	5' - 3' sequence	T_m [°C]
pa0132	F (*Nde*I)	GAACATATGAACCAGCCGCTCAAC	57
(1347 bp)	R (*Xho*I)	GCACTCGAGTCAGGCGATGCCGTTGAG	57
ydfG	F (*Nde*I)	CGCCATATGATCGTTTTAGTAACTGGAGC	55
(747 bp)	R (*Xho*I)	CTGCTCGAGTTACTGACGGTGGACATTCAG	55
panD	F (*Nde*I)	CATCATATGCTGCGCACCATC	60
(411 bp)	R (*Xho*I)	GTCCTCGAGCTAAATGCTTCTCGACGTCAA	59
ALT1	F (*Nde*I)	GCCCCATATGTTATCACTGTCTGCCAA	53
(1779 bp)	R (*Xho*I)	GATCTCGAGTCAGTCACGGTATTGGTCAAA	52
alr	F (*Nde*I)	CCTCATATGCAAGCGGCAACTG	57
(1080 bp)	R (*Xho*I)	CTGCTCGAGGCATCCGGCACAGACAAT	58
aspC	F (*Nde*I)	GCCCATATGTTTGAGAACATTACCGCC	55
(1191 bp)	R (*Xho*I)	GCCCTCGAGTTACAGCACTGCCACAATCG	56
pckA	F (*Nde*I)	GCCCATATGACTGACTTAAACAAACTCGTT	55
(1611 bp)	R (*Xho*I)	GCCCTCGAGTTATGCTTTTGGACCGGCG	56
xylR	F (*Bgl*II)	CGCAGATCTAATGTTTACTAAACGTCACCGC	56
(1179 bp)	R (*Xho*I)	GGCCTCGAGCTACAACATGACCTCGCTATTTAC	56

All primers were synthesized by Exxtend (Campinas, SP).

For gene amplification, a 50 µL PCR reaction was set up using 25 µL of TAQ HighFidelity Pool Master Mix 2x (Cellco), 2.5 µL of primer mix (10 pmol/µL of each primer), 50 ng of genomic DNA, and nuclease-free water to complete the reaction volume. PCR reactions were carried out in a Px2 Thermal Cycler (Thermo Electron Corporation) accordingly to the following steps: *i.* DNA denaturation at 95 °C for 5 min, *ii.* 30 cycles of DNA denaturation (95 °C for 15 s) followed by primer annealing ($T_A = T_m - 3$ °C, for 30 s) and DNA extension at 72 °C for *t* min (where *t* is dependent on the size to be amplified – 1.5 min for each 1 kb), and *iii.* DNA extension at 72 °C for 5 min. The amplicon was then purified using Wizard® SV Gel and PCR Clean-up System (Promega), and the DNA concentration was measured.

3.3.2 Restriction and ligation cloning

The genes *pa0132*, *ydfG*, *panD*, *pckA*, *aspC*, *ALT1*, *alr*, and *xylR* were cloned into the MCS of the pETM9 and pRSM6 plasmids in their *Nde*I/*Xho*I (*Bgl*II/*Xho*I for *xylR*) sites using restriction and ligation cloning.

Digestion: gene amplicons and plasmids were digested with the appropriate enzymes (FastDigest, Thermo Scientific™) using 1000 ng of plasmid/amplicon DNA, 1 µL of each enzyme, 3 µL of FastDigest 10x Green Buffer, and water to complete 30 µL. The reaction was incubated at 37 °C for 3 hours. The reaction mixture was then separated by electrophoresis (80 V, 35 min) in agarose gel (1 % in TAE buffer). The bands corresponding to the digested amplicon (insert) and the digested plasmid (vector) were excised, purified using Wizard® SV Gel and PCR Clean-up System (Promega), and the DNA concentration was measured.

Ligation: the ligation mixture was prepared using 75 ng of vector and *x* ng of insert DNA calculated with the ligation calculator from New England Biolabs ® using a 3:1 insert: vector ratio (https://nebiocalculator.neb.com/#!/ligation), 1 µL of T4 DNA Ligase (Thermo Scientific), 2 µL of 10X T4 DNA ligase buffer and water to 20 µL. This mixture was incubated at room temperature for 1 h.

Transformation: 10 µL of the ligation product was transformed into chemically competent *E. coli* DH5α cells following standard protocols (SAMBROOK AND RUSSEL, 2001).

All cloning procedures were first simulated in SnapGene® v.4.2.11. Plasmid maps and the history of cloning simulations are provided in Appendix C.

3.3.3 Circular Polymerase Extension Cloning (CPEC)

The *acs* gene (1959 bp) from *E. coli* K-12 MG1655 was cloned in the pETM9 plasmid by Circular Polymerase Extension Cloning (CPEC, QUAN AND TIAN, 2009). In this technique, vector and insert are amplified with overlapping regions at the ends, and the polymerase extends the annealed amplicons to form the final construct.

First, vector and insert were amplified from plasmid pETM9 and *E. coli* K-12 MG1655 genomic DNA, respectively, by PCR using the primer pairs Vec_acs/CPEC_F and _R for the vector, and Frag_acs/CPEC_F and R for the insert (fragment). These primers (Table 5) were designed with the Gibson Assembly tool of SnapGene® v.4.2.11 and were analyzed with the IDT OlygoAnalyzer tool for thermodynamics parameters. Vector and fragment were amplified using 25 µL of Platinum™ SuperFi II PCR Master Mix (Invitrogen), 2.5 µL of primer mix (10 pmol/µL each), 10 ng of DNA template (genomic or plasmid), and water to 50 µL. PCR reaction was performed using Platinum™ SuperFi II parameters: *i.* DNA denaturation at 98 °C for 30 s, *ii.* 30 cycles of DNA denaturation (98 °C for 10 s) followed by primer annealing at 60 °C for 30 s, and DNA extension at 72 °C for *t* min, where *t* is dependent on the size to be amplified – 30 s/kb, and *iii.* DNA extension at 72 °C for 5 min. Then, the reaction mixture containing the vector was digested with *Dpn*I (FastDigest, ThermoScientific™) at 37 °C for 1 h. The PCR products were separated by electrophoresis, followed by band excision and purification.

Table 5 - Primers used for *acs* gene cloning by Circular Polymerase Extension Cloning (CPEC). Complementary regions are underlined.

Primer	**5' - 3' sequence**
Vec_acs/CPEC_F	CGTAACCCACAATTGCCTCGAGTCTGGTAAAGAAACCG
Vec_acs/CPEC_R	TTGTGAATTTGGCTCATTGTATATCTCCTTCTTAAAGTTAAA CAAAATTATTTCTAGAGGGGA
Frag_acs/CPEC_F	AAGAAGGAGATATACAATGAGCCAAATTCACAAACACACC A
Frag_acs/CPEC_R	TCTTTACCAGACTCGAGGCAATTGTGGGTTACGATGG

The purified PCR products were pooled together, and the hybridized insert and vector were extended using each other as a template (Figure 9). For this reaction, 10 µL of Platinum™ SuperFi II PCR Master Mix (Invitrogen), 200 ng of the vector DNA, *x* ng of insert DNA

(accordingly to the NEBioCalculator for a 2:1 ratio of insert:vector), and water to complete 20 µL, were submitted to the following thermal cycles: *i.* DNA denaturation at 98 °C for 30 s, *ii.* 15 cycles of DNA denaturation (98 °C for 10 s) followed by hybridization at 60 °C for 30 s, and DNA extension at 72 °C for *t* min, where *t* is dependent on the size to be amplified – 30 s/kb, and *iii.* DNA extension at 72 °C for 5 min. After that, the PCR product was transformed into chemically competent cells, as described in section 3.3.2.

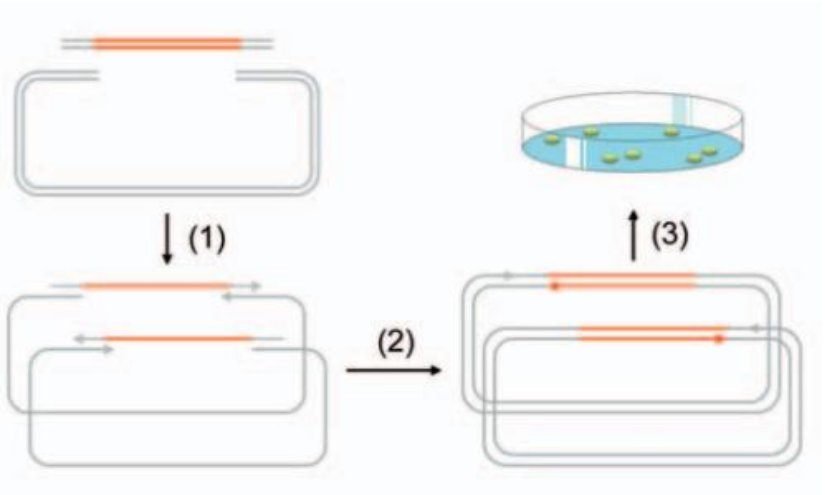


Figure 9 – A schematic diagram of gene cloning by CPEC: (1) hybridization after DNA denaturation, (2) DNA extension by polymerase, and (3) recombinant colonies transformed with CPEC reaction. From Quan and Tian, 2009.

3.3.4 Multigenic pathway construction

Following synthetic biology principles and BioBrick standards, the *ePathBrick vectors* were designed by Xu *et al.* (2012) to allow the assembly of a set of genes in three possible configurations: operon, in which all genes are transcribed and regulated together; pseudo-operon, where each gene is under the control of an independent promoter but sharing the same terminator region; and monocistronic, in which each gene have its own promoter and terminator region. Using these three configurations, gene expression in these plasmids can be differently regulated, a feature that can be very useful for metabolic engineering applications. The assembly of different configurations is possible because of the arrangement of four isocaudamer enzymes (*XmaJ*I, *Xba*I, *Bcu*I, and *Nhe*I) in the plasmids: *Xba*I right before the promoter region, *XmaJ*I right after the RBS sequence (this order is present only in pETM9 and pRSM6 plasmids), *Bcu*I just before the terminator sequence and *Nhe*I right after the terminator sequence. These

four enzymes generate the same four nucleotides cohesive ends (CTAG). However, if the digestion is done by different enzymes, the ligation product generates a sequence not recognized by any of them. In this way, the gene cluster is assembled by restriction and ligation cloning, and the enzymes choice is based on the desired configuration. The pseudo-operon configuration was selected for this work since it was reported to lead to the highest product formation when the authors constructed a flavonoid production pathway (XU *et al.*, 2012). In the pseudo-operon configuration, the gene placed near the terminator sequence is more expressed than the gene farther from it, as shown in Figure 9.

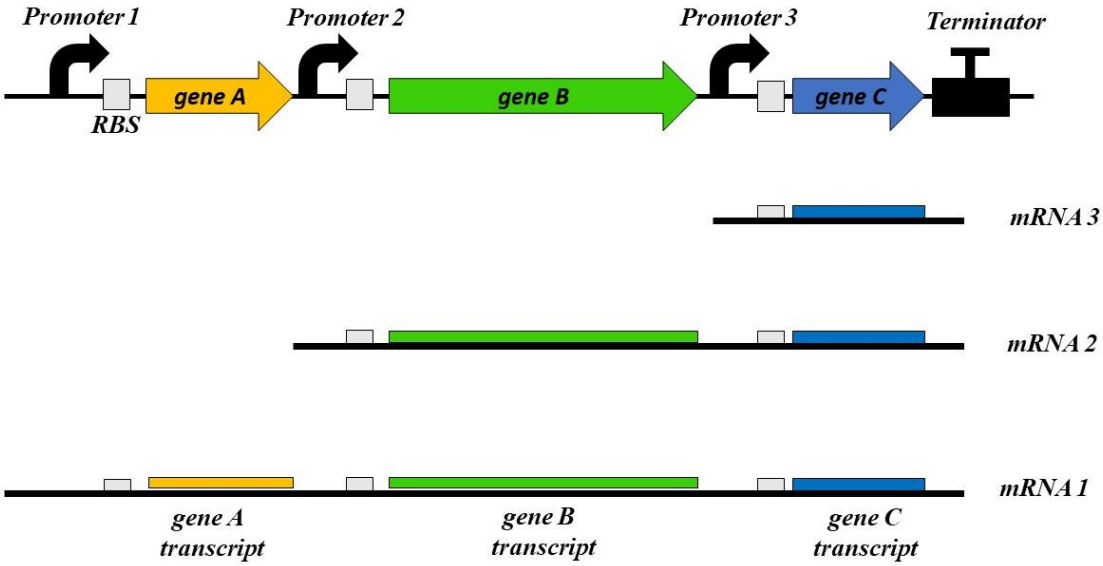


Figure 10 - Pseudo-operon gene cluster and how gene expression is regulated in this configuration based on gene proximity with terminator and mRNA transcription. RBS: ribosomal binding site.

The genes of interest in this work were assembled in two plasmids: one for the *de novo* β-alanine pathway for 3-HP production (in the pETM9 backbone) and the other for the pathway optimization (in the pRSM6 backbone). The pEbtyGpD plasmid (pETM9 harboring *pa0132*, *ydfG*, and *panD*, in this order) has the genes necessary for 3-HP production in *E. coli* cells. The *acs* gene was also added to this plasmid to test if it could prevent acetate accumulation. The other target genes (*ALT1*, *alr*, *aspC*, and *pckA*) were cloned in different combinations to test which of them could favor the 3-HP production. The cloning steps and enzymes used for each construct are shown in detail in Appendix C.

3.3.5 Positive clones screening

Positive clones, meaning *E. coli* DH5α cells carrying successfully cloned genes, were screened by PCR directly from transformant colonies or purified plasmids and by restriction digestion. All the constructs will be sent for sequencing confirmation.

In this step, PCR reactions were done in a volume of 6 µL using the GoTaq ® Green Master Mix (Promega). For endogenous cloned genes (*acs*, *ydfG*, *alr*, *aspC* and *xylR*), colony PCR was run with a pair of primers annealing the plasmid MCS sequence (Screening_MCS_F GACTCCTGCATTAGGAAGCAG and Screening_MCS_R GCTTGTCGACTCCTCCTTTCG) to avoid false positives that could be amplified from *E. coli* genomic DNA. Products from these reactions had the size of the cloned gene plus the MCS sequence from the plasmid backbone. For heterologous genes (*ALT1*, *pckA*, *panD*, and *pa0132*), PCR reactions were done with the same primers used in the cloning steps.

Positive clones were grown in 5 mL LB medium at 37 °C and 225 rpm for 16 h for plasmid extraction and cell stock preparation (720 µL of cell culture, 480 µL of glycerol 50 % v/v, stored at - 80 °C).

3.4 Site-directed mutagenesis

Site-directed mutagenesis of the *panD* and *xylR* genes in the plasmids pEpD and pRx was carried out using round the horn (RTH) PCR amplification of the plasmid, based on the work of Ochman, Gerder, and Hartl (1988). In this technique, primers are designed pointing away from each other and produce a linear amplicon of all the template plasmid that has its edges further joined together before transforming it into the cells (Figure 11). One of the primers will carry the mutagenesis in its 5' portion. The other must be phosphorylated for ligation. The 5' ends of both primers must start on a consecutive complimentary basis, so no base is left behind during PCR.

The primers used for *panD* Glu56Ser and *xylR* Pro363Ser mutagenesis are shown in Table 6 and were designed in SnapGene® v.4.2.11. Additionally, other bases were changed to create a *Sac*I restriction site inside the *panD* gene, and a *Pst*I restriction site inside the *xylR* gene without changing the polypeptide sequence and the codon bias usage. However, this step is not always necessary since this technique has a high yield of success. Anyhow, the mutant plasmid needs to be checked, and restriction digestion is an easy and rapid way of doing it before the plasmid is sent for sequencing.

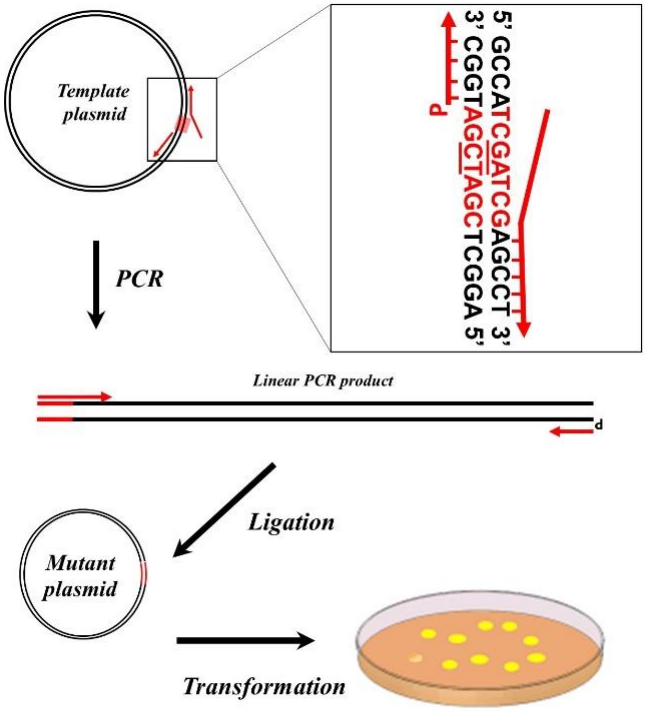


Figure 11 - The mutagenesis procedure with round the horn amplification. One of the primers has its 5' end phosphorylated for ligation, and the other contains the basis that will generate the mutant plasmid. The PCR produces a linear product that has its edges ligated before transformation into competent cells.

Table 6 - Primers used for *panD* G56S and *xylR* P363S site-directed mutagenesis through round the horn. *Sac*I and *Pst*I restriction sequences are underlined. P- stands for the phosphorylated end. Changed bases are shown in bold.

Primer	**5'-3' Sequence**
mut_*panD*_F	GG**A**GCTCGTCTG**TC**AACTTATGTCATTGTGGGCGACG
mut_*panD*_R	P-GTTGGTGATGTCTACGATGGCAAC
mut_*xylR*_F	GGTTAT**T**CATCGCTGCA**G**TATTTCTACTCTGTTTTTAAAA
mut_*xylR*_R	P-GCACATTTGCGATATCTCATTGATCGACAA

For the PCR reaction, 10 µL of Platinum™ SuperFi II PCR Master Mix (Invitrogen), 10 ng of the template plasmid (pEpD), 1 µL of primer mix solution (10 pmol/µL of each primer), and water to 20 µL were added to a 200 µL PCR microtube. The reaction was carried in a thermal cycler accordingly to the following steps: *i.* DNA denaturation at 98 °C for 2 min,

ii. 30 cycles of DNA denaturation at 98 °C for 8 s, primer annealing at 60 °C for 10 s and DNA extension at 72 °C for 3 min and *iii.* another cycle of DNA extension at 72 °C for 5 min. The PCR product was digested with 1 µL of *Dpn*I for template methylated plasmid degradation. The reaction was incubated at 37 °C for 1 hour.

After *Dpn*I digestion, 1 µL of the PCR linear product was mixed with 1 µL of T4 DNA Ligase (Thermo Scientific), 2 µL of 10X T4 DNA Ligase Buffer (Thermo Scientific), and water to 20 µL. The reaction was incubated at room temperature overnight.

For ligated plasmid transformation by heat shock, 5 µL of the ligation reaction was added to the 100 µL aliquot of chemically competent cells.

3.5 Transformation of *E. coli* K-12 MG1655 with the recombinant plasmids by electroporation

E. coli K-12 MG1655 was chosen to be the 3-HP producer strain and was transformed with the constructed plasmids by electroporation. Cells were grown in LB medium to an OD_{600nm} of 0.4 – 0.5 and then made electrocompetent by concentrating 20-fold and washing three times with ice-cold 10% glycerol.

Aliquots of 100 µL of electrocompetent cells were transformed with 100 ng of plasmid in a 0.2 mm sterile electroporation cuvette with a pulse of 2,5 kV, 200 Ω, and 25 µF (GenePulser Xcell – BioRad).

3.6 Genome editing

The genome editing procedures were performed using the No-SCAR method (Scarless Cas9 Assisted Recombineering; REISCH AND PRATHER, 2015). The plasmids used in this method, pKDsgRNAp15 and pCas9cr4, were donated by Prof. Mattheos A. G. Koffas from Rensselaer Polytechnic Institute. The plasmid maps are shown in Appendix C.

In this method, the first step is defining the Cas9 target and designing the single-strand DNA (ssDNA) that will anneal in the genomic DNA during its replication to promote gene deletion. Then, the Cas9 guide RNA (sgRNA) for each target is cloned in the pKDsgRNA plasmid. After that, recombineering is promoted by the λ-Red System (Exo, Bet, and Gam enzymes from the λ-Red phage), and Cas9 is used as counterselection for non-mutant cells. In these plasmids, the λ-Red System is induced by L-arabinose (P_{araBCD}), and Cas9 and sgRNA is induced by anhydrotetracycline (aTc) (P_{TetR}), as described in the next sections.

3.6.1 Primers and ssDNA design

For *ppc*, an 826 bp sequence in the N-terminal region was chosen to be deleted. For *ilvE*, an 890 bp sequence was selected. A 20 bp target sequence for Cas9 counterselection inside these sequences was chosen for each target gene using the online program CRISPOR (http://crispor.tefor.net/crispor.py, CONCORDET AND HAEUSSLER, 2018). These 20 bp target sequences should be followed by an NGG sequence (PAM for Cas9 from *Streptococcus pyogenes*). The sequences chosen for *ppc* and *ilvE* deletion are shown in Table 7 (see Appendix C, Figures C14, and C15 for more details). The next step is to design a single-strand DNA (ssDNA) oligonucleotide flanking the region to be deleted. The λ-red assisted ssDNA annealing to the target sequence's neighboring nucleotides will promote the gene deletion by homologous recombination. The ssDNA must have between 60 to 90 bp with at least 15 bps of homologous sequences to the target region on both the 5' and 3' ends, and $\Delta G \geq -12.5$ kcal/mol for secondary structure formation, which can be checked with mfold tool (http://unafold.rna.albany.edu/q-mfold/DNA-Folding-Form). Also, to increase the recombination efficiency, the ssDNA sequences were designed to target the lagging strand of replicating DNA: if the gene is on replichore 1 of the chromosome (> 3.92 Mb or < 1.58 Mb), the ssDNA should anneal to the 5'-3' sequence of the + strand, and if it is on replichore 2 (> 1.58 Mb and < 3.92 Mb), the ssDNA should be a copy of the 5'-3' sequence of the + strand. The ssDNA should also have two to four phosphorothioate bonds (represented by *) in its 5' end to avoid degradation by endonucleases. Table 7 shows the ssDNA sequences designed for *ppc* and *ilvE* deletion.

A pair of primers to check the knock-out of each gene was also designed (Table 6). Using these primers, the wild-type and the mutant strain amplification products should have different band sizes, easily distinguished in a 1 % agarose gel.

3.6.2 pKDsgRNAxxx cloning

The pKDsgRNA*xxx* plasmids, that is, the pKDsgRNA plasmids containing the sgRNA that targets a sequence inside the gene to be deleted, were cloned using the RTH PCR amplification procedure. The forward primer contains the 20 bp protospacer for each gene (shown in Table 7) and the sequence that anneals to the plasmid (underlined). The reverse primer anneals to the plasmid and has its 5' end phosphorylated (PTetR, PO_4-GTGCTCAGTATCTCTATCACTGA). This cloning step was carried out as described in section 3.4.

Table 7 - Primers and ssDNA sequences designed for *ppc* and *ilvE* knock-out with No-SCAR. The underlined sequence in the sgRNA primers corresponds to the sequence of the pKDsgRNAp15 plasmid.

Target	Primer	5'-3' sequence
ppc	sgRNA_*ppc*_F	AGACACCATCAAAAAAGCAGGTTTTAGAGCTAGAAATAGCAAG
	ssDNA[a]	ACCATCGACAGTTCAGAAACCAGCACCTGAATATC*T*T*T*CATTGTTCGTTCATATTACCCCAGACACCCCATCTTATCGTT
	Check_KO_F	GGATACAGGGCTATCAAACGATAAGA
	Check_KO_R	CGAGGGTGTTAGAACAGAAGTAT
ilvE	sgRNA_*ilvE*_F	TCGATGCGATGGTTTCCTCCGTTTTAGAGCTAGAAATAGCAAG
	ssDNA	T*T*A*T*TGATTAACTTGATCTAACCAGCCCCATTTATCTTCGTTTTATATTCCTTTTGCGCTCAGGCGCGGATTTGTTGTGA
	Check_KO_F	GATAACCCCGATTTCCTCATG
	Check_KO_R	GGCAATCCCATCATCCAC

3.6.3 λ-Red System-based recombineering and Cas9 counterselection

For recombineering, the plasmids pCas9cr4 and pKDsgRNA*xxx* were sequentially transformed into chemically competent *E. coli* K-12 MG1655.

One to three colonies of the strain harboring both plasmids were then incubated in 4 mL of SOB containing $MgCl_2$ 10 mM and chloramphenicol (25 µg/mL), streptomycin (50 µg/mL) at 30 °C and 225 rpm, till OD_{600nm} reached 0.4 – 0.5. Then, 0.2 % of L-arabinose was added to the culture, and cells were grown for more 15 min. Then, cells from 2 mL of culture were made electrocompetent by concentrating 20-fold with two wash steps with ice-cold 10% glycerol and the final wash step with cold water. A 100 µL aliquot was electroporated with 100 µM of the ssDNA as described in section 3.5. After electroporation and SOC addition, cells were incubated at 30 °C for 2.5 – 3.0 h, and 150 µL were plated in LB agar medium containing chloramphenicol (25 µg/mL), streptomycin (50 µg/mL), and anhydrotetracycline (100 ng/mL).

3.6.4 Recombineering confirmation and pKDsgRNA cure

Gene knock-out was confirmed by colony PCR. Positive clones were inoculated in LB medium containing chloramphenicol at 37 °C and 225 rpm overnight. Then, an aliquot of the inoculum was streaked on LB agar with chloramphenicol and incubated overnight at 42 °C for pKDsgRNA*xxx* cure, as this plasmid replication is temperature-sensitive. Up to 20 isolated colonies were selected and transferred to two other LB agar plates: one containing chloramphenicol and streptomycin and another containing only chloramphenicol. The plates were incubated overnight at 30 °C, and the colonies grown only on the LB agar chloramphenicol plate had pKDsgRNA*xxx* cured. These colonies were inoculated at 37 °C for pCas9cr4 cure or another round of genome editing. In the last case, grown colonies were transformed with the pKDsgRNA*xxx* for the new target, and this new deletion was carried out as described in section 3.5.3.

3.6.5 pCas9cr4 cure

For pCas9cr4 cure, cells were transformed with pKDsgRNAp15, which targets the p15A origin of replication of pCas9cr4. After SOC addition, cells were grown at 37 °C for 2 – 3 h, and 1 µL of aTc 100 µg/mL was added to induce Cas9 and the sgRNA. Cells were grown at 30 °C and 225 rpm for more 2 h, plated on LB agar medium containing streptomycin and aTc, and incubated at 30 °C overnight. From this Petri dish, up to 20 colonies were selected and transferred to two other LB agar plates: one containing chloramphenicol and another without any antibiotic addition. The plates were incubated overnight at 37 °C, and the colonies grown only on the LB agar plate had the pCas9cr4 cured. These colonies were then inoculated at 37 °C and 225 rpm overnight for pKDsgRNAp15 cure according to the method described before for pKDsgRNA*xxx* cure.

3.7 Shake flasks cultivation

The engineered strains were inoculated for production tests in 5 mL of LB medium at 37 °C and 225 rpm. Then, 1 mL of this inoculum was transferred into 50 mL of culture medium in 250 mL baffled flasks. Cells were grown until OD_{600nm} reached 0.8 – 1.0 when IPTG 1 mM was added for P_{lacUV5} induction. Then, they were cultivated for up to 30 h. Samples of 2 mL of

culture broth were taken periodically for OD_{600nm} measure and metabolite quantification. All experiments were performed in duplicates.

The culture medium consisted of a modified MR medium, adapted from Song *et al.* (2016), containing glucose (or xylose, or a mixture of both) 15 g/L, yeast extract 3 g/L, $NaHCO_3$ 3 g/L, $(NH_4)_2SO_4$ 9 g/L, KH_2PO_4 6,67 g/L, $(NH_4)_2HPO_4$ 4 g/L, $MgSO_4.7H_2O$ 0.8 g/L, citric acid 0.8 g/L and 0.5 % of stock trace metal solution. Trace metal solution (TMS) contains $FeSO_4.7H_2O$ 10 g/L, $CaCl_2.2H_2O$ 2 g/L, $ZnSO_4.7H_2O$ 2.2 g/L, $MnSO_4.H_2O$ 0.38 g/L, $CuSO_4.5H_2O$ 1 g/L, $Na_2MoO_4.2H_2O$ 0.14 g/L, $Na_2B_4O_7.10H_2O$ 0.02 g/L. To prevent unwanted reactions during sterilization, the medium components were separated into six solutions: glucose, magnesium sulfate, sodium bicarbonate, ammonium sulfate, yeast extract, and the sixth containing the remaining components (in which TMS should be added after dissolving all the other salts). The latter solution had its pH adjusted to 6.5 with NaOH 2 M. The sodium bicarbonate solution was sterilized by filtration. The others were autoclaved at 1 bar and 121 °C for 15 min. After cooling, the medium components were mixed and stored for use.

3.8 Analytical procedure

Cell growth was analyzed using spectrophotometry. Culture broth had its optical density measured at 600 nm (OD_{600nm}) in a Genesys 10-S UV-Vis (Thermo Scientific) spectrophotometer.

The metabolites (glucose, acetate, and 3-HP) in the fermentation broth were analyzed using High-Performance Liquid Chromatography (HPLC – Waters Co System; HPLC 510 pumps, W717 Injector, with a W410 refractive index detector and UV PDA W996 reader) with a Rezex™ ROA-Organic acid H+ (Phenomenex®) column at 65 °C. The mobile phase consisted of H_2SO_4 5 mM at 0.6 mL/min. All samples were centrifuged for 6 min at 15,800 x g and then filtered using PTFE 0.2 µm filters before injection (LIU *et al.*, 2018b).

4 RESULTS AND DISCUSSION

4.1 Heterologous route for 3-HP production from L-aspartate

E. coli is not a natural producer of 3-HP and, amongst the many metabolic routes proposed for its production in this microorganism, the β-alanine pathway is one of the most promising and was evaluated in this work. For this pathway construction in *E. coli*, the reaction converting β-alanine into malonic semialdehyde needed to be inserted into its metabolism. To that end, Song *et al.* (2016) tested different β-alanine:pyruvate aminotransferases – BAPAT – (from *Pseudomonas putida*, *Pseudomonas aeruginosa*, and *Bacillus cereus*) to catalyze this reaction and reported that the enzyme that showed the best results for 3-HP production was the BAPAT from *P. aeruginosa*. Therefore, in this work, the BAPAT enzyme from *P. aeruginosa* PA01 (encoded by its *pa0132* gene) was cloned in *E. coli* K-12 MG1655, thus constructing the β-alanine pathway for 3-HP production.

However, according to the literature, the genes for the last three reactions (L-Asp → β-ala → MSA → 3-HP) of the pathway should be overexpressed so *E. coli* could excrete this acid in the fermentation broth. Therefore, to favor the conversion of L-aspartate into β-alanine, the L-aspartate 1-decarboxylase – ADC – an enzyme from *C. glutamicum* (encoded by its *panD* gene) was chosen to be cloned based on some reports of its expression for the production of 3-HP and β-alanine (SONG *et al.*, 2015; SHEN *et al.*, 2014; QIAN *et al.*, 2018). 3-Hydroxyacid dehydrogenase – HADH – from *E. coli* (encoded by its *ydfG* gene) was chosen to be cloned and, therefore, reduce MSA into 3-HP (NAPDH-dependent reaction). This enzyme is reported as the one that yields the best results for 3-HP production (WANG *et al.*, 2014; BORODINA *et al.*, 2015; SONG *et al.*, 2016).

These three genes were assembled in the pETM9 backbone forming the pEbtyGpD plasmid, which was transformed into *E. coli* K-12 MG1655 to obtain the PS100 strain. The pseudo-operon configuration chosen for assembling the genes favors the transcription of the last genes in the cluster since those are preceded by more than one promoter sequence. The *pa0132* was chosen to be the first gene in the cluster since the BAPAT reaction produces MSA, which is toxic for the cells and should not accumulate. On the other hand, the *panD* gene was placed in the last position, right before the terminator sequence. Therefore, since ADC was reported as the pathway bottleneck (LACMATA *et al.*, 2017), the *panD* position in the plasmid should lead to a higher expression of this reaction. Figure 12 depicts the cloning confirmation of each gene in the intermediary plasmids pEbt, pEyG, pEpD, pEbtyG, and the final construct pEbtyGpD.

After the obtention of the first engineered strain, shake flask cultivation was performed in a rotary shaker to evaluate the capacity of PS100 to produce 3-HP during growth on glucose. The inoculum was grown overnight in LB medium, and a small aliquot of 1 mL was transferred to a 250 mL baffled-flask containing 49 mL of MR medium with ampicillin. The cultures were grown until the early exponential phase (OD_{600nm} = 0.8 – 1.0) when they were induced with IPTG 1 mM to activate the expression of the genes carried in the pEbtyGpD vector (the best moment of induction was evaluated, and results are available in Appendix D). Samples were taken periodically, for up to 30 h after induction, to follow cellular growth, production profile, and substrate consumption of the cultures (Figure 13).

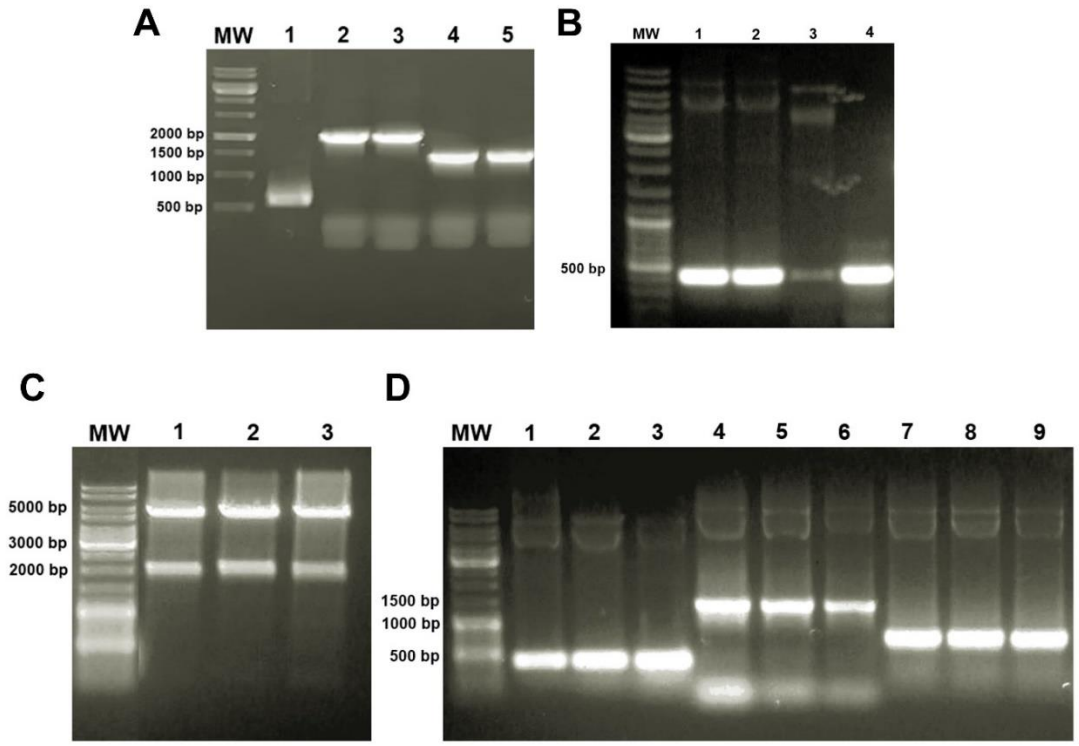


Figure 12 – Cloning of individual genes and pEbtyGpD assembly confirmed by colony PCR and restriction digestion. (A) Amplification of the MCS sequence from the plasmid pETM9 (710 bp) plus the individual cloned genes. Negative control (lane 1), *pa0132* (1347 bp, lanes 2 and 3) and *ydfG* (747 bp, lanes 4 and 5); (B) *panD* individually cloned into pETM9 (411 bp, lanes 1, 2 and 3), and positive control for *panD* amplified from *C. glutamicum* genomic DNA (lane 4); (C) pEbtyG (pETM9 carrying *pa0132* and *ydfG*) digestion with *Sal*I – expected band sizes of 5526 bp and 2011 bp; (D) pEbtyGpD (pETM9 carrying *pa0132*, *ydfG*, and *panD*). Amplification of each gene from plasmids extracted from 3 colonies (*panD*, lanes 1-3, *pa0132*,

lanes 4-6 and *ydfG*, lanes 7-9). All bands have the expected size. MW: molecular weight marker Gene Ruler DNA ladder mix (Thermo Scientific).

After 20 h of induction, all glucose was depleted and no changes in the metabolite's concentration were observed, as well as for the biomass, which had reached the stationary phase of growth after 12 h. The final titer of 3-HP produced by the PS100 in this first trial was 0.338 ± 0.044 g/L after 30 h of induction. Figure 14 shows the chromatogram profile using a refraction index detector, confirming the presence of 3-HP in the culture broth. This result is superior to the 3-HP titer of approximately 0.150 g/L reported by Song *et al.* (2016) in their first production strain expressing the BAPAT from *P. putida* (encoded by its *pp0596* gene) and 3-hydroxyisobutyrate dehydrogenase (HIBDH, for converting MSA into 3-HP) from *B. cereus* (encoded by its *bc4042* gene). In their work, these enzymes were further substituted for the enzymes used here when they assessed the best enzyme combinations for the last three reactions in the β-alanine pathway. However, this substitution was done in a strain that already had other metabolic modifications (Δ*iclR* Δ*fumA*, Δ*fumB*, Δ*fumC*, Δ*ptsG*, Δ*lacI*, *PaspA::Ptrc*, overexpressing *aspA* from *E. coli*) and that was able to produce 3.31 g/L of 3-HP. This result was an 8-fold improvement from the strain that also had the modifications above but overexpressed *pp0596* and *bc4042*.

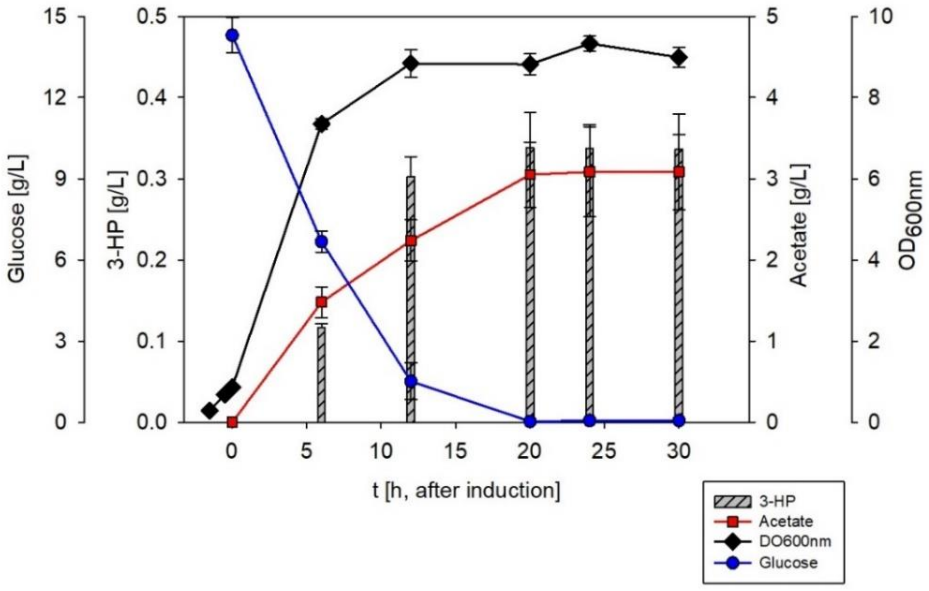


Figure 13 - Growth and 3-HP production profile during shake flasks culture of PS100 strain (*E. coli* K-12 MG1655 harboring the pEbtyGpD plasmid). Medium values are shown, and error bars represent standard deviations.

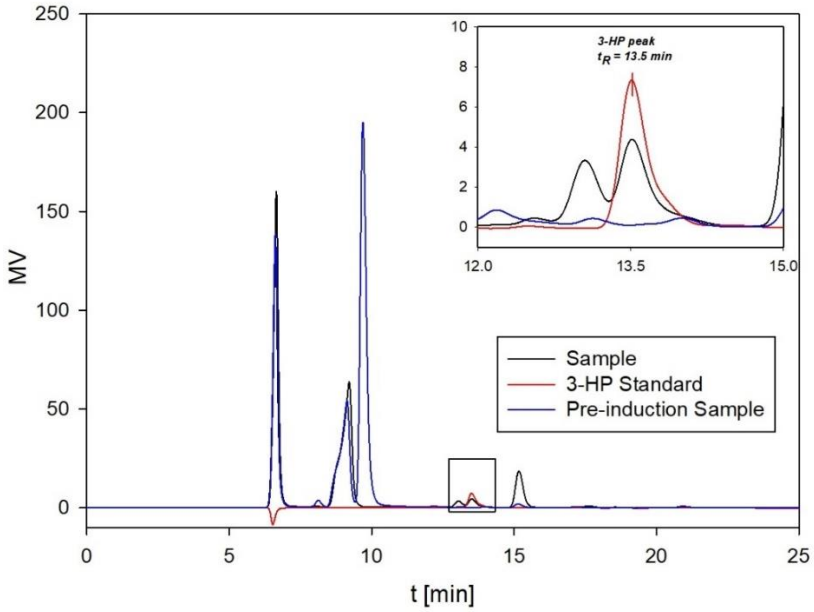


Figure 14 – Refraction index chromatogram profile of PS100 culture broth after 30 h of induction (black line) compared to the standard 3-HP compound (1.0 g/L, red line). Comparison of the 3-HP peak (residence time of 13.5 min), in detail, confirming the presence of 3-HP in the broth.

Acetate, a common fermentation byproduct of *E. coli* growth, reached 3.553 ± 0.511 g/L at the end of cultivation. As acetate formation is linked to the cell growth rate, PS100 cultivation was repeated at 30 °C and 37 °C to evaluate if a temperature decrease could alleviate this byproduct formation. The results are shown in Table 8 and, although acetate production decreased at 30 °C, so did the 3-HP titer.

Table 8 - 3-HP and acetate titer after 24 h of induction of PS100 fermentation at different temperatures.

Product	**Temperature**	
	30 °C	**37 °C**
3-HP [g/L]	0.320 ± 0.011	0.402 ± 0.007
Acetate [g/L]	3.611 ± 0.150	4.423 ± 0.420

Acetate is the main byproduct of *E. coli* growth and is commonly observed even under aerobic conditions due to an imbalance of the acetate production/uptake node. The enzymes

directly involved in the acetate metabolism are acetate kinase (ACKA, codified by the *ackA* gene), phosphate acetyltransferase (PTA, codified by the *pta* gene), and acetyl-CoA synthetase (ACS, codified by the *acs* gene). PTA converts acetyl-CoA and Pi into CoA and acetyl-P, the latter being converted by ACK into acetate and ATP. ACS converts acetate back into acetyl-CoA through an acetyl-adenylate (acAMP) intermediary (KUMARI *et al.*, 1995; KUMARI *et al.*, 2000).

Acetate accumulation is undesirable as it is harmful to cellular growth, and its formation deviates carbon flux from the substrate. Some strategies have been evaluated to prevent its accumulation, with the most common being the *pta-ackA* knock-out. However, this strategy leads to adverse effects on cell growth. Another strategy, first proposed by Lin *et al.* (2006) and applied to this work, is *acs* overexpression since the enzyme codified by this gene, ACS (E.C. 6.2.1.1), is the enzyme responsible for scavenging acetate from culture broth due to its high affinity for acetate. During metabolism overflow, triggered by the complex cAMP-CRP in the exponential growth phase, acetyl-CoA is directed to the ACKA-PTA node leading to acetate accumulation because of ACS down-regulation (NOVAK *et al.*, 2018; VALGEPEA *et al.*, 2010). Therefore, an increase in ACS expression level could bypass this effect and prevent acetate build-up.

In this work, the *acs* gene was cloned in the pETM9 plasmid by CPEC. After its cloning, the insert containing the *lacUV5* promoter, RBS sequence, and *acs* gene was placed at the beginning of the gene cluster of pEbtyGpD, thus creating the pESbtyGpD plasmid. Figure 15 shows each confirmation step of the cloning procedure (A and B: *acs* in the pETM9, and C: pESbtyGpD). This plasmid was transformed into *E. coli* cells to evaluate if the *acs* overexpression can prevent acetate accumulation and if it would benefit 3-HP production.

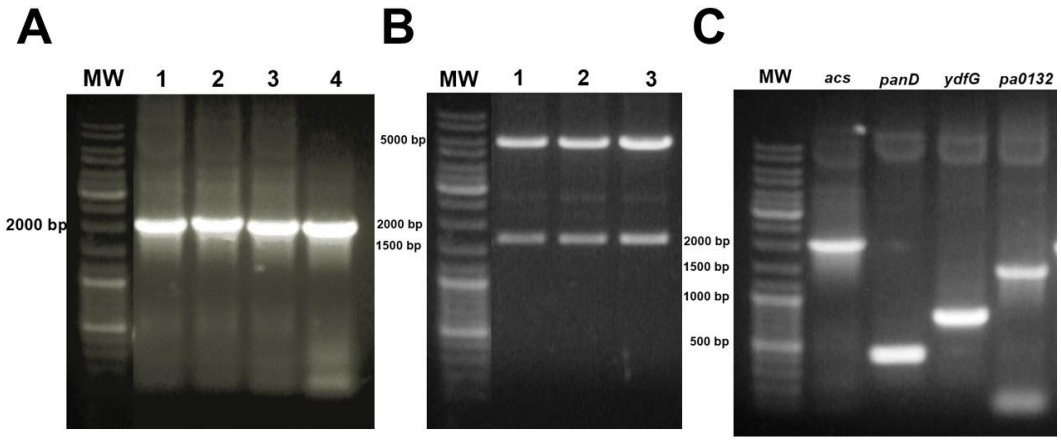


Figure 15 – Cloning of *acs* gene and pESbtyGpD assembly confirmation. (A) *acs* (1959 bp) amplification from purified pES plasmid (pETM9 harboring the *acs* gene); (B) restriction

digestion of the pES construct with *Xho*I to confirm *acs* cloning into pETM9 – expected bands with 5617 bp and 1598 bp; (C) pESbtyGpD confirmation by amplification of each cloned gene from the purified plasmid (*acs*, lane 1, *panD*, lane 2, *ydfG*, lane 3, and *pa0132*, lane 4). All bands have the expected size. MW: molecular weight marker Gene Ruler DNA ladder mix (Thermo Scientific).

Another strategy that was evaluated in this work to increase the 3-HP titer was an attempt to improve ADC activity. Zhang *et al.* (2018) showed that replacing the Glu56 amino acid residue for Ser could improve ADC activity from *B. subtilis*. Based on their work, we aimed to evaluate the effect of this same mutation on the ADC enzyme from *C. glutamicum*. The Glu56 residue is directly related to the catalytic stability of ADC from *B. subtilis*, and as this residue is conserved in the *C. glutamicum* enzyme, there was a chance of this substitution have a similar role in the latter. Hence, site-directed mutagenesis of *panD* was done in the pEpD plasmid using round the horn PCR amplification. The mutant gene was inserted in the pEbtyG plasmid, resulting in the pEbtyGpD* construct.

The *acs* gene was also inserted in this last plasmid to generate the pESbtyGpD* plasmid, which may be beneficial both for 3-HP production and acetate uptake. The confirmation steps of pEbtyGpD* and pESbtyGpD* assembly can be visualized in Figure 16.

The pEbtyGpD*, pESbtyGpD, and pESbtyGpD* plasmids were individually transformed into *E. coli* K-12 MG1655, thus creating PS200, PS300, and PS400 strains, respectively, that were evaluated for 3-HP and acetate production by shake flasks cultivation.

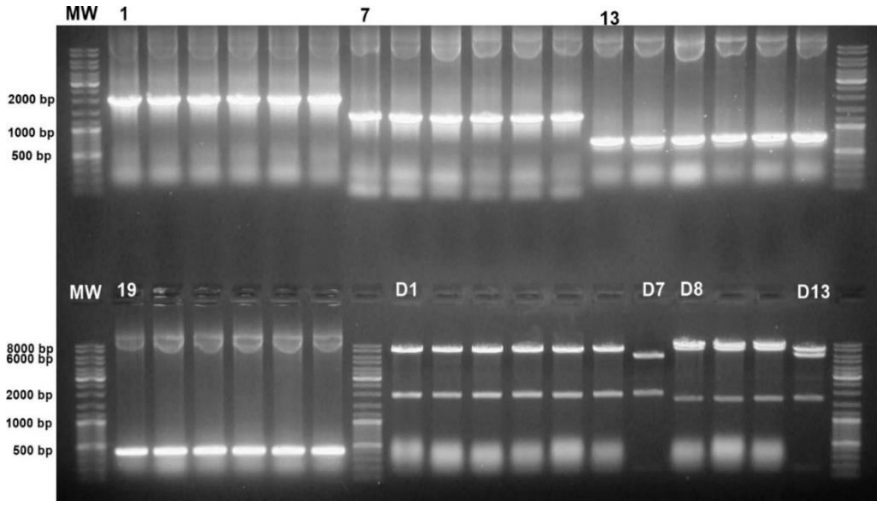


Figure 16 – Plasmids pESbtyGpD and pESbtyGpD* assembly confirmed by plasmid PCR and restriction digestion. Amplification of *acs* in pESbtyGpD (lanes 1-3) and in pESbtyGpD* (lanes

4-6¨). Amplification of *pa0132* in the pESbtyGpD (lanes 7-9) and in the pESbtyGpD* (lanes 10-12). Amplification of *ydfG* in plasmid pESbtyGpD (lanes 13-15) and in pESbtyGpD* (lanes 16-18). Amplification of *panD* in pESbtyGpD (lanes 19-21) and in pESbtyGpD* (lanes 22-24). Lanes D1-D6, restriction digestion of pESbtyGpD with *Nhe*I and *Bgl*II (expected bands of 7913 bp, 2224 bp, and 174 bp), and lane D7, restriction digestion of pEbtyGpD with the same enzyme pair for control (expected bands of 5747 bp, 2224 bp, and 174 bp). Lanes D8-D10, restriction digestion of pESbtyGpD* with *Sac*I and *Bgl*II (expected bands of 8404 bp, 1733 bp, and 174 bp), and lane D13, restriction digestion of pEbtyGpD* with *Sac*I and *Bgl*II for control (expected bands of 6238 bp, 1733 bp, and 174 bp). All bands have the expected size. MW: molecular weight marker Gene Ruler DNA ladder mix (Thermo Scientific).

Figure 17 depicts the final titer of 3-HP and acetate obtained from PS100, PS200, PS300, and PS400 shake flasks cultivations after 24 h of induction. Strains PS200 and 400 carry the E56S mutagenesis in the ADC gene from *C. glutamicum*, and strains PS300 and 400 carry a copy of *acs* gene from *E. coli* to overproduce acetyl-CoA synthetase. Compared to the PS100 strains, all other strains had their 3-HP final titer diminished, even in the strain that did not have any mutation in the ADC enzyme, PS300. However, as PS300 carries the *acs* gene, the metabolic burden of the pESbtyGpD plasmid may have played a crucial role in this titer reduction. Results from strains PS200 and PS400 indicated that the change in the Glu56 amino acid residue of ADC did not improve its stability. Actually, this mutation probably caused the opposite effect, making the enzyme structure unstable and reducing its activity towards L-aspartate decarboxylation.

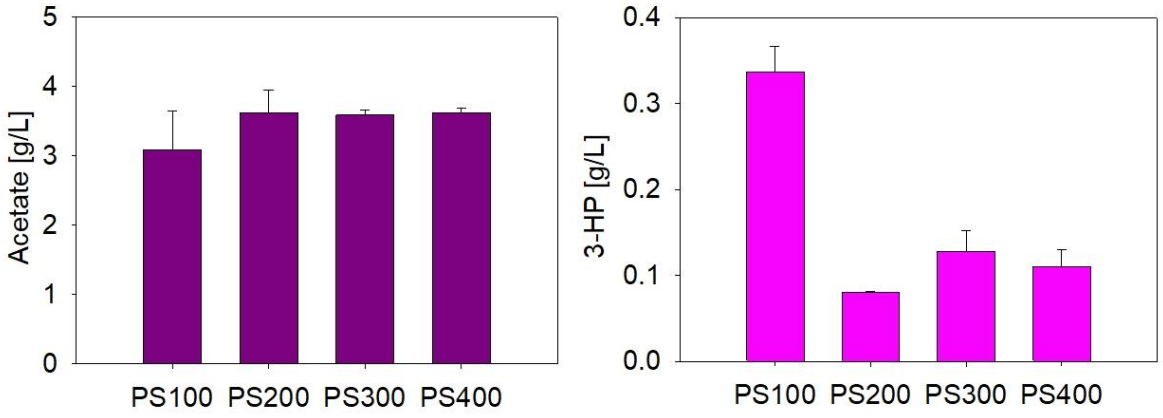


Figure 17 – Acetate and 3-HP production from PS100, PS200, PS300, and PS400 strains. Medium values are shown and error bars represent standard deviations.

In their work, Zhang *et al.* (2018) claim that the shift of a Glu56 amino acid residue for a Ser improves the stability of ADC from *B. subtilis* by increasing the polar contact of this residue with the Gln45 (Figure 18-A), besides making the entrance to the active site pocket less hindered for L-aspartate. An analysis of this pocket, within a range of 4.0 Å, revealed that this residue of Gln45 is not within this range for ADC from *C. glutamicum* (Figure 18-B). Besides, Glu56 is one of the four amino acid residues (Val23, Ile26, Thr27, Glu56) identified by Mo *et al.* (2018) that are important for the self-cleavage process at Gly24-Ser25 that ADC undergoes to be catalytically active. These four amino acids residues are present in the ADC from *B. subtilis*, but the ADC from *C. glutamicum* does not have the Ile26 residue. Therefore, the substitution of Glu56 may be more critical for the *C. glutamicum* enzyme compared to the same modification in the ADC from *B. subtilis*, so the Glu56Ser modifications may have interfered in the interactions within the polypeptide structures, making the enzyme more unstable and less catalytically active.

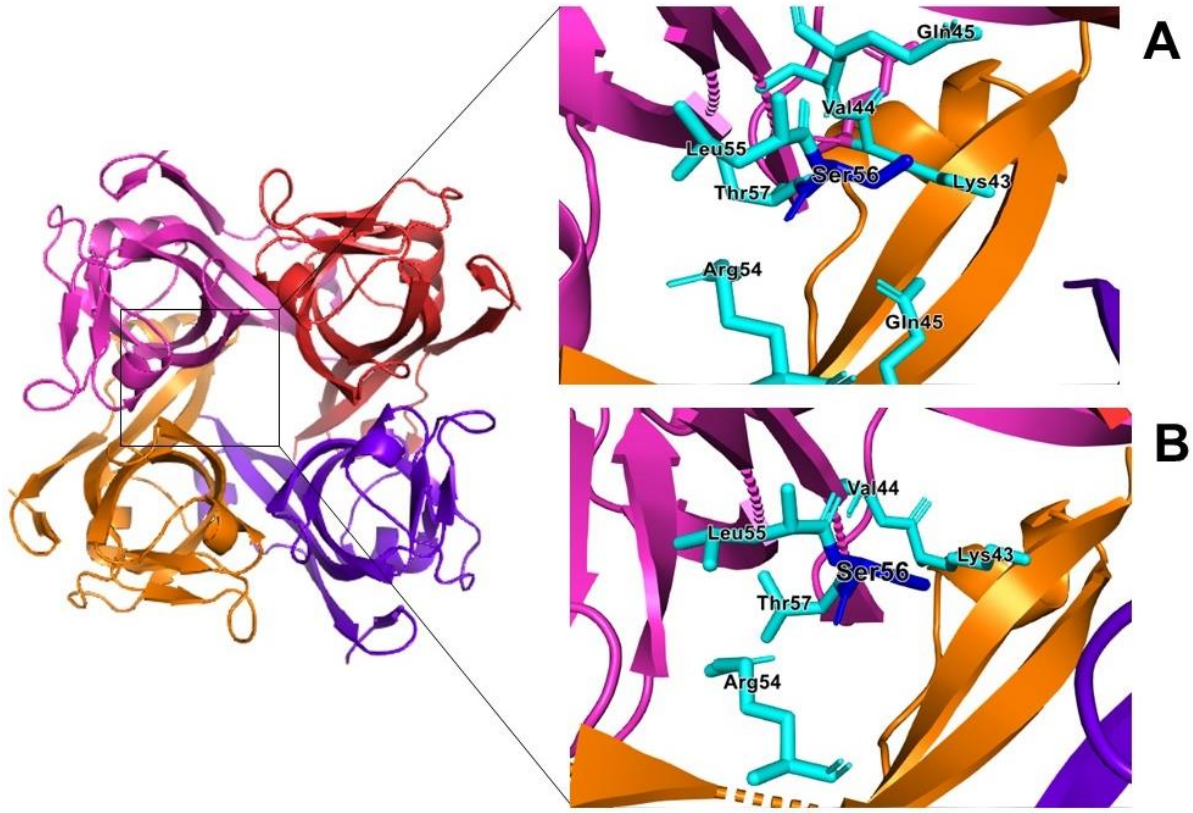


Figure 18 – ADC homotetramer from *Bacillus subtilis* and Ser56 pocket of (A) ADC from *Bacillus subtilis* (B) ADC from *C. glutamicum*, in detail. The structure models of these ADCs were constructed by the homology modeling of ADC from *E. coli* (PDB ID: 1PQE) using

SWISS-MODEL (https://swissmodel.expasy.org/interactive) workspace and SAVES v6.0 from UCLA (https://saves.mbi.ucla.edu/) for quality checking, as described by Zhang *et al.* (2018).

The attempt to reduce acetate accumulation in the fermentation broth by cloning the acetyl-CoA synthetase (ACS) did not work. The strains PS300 and PS400 that harbor pESbtyGpD and pESbtyGpD* plasmids containing the *acs* gene produced 3.553 ± 0.511 g/L and 3.655 ± 0.141 g/L of acetate, respectively, while PS100 produced 3.084 ± 0.460 g/L. In the work of Lin *et al.* (2006), the *acs* overexpression produced a 9-fold increase to the ACS catalytic activity, which resulted in lower acetate accumulation and faster acetate assimilation by *E. coli* MG1655 cells. As our fermentations with *acs* overexpression had a different outcome from Lin *et al.* (2006) in the 3-HP production strains, we decided to transform a plasmid carrying only the *acs* gene (plasmid pES) into *E. coli* K-12 MG1655 cells, to assess the acetate production of this new strain (named KS) compared to the WT strain. KS and WT cells were grown in glucose for 30 h and their acetate production was evaluated (Figure 19).

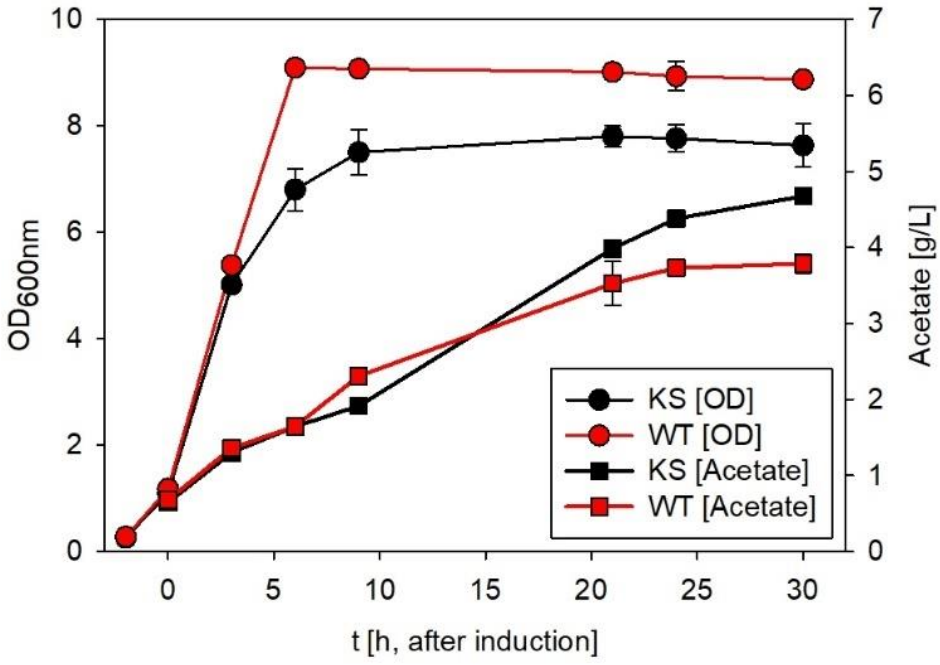


Figure 19 – Cell growth profile and acetate production of KS (*E. coli* K-12 MG1655 harboring pES) and WT (*E. coli* K-12 MG1655) strains in glucose. Medium values are shown and error bars represent standard deviations.

The KS strain did not have its acetate production decreased by *acs* overexpression, such as PS300 and PS400 strains. Actually, the KS strain produced even more acetate than WT, 4.674 ± 0.039 g/L against 3.786 ± 0.129 g/L. As the pES plasmid and was not sequenced and no other test were carried out to analyze ACS activity yet, no further conclusions could be taken from these results.

Finally, as the best production strain so far was PS100, it was selected to be further engineered to improve the 3-HP final titer.

4.2 Engineering flux from phosphoenolpyruvate towards L-aspartate

To channel the carbon flux to the L-aspartate branch towards 3-HP production, the first two reactions of the β-alanine pathway might be engineered. As the pathway precursor can be either phosphoenolpyruvate or pyruvate, the classical approach adopted by Song *et al.* (2016) and Borodina *et al.* (2015) was to overproduce phosphoenolpyruvate carboxylase (PPC) and pyruvate carboxylase (PYC), respectively. However, a third option that has not been explored so far to produce 3-HP, could couple the conversion of PEP into OAA with ATP formation, using the gluconeogenic enzyme phosphoenolpyruvate carboxykinase (PCK). According to Kumar, Ashok, and Park (2013), this pathway is among the most promising for 3-HP production, as it is redox-neutral and has a net ATP yield of 1 mol of ATP per 3-HP mol produced, being energetically and thermodynamically favorable.

PCK catalyzes the interconversion of OAA and PEP in the gluconeogenic pathway. In *E. coli*, PCK has OAA decarboxylation as the preferential direction because of the enzyme's lower affinity for carbonate than PPC (irreversible), which is the other enzyme responsible for this anaplerotic reaction. However, in some microorganisms such as *Mannheimia succiniciproducens*, *Actinobacillus succinogenes*, and *Anaerobispirillium succiniciproducens*, PCK is the only anaplerotic enzyme responsible for the carboxylation of PEP into OAA. Some works have shown promising results using the PCK instead of PPC when the medium is supplemented with bicarbonate or PPC is inactivated (MENG *et al.*, 2016), and these strategies were chosen to be evaluated in this work.

The *pckA* gene from *A. succinogenes*, encoding its PCK, was cloned in the pRSM6 vector (Figure 20 – A and B), thus generating the pRk plasmid. This construct was transformed into the PS100 strain to evaluate the pathway with PCK overexpression for 3-HP production, creating the PS150 strain. Two alternatives were assessed: overexpression of PCK from *A. succinogenes* with and without inactivation (knock-out) of PPC in *E. coli*.

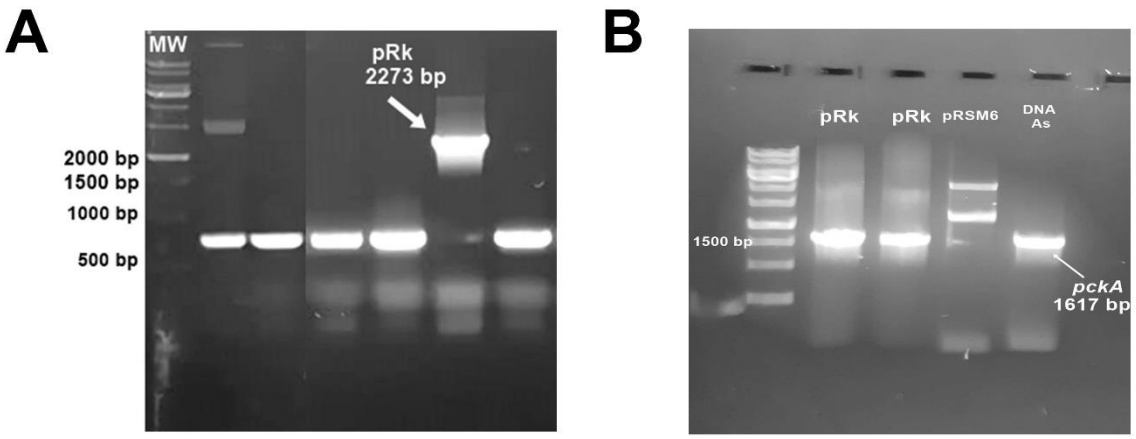


Figure 20 - Cloning of *pckA* confirmation by PCR: (A) amplification of the MCS sequence from the plasmid pRSM6 (710 bp) plus the individual *pckA* cloned gene (1647 bp) forming the pRk plasmid. (B) amplification of *pckA* gene by PCR with pRk as template and from *A. succinogenes* DNA (lane 4, after MW). All bands shown are in agreement with simulations done in SnapGene® v.4.2.11, therefore having the expected size. MW: molecular weight marker 1-kb DMA ladder (Sinapse).

The knock-out of PPC, encoded by the *ppc* gene in *E. coli*, was performed using the No-SCAR protocol for genome editing based on homologous recombination using the λ-Red system (recombineering) and the Cas9 from *Streptococcus pyogenes* for counter-selection. Recombineering is promoted by three proteins from the λ-Red prophage: Gam, Bet, and Exo (the Red system). Gam acts inhibiting RecBCD exonuclease V from the host (responsible for linear DNA degradation) so that Bet and Exo can promote the homologous recombination of the host DNA. This homologous recombination incorporates an oligonucleotide (ssDNA) into the genomic DNA. This ssDNA carries the desired genomic modification (point mutations, insertions, or deletions) and is transformed into the cell, binds to Bet and anneals with the genomic DNA. After DNA replication and cell division, one daughter cell carries the mutated genome and the other the wild type. Following recombineering, Cas9 is expressed in both cells and targets the genomic DNA sequence without the desired modification. Cas9 is guided by sgRNA to any target sequence in the DNA, known as protospacer, that anneals with sgRNA and is followed by a protospacer adjacent motif (PAM sequence). This sgRNA is designed so that Cas9 targets the DNA from non-mutant cells. In the DNA protospacer, Cas9 causes a DNA double-strand break (DSB) that eventually leads to cell death (REISCH AND PRATHER, 2015; REISCH AND PRATHER, 2017).

Before the recombineering step, the cells must be transformed with two plasmids of the No-SCAR system. The first is responsible for the tightly regulated Cas9 expression (pCas9cr4, Cm^R), and the other is responsible for the λ-Red System and sgRNA expression (pKDsgRNA, $Spec^R$, and $Strep^R$) (Figure 21-A). Cas9 and sgRNA expression are controlled by the P_{TetR} promoter and induced only in the presence of aTc (anhydrotetracycline), which binds to the repressor TetR and releases P_{TetR} for RNA polymerase binding. The λ-Red System expression is controlled by the P_{araBAD} promoter. Both are low copy number plasmids since λ-Red System and Cas9 expression are toxic for the cells. After cell transformation with pCas9cr4 and pKDsgRNA, λ-Red System expression is induced by L-arabinose. Then, cells are transformed with the oligonucleotides that will be incorporated into the chromosome by Bet. After recombineering, Cas9 and sgRNA expression are induced, and the complex Cas9/sgRNA causes non-mutant cell death (Figure 21-B). When using No-SCAR for genome editing, oligonucleotides and the sgRNA sequence must be carefully designed so that Cas9 only causes DSB in the wild type DNA.

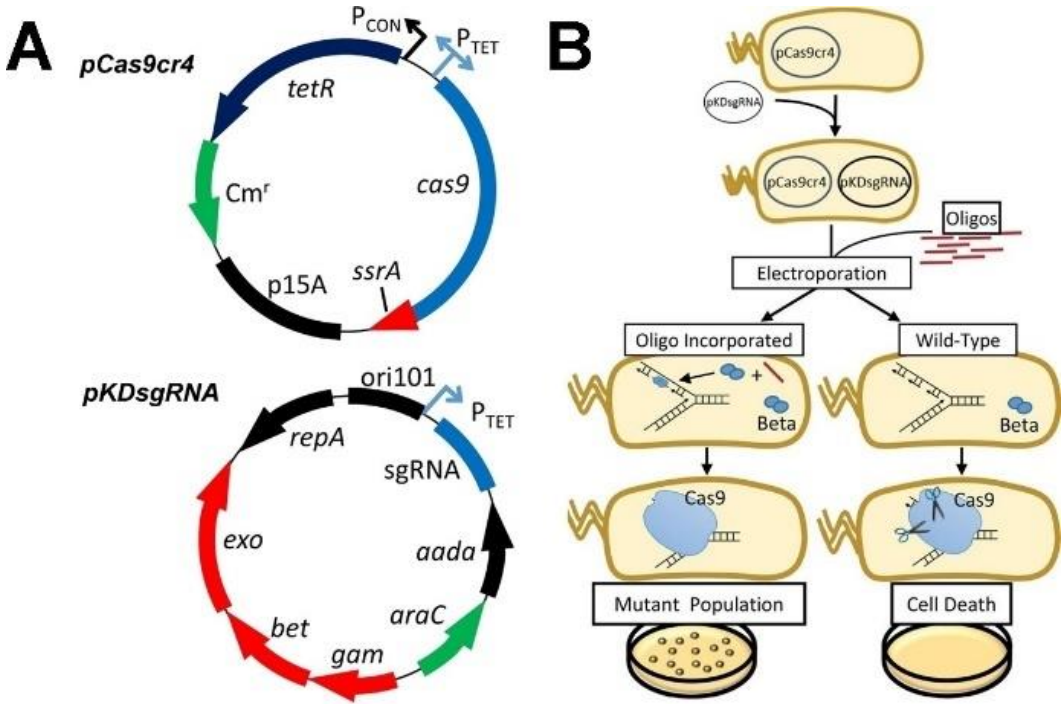


Figure 21 - The No-SCAR system plasmids pCas9cr4 and pKDsgRNA, for Cas9/sgRNA complex and Red-system expression (A) and workflow (B). Adapted from Reisch and Prather (2015) and Reisch and Prather (2017).

For *ppc* gene knock-out (K.O.), the oligonucleotide was designed with two homology arms that anneal to the *ppc* gene flanking sequences. However, since recombineering efficiency decreases with increasing deletion size, the oligo for *ppc* K.O. was designed to delete only a part of the gene N-terminal region (826 bp deletion out of 2652 bp, shown in Figure 22) and also inserting a stop codon inside the gene sequence.

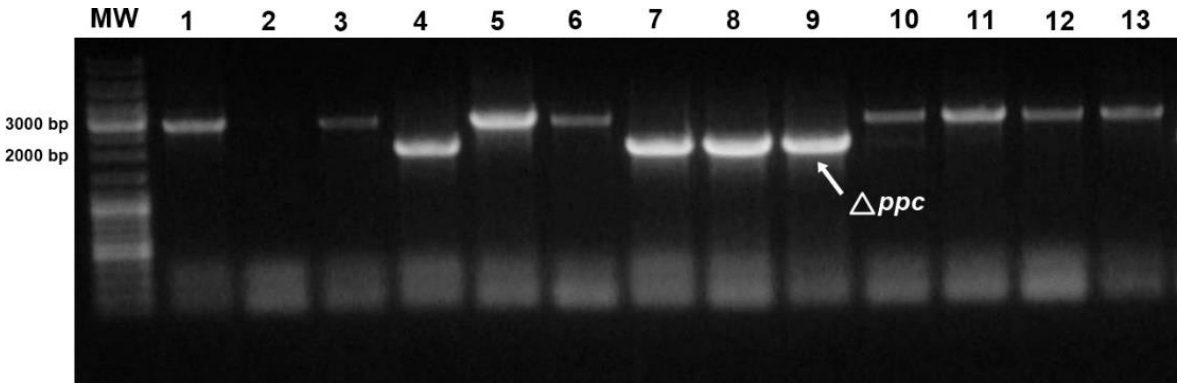


Figure 22 – Knock-out of *ppc* gene in *E. coli* K-12 MG1655 genome by applying No-SCAR technique. Amplification of the gene by colony PCR. Negative control (WT, lane 13), mutant strains (lanes 4 and 7-9), and wild-type strains (lanes 1-3, 5, 6, and 10-12). The primers used for screening were design to amplify a 2771 bp region in the wild-type strain and a 1945 bp region in the *ppc* mutant strains. MW: molecular weight marker Gene Ruler DNA ladder mix (Thermo Scientific).

After the *ppc* knock-out that created the PS001 strain (*E. coli* K-12 MG1655 Δ*ppc*), the pEbtyGpD and pRk plasmids were transformed into this mutant strain to create PS151. Therefore, to assess if PCK overexpression can redirect carbon flux from PEP towards OAA, and ultimately 3-HP, and if PPC must be inactive to favor PCK reaction, PS150 and PS151 were tested in shake flasks cultivations as described before.

However, before PS150 and PS151 fermentation trials, the experiments were suspended due to the COVID-19 pandemic. After returning to laboratory activities, we had to go on cloning other targets before assessing the 3-HP production from PS150 and PS151 strains, as the HPLC laboratory remained closed for a much longer period. Based on literature reports suggesting that PCK overexpression could benefit the 3-HP production, we decided to test the next targets along with PCK.

Hence, we constructed the pRak plasmid that also overexpressed the *aspC* gene from *E. coli* besides *pckA* from *A. succinogenes.* The *aspC* gene encodes the L-aspartate

aminotransferase – AAT – that converts OAA into L-aspartate. Piao *et al.* (2019) reported the overexpression of AAT from *E. coli* to catalyze this reaction and showed that although this enzyme can act on a broad range of substrates leading to byproducts formation, this can be circumvented if high amounts of oxaloacetate are present. The *aspC* gene was cloned in the pRSM6 to construct the pRa plasmid (Figure 23 – A and B). We hypothesized that the PCK overexpression would increase the carbon flow towards OAA, which could help overcome this enzyme's lack of specificity. Additionally, as will be discussed in section 4.3, other optimization strategies applied to this study may significantly increase the L-glutamate pool (the amino donor for OAA in the AAT reaction), thus contributing to driving the AAT reaction in the desired direction.

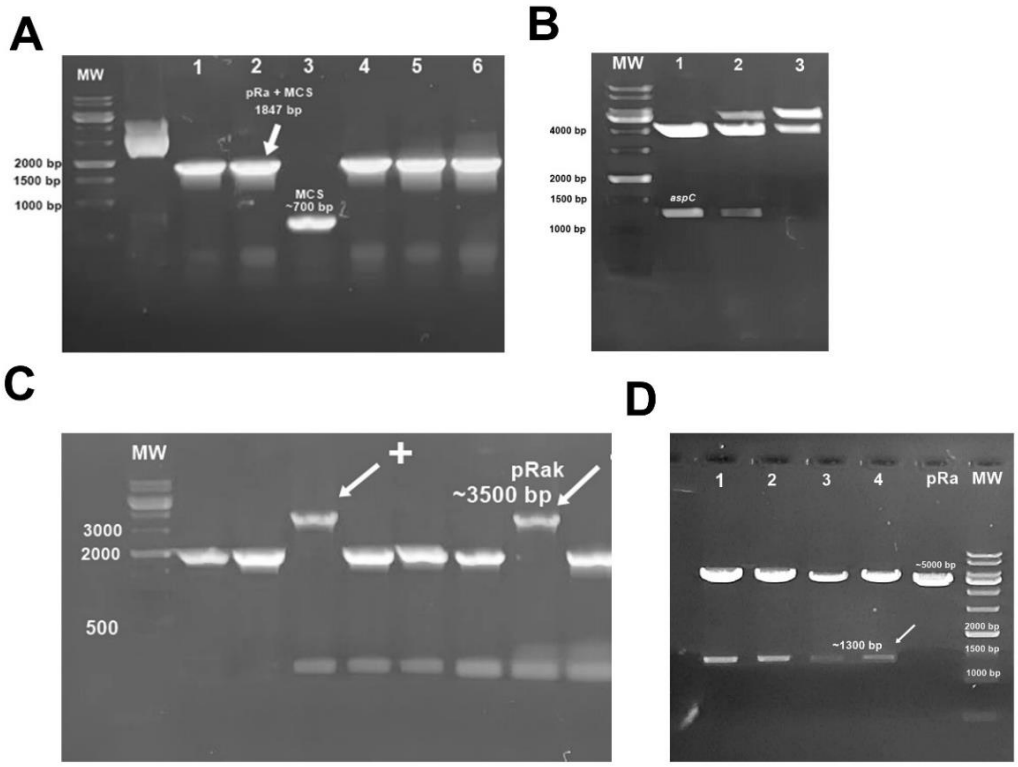


Figure 23 – Cloning of individual *aspC* gene and pRak assembly confirmation by colony PCR and restriction digestion. (A) Amplification of the MCS sequence from the pRSM6 plasmid (710 bp, lane 3) plus the *aspC* cloned (lanes 1, 2, 4-6). (B) Restriction digestion of pRa with *Nde*I and *Xho*I – expected band sizes of 1193 bp and 3818 bp. (C) Amplification of the MCS region in pRak plasmid (pointed, expected size of 3661 bp). (D) Restriction digestion of pRak with *Nde*I to confirm *aspC* and *pckA* assembly – expected size of 5437 bp and 1388 bp. All bands have the expected size. MW: molecular weight marker 1-kb DMA ladder (Sinapse).

With that in mind, the pRak plasmid was constructed in the pseudo-operon configuration placing the *pckA* gene in the last position (closest to the terminator) for higher PCK activity, and therefore, favoring the carbon flux from PEP to OAA. This plasmid was transformed into the PS100 and PS001 (along with pEbtyGpD for this later) to test if the AAT overexpression could improve its capacity to produce 3-HP, creating the PS160 (PS100 harboring pRak) and PS161 (PS001 harboring pEbtyGpD and pRak) strains.

Figure 24-A shows the growth profile of PS150, PS151, PS160, and PS161 strains. The strains PS151 and PS161, which had the *ppc* gene knocked-out, grew poorly, reaching a maximum OD600nm of 2. These strains also did not produce 3-HP. As PEP carboxylation is an important anaplerotic reaction, one could expect that its deletion would have harsh effects on cell growth. The fact that PS151 and PS161 had their growth diminished by almost 3-fold compared to the strains harboring the same plasmids but with no inactivation of PPC was indicative that PCK overexpression could not restore the effect of PPC deletion. Hence, to further test this hypothesis, we transformed the mutant strain (PS001) strain with pRk, creating the PS051 strain (*E. coli* K-12 MG1655 Δ*ppc* harboring pRk), and compared its growth with the wild-type and PS001 strains (Figure 24-B).

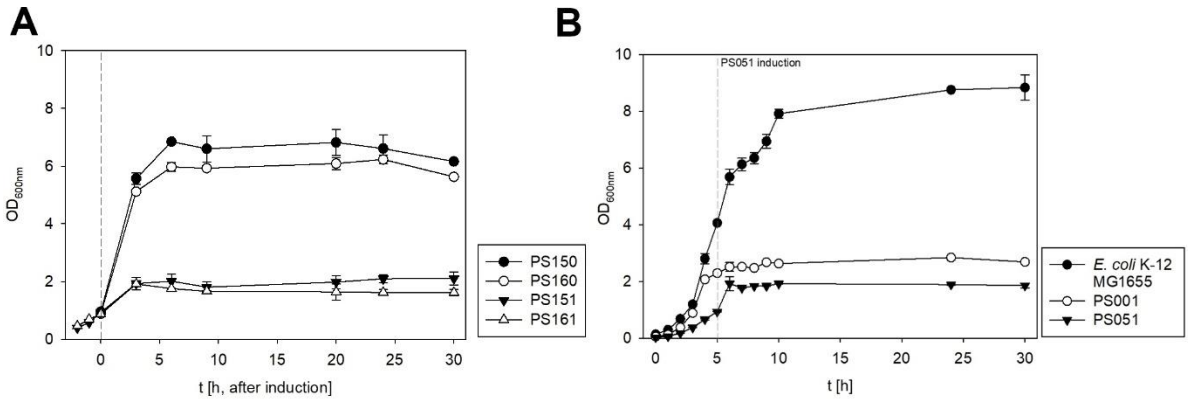


Figure 24 – (A) Growth profile of PS150 (*E. coli* K-12 MG1655 harboring pEbtyGpD and pRk), PS151 (*E. coli* K-12 MG1655 Δ*ppc* harboring pEbtyGpD and pRk), PS160 (*E. coli* K-12 MG1655 harboring pEbtyGpD and pRak), and PS161 (*E. coli* K-12 MG1655 Δ*ppc* harboring pEbtyGpD and pRak) strains. (B) Growth profile of *E. coli* K-12 MG1655 (wild-type strain), PS001 (*E. coli* K-12 MG1655 Δ*ppc*), and PS051 (*E. coli* K-12 MG1655 Δ*ppc* harboring pRk). The dashed line represents induction with IPTG 1 mM. Medium values are shown and error bars represent standard deviations.

The growth profile of PS051 further confirmed that *pckA* overexpression could not restore the negative effect of inactivating PPC. Both PS001 and PS051 strains grew poorly even after induction of *pckA* expression. The effect of *ppc* inactivation was even more evident in PS051, possibly because of the burden of harboring a high copy number plasmid with defective growth. These results are in disagreement with some previously reported works. Kim *et al.* (2004) verified that succinate production could only be enhanced in an *E. coli* K-12 strain overexpressing PCK from *A. succinogenes* after *ppc* gene deletion. Meng *et al.* (2016) replaced *ppc* for *pckA* gene from *A. succinogenes* and obtained an improvement over specific growth rate and glucose uptake rate in an *E. coli* strain engineered for succinate production. Liu *et al.* (2012) and Jiang *et al.* (2014) reported that overexpression of PCK from *Bacillus subtilis* 168 in an *E. coli ΔppcΔldhAΔpflB* strain restored cell growth and improved succinate production. These works show that PCK can be recruited for PEP carboxylation coupled with energy production in *E. coli* and, therefore, replace PPC.

In this work, *pckA* overexpression did not improve the 3-HP final titer after 24 h of induction. On the contrary, the PS150 strain produced 0.149 g/L ± 0.005 g/L of 3-HP. Therefore, we hypothesized that, in our case, PCK is favoring OAA conversion back into PEP (the gluconeogenic direction). This increase in the PEP pool of PS150 in comparison to PS100 may be the reason why PS150 produced almost twice acetate than PS100 (6.386 ± 0.066 g/L). The PS160, which overexpress *aspC* gene besides *pckA,* had an even lower 3-HP final titer, reaching 0.079 ± 0.004 g/L after 24 h of induction. As the PCK reaction seems to be converting OAA into PEP, we suppose that the AAT reversible transaminase reaction is possibly consuming L-aspartate to replenish the OAA pool, impairing the 3-HP production.

4.3 *In silico* prediction of non-intuitive metabolic targets for 3-HP production

The *E. coli* metabolism can produce 3-hydroxypropanote through the uracil degradation pathway, in which this pyrimidine base is reduced to 3-aminoacrylate, CO_2, and ammonia. The intermediate 3-aminoacrylate is further converted to 3-hydroxypropanoate by aminoacrylate hydrolase and 3-hydroxyacid dehydrogenase (HADH, NADPH-dependent). This pathway was first reported by Loh *et al.* (2006) and is present in the *i*ML1515 model (see uracil degradation pathway in Appendix A). However, this is a long pathway to produce 3-HP from sugars, and a shortcut can be taken by adding the exogenous reaction that converts β-alanine into malonic semialdehyde. This reaction, named R_APATr in the SBML language and corresponding to the BAPAT reaction, was added to the model (Figure 25), generating its modified version,

containing the β-alanine pathway for 3-HP production (recalling: PEP → OAA → L-asp → β-ala → MSA → 3-HP).

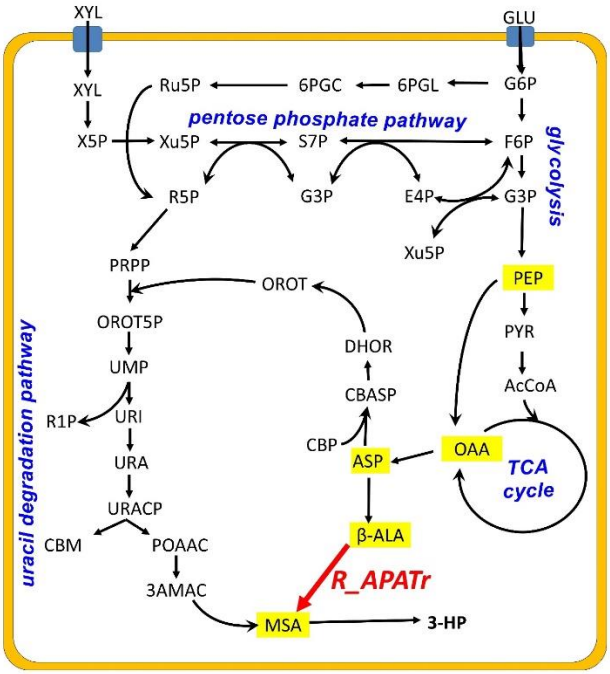


Figure 25 – Insertion point of R_APATr (red arrow) in the central metabolism of *E. coli* and its integration to the uracil (URA) degradation pathway for 3-HP production. The established β-alanine pathway for 3-HP production is highlighted in yellow.

Six simulations were run in *Optflux* to estimate the maximum theoretical production of 3-HP from glucose or xylose in the original and modified version of the model and the maximum growth rate using different carbon substrates (Table 9). For simulations 1 and 2, the biomass formation was maximized as the objective function – OF – in the original model and, for simulations 3-6, the 3-HP exchange reaction was set as OF, in the modified model.

The biomass growth rate (μ) was found to be approximately the same for the simulated growth on glucose and xylose for the same substrate uptake rate on a C-mol basis. The minimal difference observed is due to the different assimilation pathways. Gonzalez, Long, and Antoniewicz (2017) measured the uptake flux for glucose and xylose in *E. coli* BW21135 grown aerobically using C^{13}-MFA and found the values of 8.8 ± 0.5 and 9.5 ± 0.5 mmol.$gDW^{-1}.h^{-1}$ for each sugar, respectively. These values are essentially the same on a C-mol basis (considering the standard deviation), agreeing with the simulations. However, they did not reflect similar values for μ, like in the *in silico* analysis, being experimentally measured as 0.70 ± 0.01 and 0.50 ± 0.02 for growth on glucose and xylose, respectively. This difference between

simulated and experimental results is expected and comes from the limitations of GEMs, as these metabolic models do not account for kinetic and regulatory aspects.

The maximum flux for 3-HP production was approximately the same for both carbon sources. However, much higher fluxes were observed for the modified model, indicating that the heterologous β-alanine pathway is a better alternative to produce 3-HP in *E. coli* compared to the endogenous uracil degradation pathway. Also, for simulations 3-6, no other byproduct was formed, only 3-HP and CO_2. This later increased more than 6-fold for simulations 3 and 4 compared to simulations 5 and 6. We also performed two other simulations to assess the maximum theoretical production of 3-HP through the β-alanine pathway considering NADH as a cofactor for the last reaction. These two other simulations had only a slight increase over 3-HP production flux, 19.2359 and 19.1609 mmol.gDW^{-1}.h^{-1} for glucose and xylose conditions, respectively.

Table 9 - Simulations results from the original and modified *i*ML1515 model to assess the biomass growth rate and 3-HP production.

Sim. #	**Goal**	**Environmental Conditions***	**Biomass growth rate [h^{-1}]**	**3-HP production [mmol.gDW^{-1}.h^{-1}]**
1	Assess biomass growth rate	Glucose	0.8770	-
2		Xylose	0.8657	-
3	Assess 3-HP maximal production in the original model	Glucose	-	11.7439
4		Xylose	-	11.6527
5	Assess 3-HP maximal production in the modified model	Glucose	-	18.6171
6		Xylose	-	18.5814

*Glucose and xylose uptake rate set as -10 and -12 mmol.gDW^{-1}.h^{-1}, respectively. Under xylose growth, the reaction of glucose uptake (R_EX_glc__D_e) was inactivated setting its lower bound to 0 mmol.gDW^{-1}.h^{-1}).

To further increase the 3-HP production in *E. coli*, several optimizations were performed using the Evolutionary Optimization tool of *Optflux*. The objective was to find a set of modifications that could improve 3-HP production coupled with biomass growth. To do so, the

optimization was done setting the option BPCY (biomass coupled yield) as the OF. Although each optimization carried 30,000 iterations, *Optflux* has an algorithm that simplifies the final result and, therefore, each optimization returned a set of approximately 100 values of BPCY. Thus, all the generated data were screened to find the best suitable metabolic targets, prioritizing the highest BPCY values, the biological meaning of the targets, the feasibility of the proposed modifications, and the frequency that a given modification appeared. Also, to each set of ≤ 10 metabolic modifications that corresponded to a particular BPCY value, mutant simulations were performed assessing combinations of subsets inside them, excluding those modifications that would not fit in one of the screening parameters. For example, in many optimized conditions, the reaction R_ATPM appeared as a suggestion for improvement. However, this reaction is the ATP maintenance requirement of the model, and no direct biological target can be considered for this prediction.

The best set of metabolic modifications identified by the evolutionary algorithm was a combination of L-alanine aminotransferase reaction underexpression (R_ALATA_L, ALAT, which had its flux multiplied by 0.03125), alanine racemase reaction overexpression (R_ALAR, ALAR, which had its flux multiplied by 32.0), and L-valine transaminase reaction knock-out (R_VALTA, VALTA, which had its flux multiplied by 0.0). These reactions are related to the β-alanine pathway in its two transamination steps: AAT and BAPAT reactions (Figure 26). Although the optimization results pointed to the underexpression of ALAT, this modification will be referred to as overexpression from now on since the way the reaction is written in the model and in the databases is inverted (pyruvate + L-glutamate ↔ L-alanine + α-ketoglutarate, for the model representation). Both modifications (either underexpression in the model or in vivo overexpression) should have the same effect of increasing pyruvate and decreasing L-alanine pools, whose implication will be further discussed. Together, these three modifications account for new non-intuitive metabolic targets that have great potential to further improve 3-HP production through the β-alanine pathway in *E. coli* cells.

The mutant strain with the three aforementioned modifications was able to produce 3-HP coupled with its growth when simulations were run with biomass as OF, indicating that the modified strain produces 3-HP as a byproduct of cellular growth. No other byproduct was predicted in the simulation. The μ values for the mutant strain were slightly lower than the wild-type simulations (0.7878 and 0.7778 h^{-1} for growth with glucose and xylose, respectively) due to the carbon flux deviation to product formation. 3-HP exchange flux values were 1.9116 $mmol.gDW^{-1}.h^{-1}$ when glucose was set as carbon source and 1.8870 $mmol.gDW^{-1}.h^{-1}$ for xylose. Although they account for a relatively small product flux when compared to the maximum

exchange values from simulations 5 and 6 (Table 9), it is believed that these modifications will increase the product formation in an engineered production strain with other metabolic modifications.

The genetic targets chosen from the *in silico* optimization results (Figure 26) directly impact L-glutamate and L-alanine intracellular pools, affecting the equilibrium of the transamination reactions present in the β-alanine pathway. The ALR and ALAT overexpression may reduce the L-alanine pool, while VALTA knock-out may increase the L-glutamate pool. Together, these modifications may favor AAT and BAPAT reactions towards the 3-HP formation. Moreover, ALR overexpression favors L-alanine isomerization to D-alanine, which can be further converted into pyruvate by D-amino acid dehydrogenase. This reaction had its flux increased from 0 to 1.1357 mmol.gDW^{-1}.h^{-1} in the mutant simulations (compared to the wild-type).

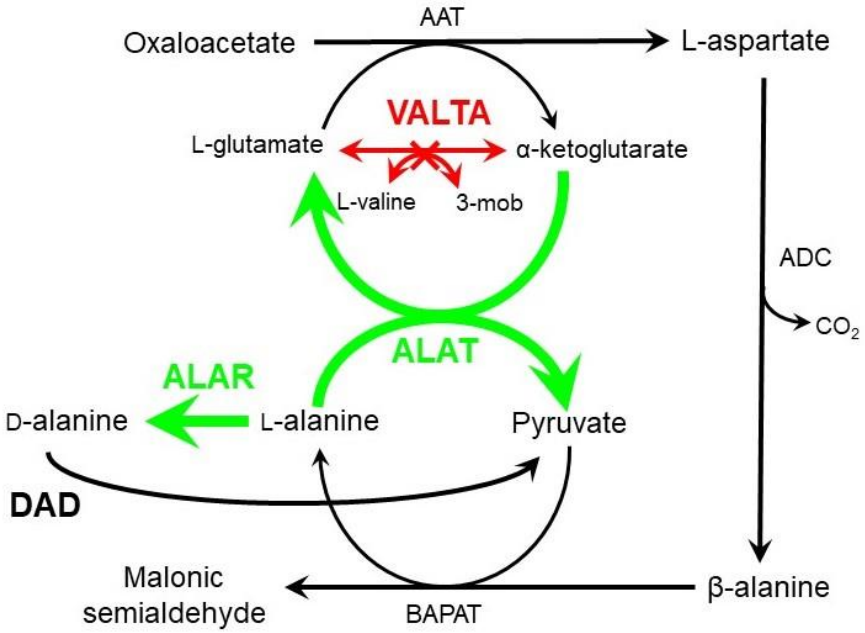


Figure 26 - Target reactions identified by the evolutionary algorithm in *Optflux* and their relation to the β-alanine pathway for 3-HP production. AAT, L-aspartate aminotransferase; ADC, L-aspartate 1-decarboxylase, BAPAT, β-alanine:pyruvate aminotransferase; VALTA, L-valine transaminase; ALAT, L-alanine aminotransferase, ALR, alanine racemase and DAD, D-amino acid dehydrogenase. Thicker green arrows stand for reaction overexpression, and red arrows with the red X stand for reaction deletion.

Figure 27 depicts other changes in the carbon flux of the β-alanine pathway reactions and the TCA cycle for the mutant strain. It can be seen that the first two reactions of the β-alanine pathway, PPC and AAT reactions, had their fluxes increased almost 2-fold to boost OAA and L-aspartate pools. The mutant simulation showed lower fluxes in the TCA cycle, possibly because of the carbon flux diverted from OAA to 3-HP formation. There is a minor carbon flux in the wild-type simulation to produce β-alanine (ADC reaction, 0.0005 $mmol.gDW^{-1}.h^{-1}$) because of the cell requirement of pantothenic acid (vitamin B5). The same reaction had its flux highly increased in the mutant simulations because of the upstream flux enhancement and the metabolic modifications, and this carbon flux followed the subsequent reactions in the pathway to produce 3-HP.

Regarding the ALAT reaction, which is written as pyruvate + L-glutamate ↔ L-alanine + α-ketoglutarate in the model, its flux changed from -0.5102 (WT) to -0.0159 $mmol.gDW^{-1}.h^{-1}$ (mutant). This result means that, for the mutant strain, less pyruvate and L-glutamate were consumed, and therefore, less L-alanine and α-ketoglutarate were formed.

The simulations also showed a slight increase in the R_GLUDy ($NADP^+$-dependent glutamate dehydrogenase) flux and decreased R_GLNS (glutamine synthetase) flux when the mutant was compared to the wild-type strain. These reactions are present in the NH_4 assimilation pathway and L-glutamate metabolism. The first produces L-glutamate from α-KG, and the second converts L-glutamate into glutamine. Together, these flux changes also contribute to L-glutamate pool increase and α-KG decrease, further boosting AAT equilibrium towards L-aspartate.

Another interesting result from comparing the wild-type and mutant simulations is related to L-valine:pyruvate aminotransferase (VPAT). According to Yoneyama *et al.* (2011), L-alanine is synthesized by three major aminotransferases in *E. coli*. Two of them are ALATs (encoded by the *yfdZ* and *yfbQ* genes) and, the third is VPAT (*avtA* gene), which converts, reversibly, L-valine and pyruvate into L-alanine and 3-methyl-2-oxobutanoate. This third reaction is named R_VPAMTr in the *i*ML1515 model and had its flux increased from 0.0 (in the WT simulation) to 0.3333 $mmol.gDW^{-1}.h^{-1}$ towards L-valine and pyruvate formation. It is believed that this reaction is diverted in this direction to supply L-valine cell requirements (which is compromised by VALTA knock-out) and produces, in parallel, a decrease of L-alanine pool and increase of pyruvate pool, further contributing to pushing BAPAT (R_APATr) towards MSA formation. Although this reaction was not present in the optimization results, it can be an interesting target for further evaluation.

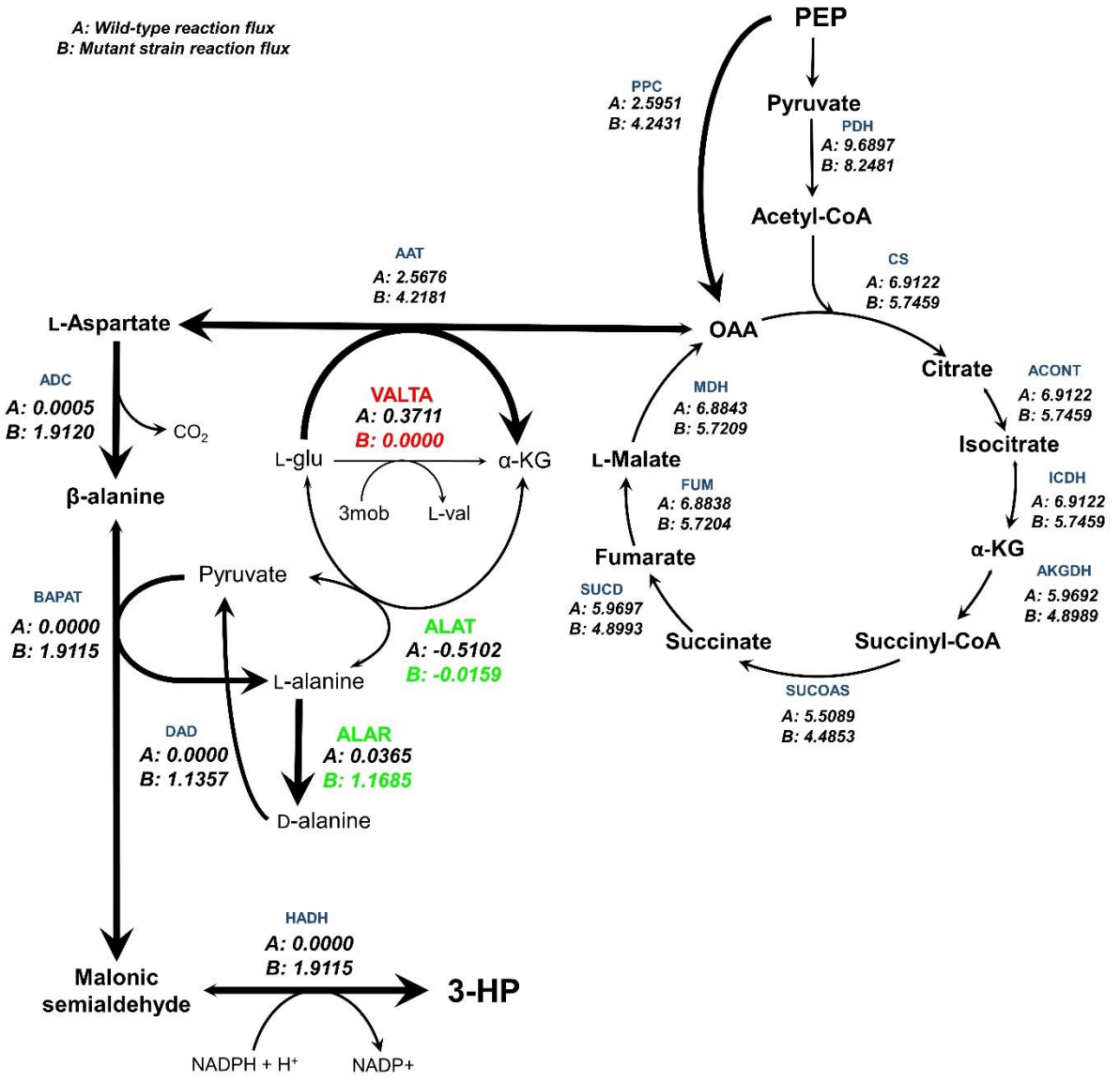


Figure 27 - The effect of the metabolic targets identified by the evolutionary algorithm over the carbon flux in the β-alanine pathway and TCA. PYK, pyruvate kinase; PDH, pyruvate dehydrogenase; CS, citrate synthase; ACONT, aconitase; ICDH, isocitrate dehydrogenase; AKGDH, α-ketoglutarate dehydrogenase; SUCOAS, succinyl-CoA synthetase; SUCD, succinate dehydrogenase; FUM, fumarase; MDH, L-malate dehydrogenase; PPC, phosphoenolpyruvate carboxylase; AAT, L-aspartate aminotransferase; ADC, L-aspartate 1-decarboxylase; BAPAT, β-alanine:pyruvate aminotransferase; HADH, 3-hydroxyacid dehydrogenase; ALAT, L-alanine aminotransferase; ALAR, alanine racemase; VALTA, L-valine transaminase and DAD, D-amino acid dehydrogenase. 3mob, 3-methyl-2-oxobutanoate. Flux values in mmol.gDW^{-1}.h^{-1}. Glucose as carbon source. Green: reaction overexpression; and

red: reaction deletion. Thicker arrows were used to address the increase of carbon flux in the β-alanine pathway.

As the product formation requires 1 mol of NADPH for each mol of 3-HP, a cofactor balance analysis was performed to investigate the effect of 3-HP production on the reactions for NADPH generation (Figure 28). The values for NADPH production fluxes for WT of 12.4905 and 12.3300 mmol.gDW^{-1}h^{-1} for glucose and xylose conditions, respectively, are in agreement with the values measured by Gonzalez, Long, and Antoniewicz (2017). In our simulations, NADPH was mainly produced from the isocitrate dehydrogenase reaction of the TCA cycle (~ 56 %) and the oxidative branch of the pentose phosphate pathway (~ 36 %). For glucose, NADPH production in the TCA cycle was slightly lower than for xylose (6.9130 against 7.1329 mmol.gDW^{-1}h^{-1}). The oxidative branch of pentose phosphate produced more NADPH in the glucose condition, 4.7107 mmol.gDW^{-1}h^{-1}, than in the xylose condition, 4.3413 mmol.gDW^{-1}h^{-1}. The main utilization of NADPH is for biomass formation, as it was consumed in many amino acid biosynthesis pathways (L-glutamate, L-valine, L-isoleucine, L-proline, L-lysine, besides the precursors' ornithine, chorismite, and L-homoserine) and, to a less extent, in some reactions of the cell wall biosynthesis. The L-glutamate biosynthesis accounted for 60 % of the NADPH utilization.

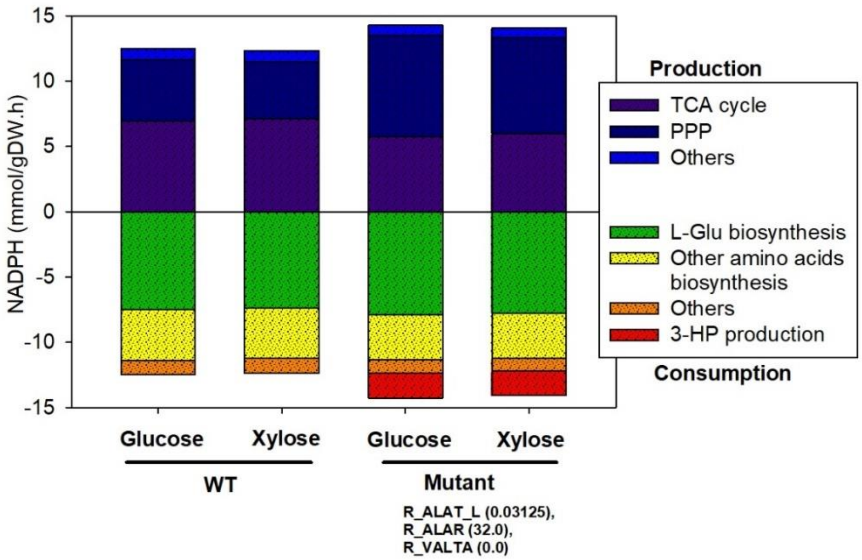


Figure 28 – NADPH balance of the simulations performed in *Optflux*. NADPH production was set as > 0 and < 0 for consumption.

In the mutant simulations, the NADPH production in the TCA cycle had a slight decrease, but a much higher flux was observed from the oxidative branch of PPP (~ 65 %). This increment on NADPH formation was required to meet its demand for 3-HP production, while the other NADPH requirements remained almost the same.

After the metabolic targets were chosen and evaluated *in silico*, the kinetic aspects for the selected reactions were analyzed from the MetaCyc and Brenda databases. For the ALAT reaction, also known as glutamate:pyruvate transaminase (GPT), the ALAT enzyme from *Saccharomyces cerevisiae* (encoded by its *ALT1* gene) was selected for cloning.

L-Alanine aminotransferase (E.C. 2.6.1.2) is a PLP-dependent transaminase that transfers the amino group from L-alanine to α-KG, producing pyruvate and L-glutamate. This reversible reaction has an equilibrium constant near one, so the reaction's preferent direction is determined by the *in vivo* substrate's concentration. *E. coli* has two functional ALAT encoded by the genes *yfbQ* and *yfdZ*. The ALAT encoded by the *yfbQ* gene is the most active (KIM, SCHNEIDER AND REITZER, 2010), but no information about this enzyme kinetics is available so far. However, Duff *et al.* (2012) reported that bacterial ALATs might be less specific than ALATs from eukaryotes. Duff *et al.* (2012) and McAllister *et al.* (2013) studied and reported the kinetic parameters of ALATs from many sources, including those from *S. cerevisiae*. This yeast has two ALAT encoding genes, but only the *ALT1* gene product is a functional enzyme (GARCÍA-CAMPUSANO *et al.*, 2009). The kinetic parameters of ScALAT seemed to favor pyruvate and L-glutamate formation (L-glutamate + PLP ↔ PMP + α-KG with K_m (α-KG) = 0.5 mM < K_m (L-glu) = 0.7 mM and PMP + pyruvate ↔ PLP + L-alanine with K_m (L-ala) = 0.3 mM < K_m (pyr) = 11.0 mM). So, *ALT1* overexpression was chosen for the first attempt of targeting the ALAT reaction in the direction that favors 3-HP formation in the β-alanine pathway.

There is only one report of ALAT overexpression for metabolic engineering purposes to the best of our knowledge. In their work, Borodina *et al.* (2015) tried to recycle L-alanine back into pyruvate by overexpressing the *ALT1* gene from *S. cerevisiae* to improve 3-HP yield in their engineered *S. cerevisiae* strain. However, the authors reported a 10 % decrease in the 3-HP production. Nevertheless, the strain they used did not overexpress the AAT reaction. Here, we propose to couple the ALAT and AAT overexpression, the latter demanding L-glutamate and producing α-KG that will be provided and consumed, respectively, by ALAT.

For the ALAR reaction (E.C. 5.1.1.1, PLP-dependent), the *alr* gene from *E. coli* was chosen for cloning. *E. coli* has two isoenzymes encoded by *dadX* and *alr* genes. The product of the *dadX* gene is known as the catabolic enzyme because its expression is induced during

growth on high concentrations of L- or D-alanine and is much more abundant than the constitutively expressed biosynthetic racemase encoded by the *alr*, involved in the synthesis of peptidoglycan in the cell wall (WILD *et al.*, 1985; JU *et al.*, 2011). By studying the *E. coli* D-alanine export system, Katsube *et al.* (2016) reported that the overexpression of *alr* enabled a higher extracellular level of D-alanine when compared to the overexpression of the *dadX* gene. Therefore, *alr* overexpression was selected to investigate the ALAR target suggested by the optimization results.

The third target is the VALTA reaction catalyzed by the branched-chain amino acid aminotransferase (BCAT, also known as transaminase B, E.C. 2.6.1.42) encoded by the *ilvE* gene. BCAT catalyzes the last reaction in the biosynthesis of L-valine, L-leucine, and L-isoleucine (branched-chain amino acids, BCAAs) and, to a less extent, L-phenylalanine (KURAMITSU *et al.*, 1985). Like other transaminases, it is also PLP-dependent and transfers the amino group from L-glutamate to a keto acid to form α-ketoglutarate and a BCAA, having pyridoxamine as intermediate. This enzyme has been targeted for metabolic engineering purposes in several works. For instance, Bückle-Vallant *et al.* (2014) deleted the *ilvE* gene in *C. glutamicum* for the overproduction of 2-ketoisocaproate, the L-leucine keto acid. Working with *S. cerevisiae*, Hammer and Avalos (2017) deleted the mitochondrial BCAT (encoded by *BAT1* gene) to increase isobutanol production (obtained from 2-ketoisovalerate). Also, for 2-ketoisovalerate production, Krause, Blombach, and Eikmanns (2010) deleted the *ilvE* gene in *C. glutamicum.* However, as this enzyme plays a key role in amino acid biosynthesis, one may argue that an Δ*ilvE* engineered strain would show auxotrophy or deficient growth in the absence of BCAAs supplementation. Nevertheless, this might not be a problem growing the cells in a complex medium. Another possible approach could be modifying the *ilvE* promotor sequence for gene down-regulation. This strategy was shown before by Holátko *et al.* (2009) that improved *ilvE* promoter for L-valine overproduction in *C. glutamicum* and by Hüser *et al.* (2005) that down-regulated it for pantothenate (synthesized from ketoisovalerate) overproduction. In our case, VALTA reaction underexpression or knock-out had the same simulation results in *Optflux*.

4.4 Improving 3-HP production based on *in silico* optimization results

To evaluate the targets from *in silico* optimization results from section 4.3, the *ALT1* gene from *S. cerevisiae*, encoding L-alanine aminotransferase, and the *alr* gene from *E. coli*, encoding one of its alanine racemases, were cloned in the pRSM6 plasmid, forming the pRA

and pRr plasmids (Figure 29 – A and B). These genes were then inserted in the pRak[2] plasmid, forming the pRAak and pRrak plasmids (Figure 29 – C), to assess the effect of ALAT and ALAR overexpression separately. Moreover, the plasmid pRArak was then constructed to simultaneously assess their overexpression (Figure 29 – D and E).

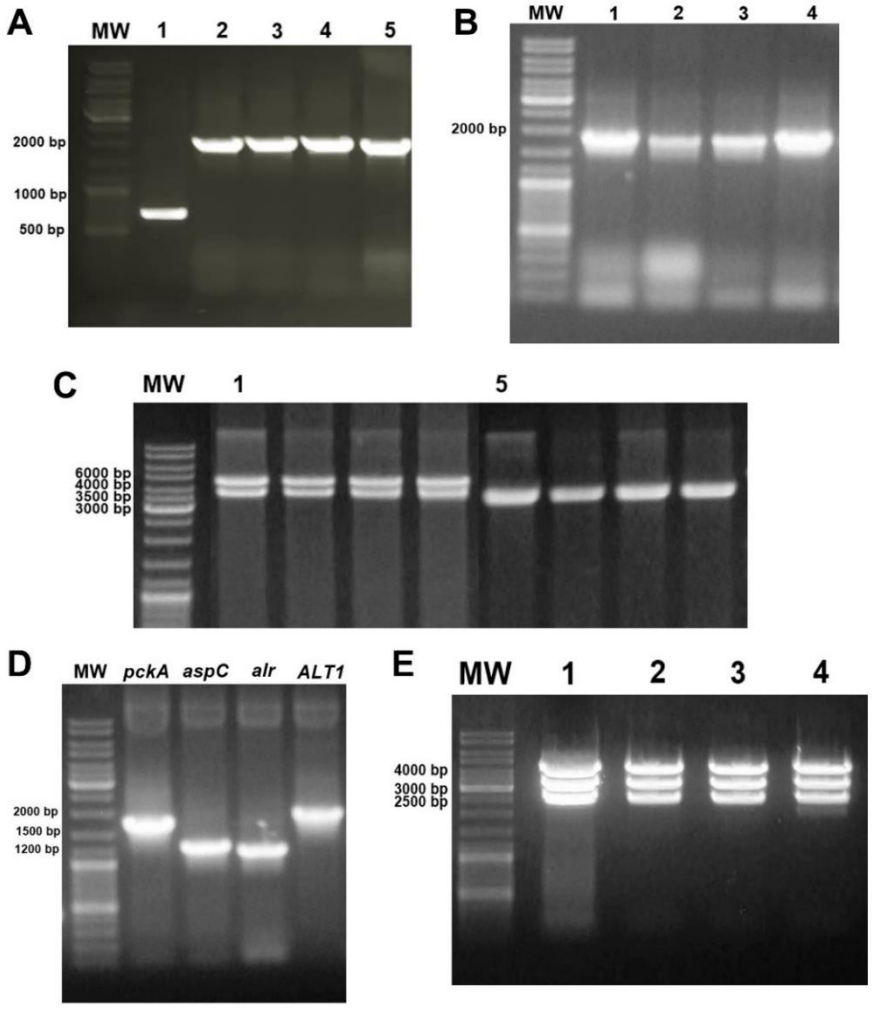


Figure 29 – Cloning of individual genes and pRAak, pRrak, and pRArak assembly confirmation by colony PCR and restriction digestion. (A) *alr* (1080 pb) by amplification of pRr MCS (lane 1: MCS as negative control; lanes 2-5: positive clones); (B) *ALT1* (1779 bp) amplified from the constructed plasmid pRA; (C) pRAak (lanes 1-4) and pRrak (lanes 5-8) restriction digestion with *Nhe*I – expected band sizes of 4793 bp and 4011 bp for pRAk, and 4225 bp and 3912 bp for pRrak. In the pRrak plasmid, the bands have similar sizes and are therefore difficult to differentiate. (D) Amplification of all genes in the pRArak plasmid. (E)

[2] These genes were inserted in the pRak plasmids because the cloning step came before fermentation trials of the previous strains, so we continued with PS160 even though the *pckA* and *aspC* overexpression did not improve the 3-HP titer (see page 53).

Restriction digestion of pRArak with *BamH*I – expected band sizes of 4396 bp, 3219 bp, and 2501 bp. All the bands have the expected size. MW: molecular weight marker Gene Ruler DNA ladder mix (Thermo Scientific).

The other metabolic target, *ilvE* gene deletion, was performed using the No-SCAR method. The ssDNA for Δ*ilvE* was designed to delete 890 bp from its 930 bp sequence (Figure 30). The resulting *E. coli* K-12 MG1655 Δ*ilvE* was named PS003. The effect of modifying ALAT, ALAR, and BCAT expression levels on 3-HP production was assessed using the last strain from the previous section, PS160, although this strain only produces small amounts of 3-HP (0.079 ± 0.004 g/L).

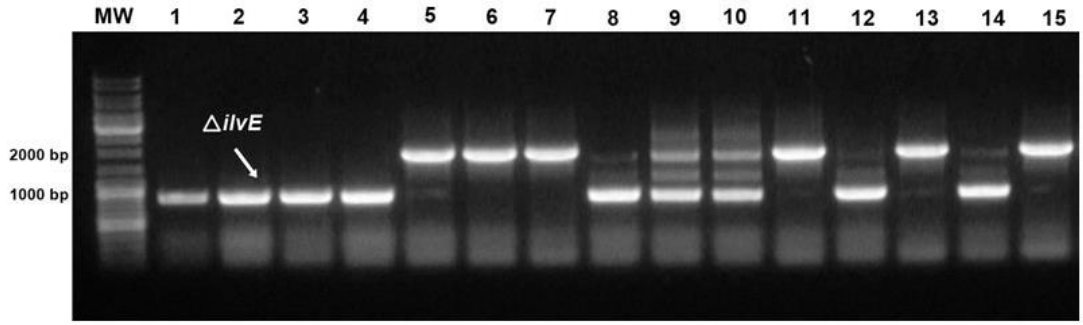


Figure 30 – Knock-out of *ilvE* gene in *E. coli* K-12 MG1655 genome by using No-SCAR technique. Amplification of *ilvE* gene by colony PCR. Negative control (lane 15), mutant strains (lanes 1-4, 8-10, 12 and 14), and wild-type strains (lanes 5-7, 11 and 13). The primers used for screening were designed to amplify a 1735 bp region in the wild-type strain and an 845 bp region in the *ilvE* mutant strains. MW: molecular weight marker Gene Ruler DNA ladder mix (Thermo Scientific).

The plasmids pRAak and pRrak were transformed into PS100 cells, thus creating PS170 and PS180, respectively. The plasmids pEbtyGpD and pRak were both transformed in PS003 cells to construct the PS163 strain. These three strains were constructed to assess the effect of each target predicted by the evolutionary optimizations separately. We also evaluated these targets in pairs, transforming the pRAak and pRrak plasmids in the PS003 strain (*E. coli* K-12 MG1655 Δ*ilvE*) already harboring pEbtyGpD, and the pRArak plasmid in the PS100 strain, thus creating the PS173, PS183, and PS190 strains. Finally, to assess the effect of all targets predicted in the optimizations, we engineered the PS193 strain by transforming PS003 with pEbtyGpD and pRArak plasmid.

The seven strains described before (PS163, 180, 170, 183, 173, 190, and 193) were grown on glucose for 24 h after induction, and the final 3-HP titer was measured (Figure 31) and compared to the titer of PS160 to evaluate if these targets could increase this acid production. The *ALT1* overexpression decreased the 3-HP final titer of PS170 compared to the titer of PS160 by almost 10 %, as reported by Borodina *et al.* (2015). The *alr* overexpression did not affect 3-HP final production, whereas BCAT knock-out improved 3-HP production t by 34 % (0.107 ± 0.010 g/L). However, together with *ALT1* expression, this reaction deletion had no effect over 3-HP production (PS173 strain). Differently, the *ilvE* knock-out with *alr* overexpression increased 3-HP final titer to 0.094 ± 0.004 g/L (PS183 strain). The overexpression of *ALT1* and *alr* genes also decreased 3-HP final titer to 0.061 ± 0.009 g/L. However, the combination of all targets predicted by the SPEA2 algorithm further increased the 3-HP final titer to 0.111 ± 0.001 g/L.

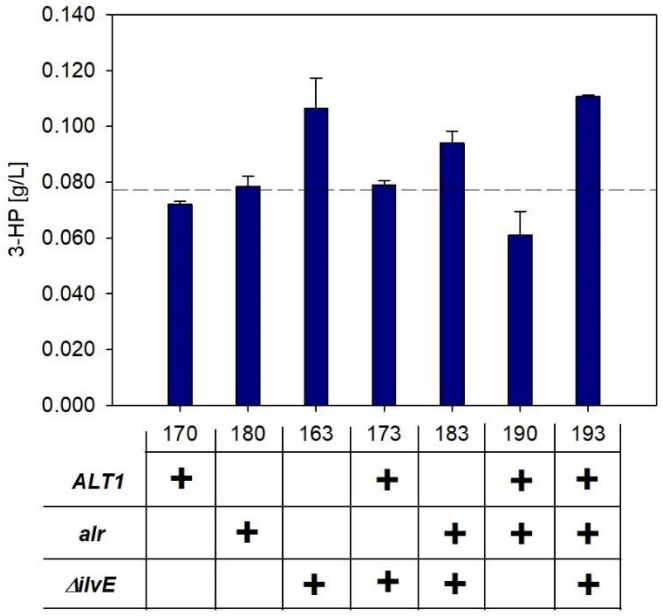


Figure 31 – 3-HP final titer of the strains harboring the targets predicted by the evolutionary optimizations: ALAT (encoded by *ALT1* gene from *S. cerevisiae*) and ALAR (encoded by *alr* gene from *E. coli* K-12 MG1655) overexpression with BCAT (encoded by the *ilvE* gene in *E. coli*) knock-out. The dashed horizontal line represents the 3-HP titer from PS160. Medium values are shown and error bars represent standard deviations.

The results presented above indicate the excellent quality of the non-intuitive metabolic targets predicted *in silico*, as their direct influence on 3-HP formation could be observed even

using a low producer strain at this point. The best result came from the combination of the three targets as indicated by the optimization results, improving the 3-HP final titer by 39 % compared to the PS160 strain. However, as this is a low titer and much less than what we first obtained with PS100 strain, we decided to test these non-intuitive targets without *pckA* and *aspC* overexpression.

Therefore, to evaluate the effect of each target again, the pRA and pRr plasmids were transformed into the PS100 cells, creating the PSO101 and PSO102 strains. The PS003 strain was transformed with pEbtyGpD to create the PSO103 strain. To test the targets in pairs and all of them together, the *ALT1* and *alr* genes were clustered in the pRAr plasmid in the pseudo-operon configuration (Figure 32).

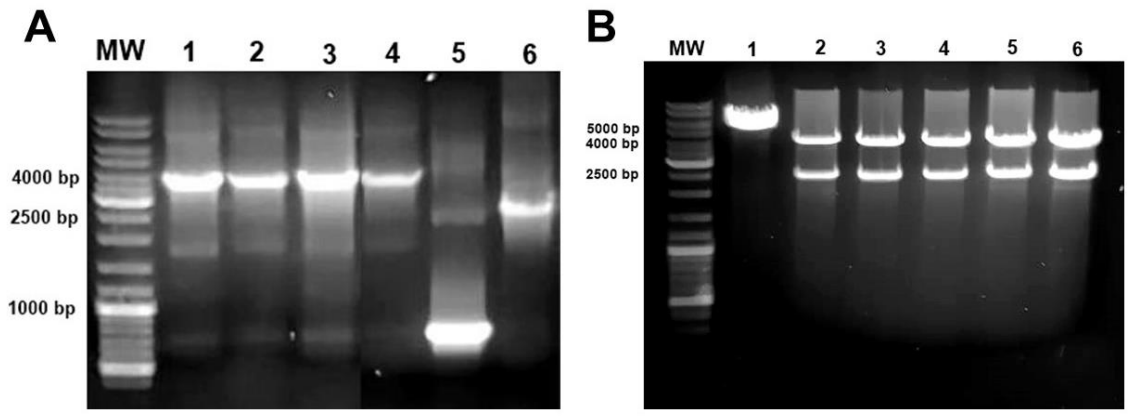


Figure 32 - Cloning confirmation of: (A) pRAr assembly by PCR colony with amplification of the MCS region of the pRAr plasmid (3750 bp, lanes 1-4), pRSM6 (710 bp, lane 5) and pRA (2438 bp, lane 6), and (B) pRAr assembly by digestion of the extracted plasmids with *BamH*I (lanes 2-6) since this enzyme have cut sites inside the *ALT1* and *alr* genes, separated by 2501 bp (lower band). The other band of 4413 bp represents the rest of the plasmid. The negative control (lane 1) was pRA digested with the same enzyme. All bands have the expected size. MW: molecular weight marker Gene Ruler DNA ladder mix (Thermo Scientific).

The pRAr plasmid was transformed into PSO103 and PS100 strains to construct the PSO107 and PSO104, respectively. PSO105 and PSO106 were constructed transforming PSO103 cells with pRA and pRr, respectively. These seven new PSOs strains were cultivated in glucose and had the final titer of 3-HP measured 24 and 30 h after induction (Figure 33). The

30 h point was taken because we noticed that glucose was not depleted in the medium after 24 h for the prior strains of this section.

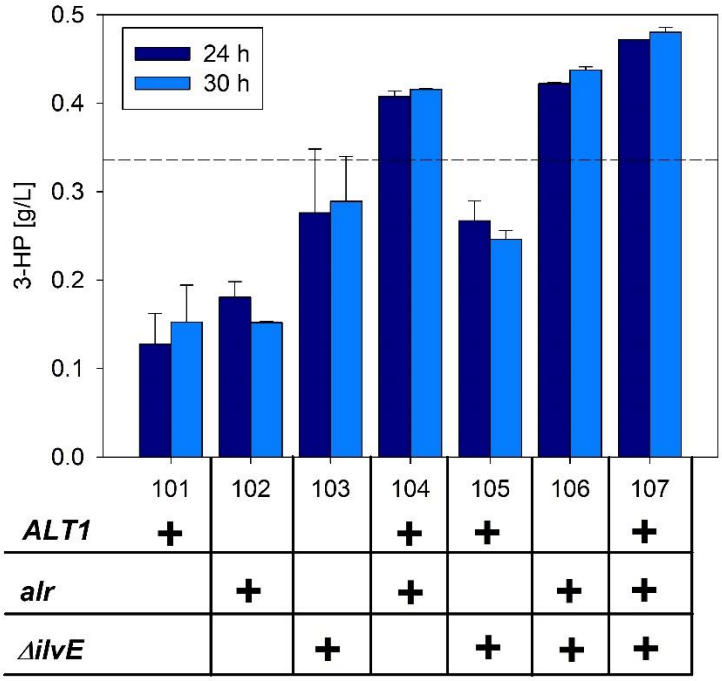


Figure 33 – 3-HP final titer of the PSOs strains, harboring *ALT1* and *alr* overexpression and *ilvE* knock-out individually (101, 102, and 103), in pairs (104, 105, and 106) and altogether (107). Medium values are shown and error bars represent standard deviations. The dashed horizontal line represents the 3-HP titer from PS100.

Overall, there was no significant difference in the 3-HP final concentration for the samples taken after 24 h and 30 h of induction. Similar to what happened for PS170, the 3-HP titer decreased in the strain harboring only the pRA plasmid for *ALT1* overexpression. For PSO102, the overexpression of just *alr* also decreased the product's final titer compared to the PS100 production. The *ilvE* knock-out alone and together with *ALT1* did not have any effect in the 3-HP final titer from PS100. However, in the case of PSO104 and PSO105, which produced 0.408 ± 0.001 and 0.422 ± 0.001 g/L, there was a 18 % and 24 %, respectively, increase in the acid formation. Again, the best-engineered strain PSO107 confirmed that the predicted targets could indeed improve 3-HP production, reaching 0.472 ± 0.0 g/L of 3-HP after 24 h and 0.480 ± 0.006 after 30 h, which is 42 % higher than the PS100 first production.

The PSO107 strain was then cultivated again in parallel with PS100 to compare their final 3-HP titers. The growth profile for PSO107 is depicted in Figure 34. In this new trial, the

PS100 strain produced 0.366 ± 0.007 g/L, whereas PSO107 produced twice this titer, 0.743 ± 0.016 g/L of 3-HP. This result further confirms that PSO107 is an optimized strain when compared to PS100.

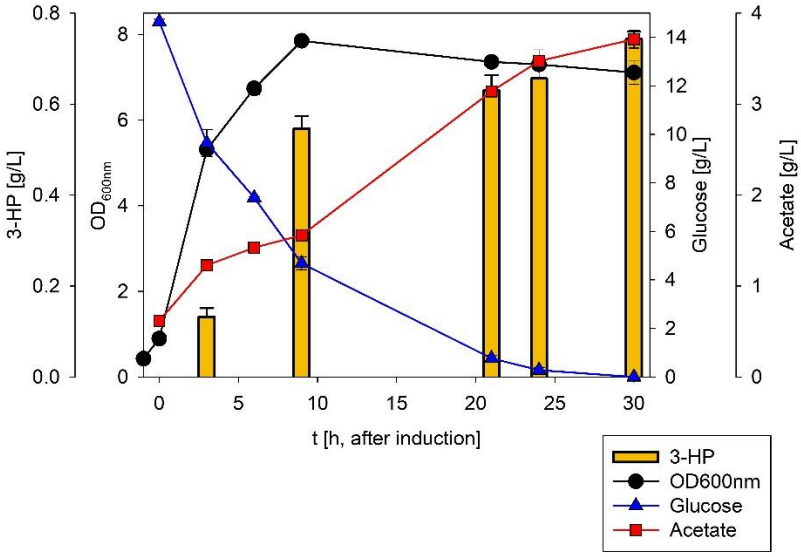


Figure 34 - Growth profile of PSO107 (*E. coli* K-12 MG1655 Δ*ilvE* harboring pRAr plasmid) in MR medium containing glucose as substrate. Medium values are shown and error bars represent standard deviations.

This last strain, PSO107 (Figure 35), harbors all the targets predicted in the evolutionary algorithm proving the usefulness of employing genome-scale metabolic models and evolutionary algorithms tools for metabolic engineering purposes.

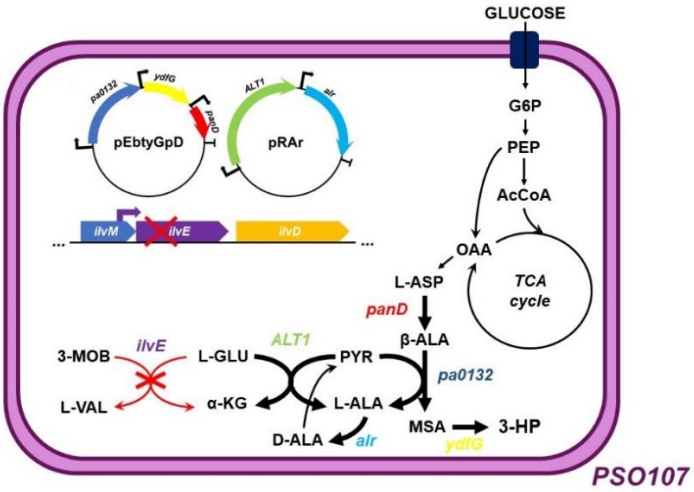


Figure 35 – PSO107 schematic representation with the respective genetic modifications (Δ*ilvE*, pEbtyGpD and pRAr) and their effect on central metabolism. Thicker arrows represent reaction

overexpression. Red arrows with a red X stand for reaction deletion. G6P: glucose-6-phosphate, PEP: phosphoenolpyruvate, AcCoA: acetyl-CoA, OAA: oxaloacetate, L-ASP: L-aspartate, β-ALA: β-alanine, MSA: malonic semialdehyde, 3-HP: 3-hydroxypropionic acid, PYR: pyruvate, L-ALA: L-alanine, D-ALA: D-alanine, L-GLU: L-glutamate, α-KG: α-ketoglutarate, L-VAL: L-valine and 3-MOB: 3-methyl-2-oxobuatnoate.

4.5 Engineering a co-sugar fermenting strain

Lignocellulosic biomass is mainly composed of glucose, arabinose, and xylose. Xylose and arabinose comprise almost 95 % of the hemicellulosic fraction (GROFF *et al.*, 2012). Therefore, when developing an *E. coli* strain capable of growing and fermenting sugars from this kind of substrate, it is crucial to take into account the need for co-utilization of sugars, since *E. coli* cells consume them in a hierarchical manner, known as diauxic growth or carbon catabolite repression (CCR), with glucose as the primary carbon source. The genes responsible for metabolism of secondary carbon sources remain repressed in the presence of glucose. Globally, this phenomenon is mediated by the phosphorylation level of $EIIA^{Glc}$, a domain of the glucose transporter in the phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS). A key factor that affects this phosphorylation level of $EIIA^{Glc}$ is the PEP:pyruvate ratio. When extracellular glucose level is low, PEP and phosphorylated-$EIIA^{Glc}$ levels increase. $P{\sim}EIIA^{Glc}$ activates cAMP synthesis in adenylate cyclase (AC), leading to a higher cAMP-CRP complex level. This complex activates the transcription of multiple operons for secondary sugar utilization, such as *ara* operon (for arabinose) and *xyl* operon (for xylose). On the other hand, when the extracellular glucose level is high, unphosphorylated $EIIA^{Glc}$ acts inhibiting/inactivating enzymes and transporters for secondary carbon sources, thus preventing the inducer formation. This mechanism is known as inducer exclusion (Figure 36-A; GÖRKE AND STÜLKE, 2008).

In this work, the project submitted to FAPESP proposed the utilization of xylose as a carbon source, based on future prospects for the engineered strain to be able of metabolizing hemicellulosic hydrolysates in the context of the ethanol biorefinery (Thematic Project Process 2016/10636-8). However, hemicellulosic hydrolysates are not just made of xylose, as other sugars are also present in its composition. Since *E. coli* cells consume these sugars in a hierarchical manner due to CCR, a strategy to alleviate CCR in the engineered strains for 3-HP production of this project were investigated.

Xylose metabolism in *E. coli* is mediated by *xylAB*, *xylFGHR*, and *xylE* genes (Figure 36-B). Firstly, xylose is transported into the cells by XylFGH, a high-affinity ABC type transporter (ATP-binding cassette), or by XylE, a low-affinity H+ symporter. Then, xylose is converted into xylulose by xylose isomerase (encoded by *xylA* gene). Xylulose is converted into xylulose-5-phosphate by xylulokinase (encoded by *xylB* gene), and xylulose-5-phosphate enters the central metabolism via pentose phosphate pathway (Figure 36-A). The P_E, P_A, and P_F promoters of the *xylE* gene and the *xylAB* and *xylFGH* operons, respectively, are under the positive control of XylR (encoded by *xylR* gene). Thus, XylR is a transcriptional activator of xylose metabolism in the presence of this sugar (NI *et al.*, 2013; SONG AND PARK, 1997). Xylose-bounded XylR activates transcription of the *xyl* operon by protein interaction with the σ70 subunit of RNA polymerase due to the proximity of their sites (XylR and -35 region of the promoter) (NI *et al.*, 2013; SONG AND PARK, 1997).

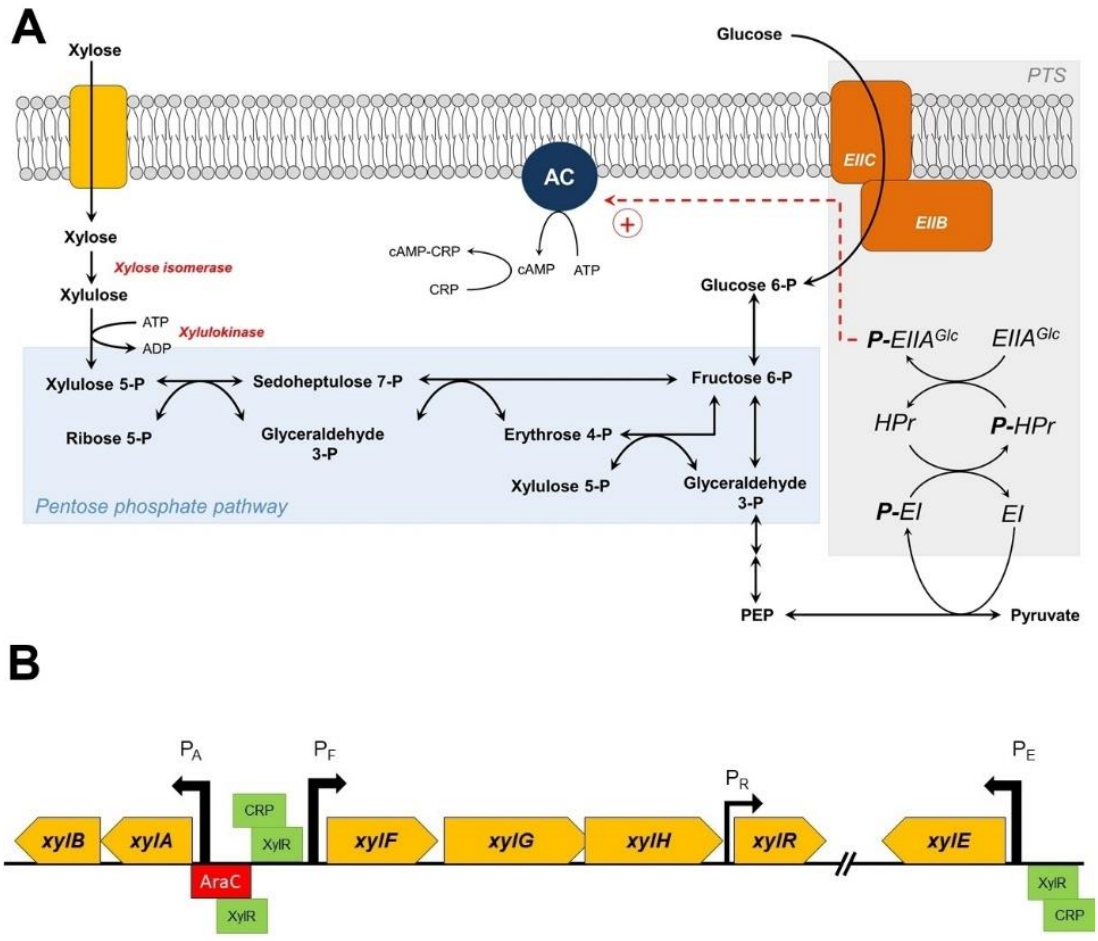


Figure 36 – (A) Carbon catabolite repression and xylose assimilation pathway. (B) *xyl* operon in *E. coli*. AC, adenylate cyclase; PTS, phosphoenolpyruvate:carbohydrate phosphotransferase system; PEP, phosphoenolpyruvate; CRP, cAMP receptor protein.

Sievert *et al.* (2017) identified two point-mutations in XylR, R121C and P363S that, whether alone or together, enable *xyl* genes up-regulation and can even relieve CCR in sugar mixtures of glucose-xylose, xylose-arabinose, and glucose-xylose-arabinose (MARTINEZ *et al.*, 2019). These point mutations enhance the xylose utilization rate and are believed to increase XylR affinity for its operator motifs. The authors showed that XylR* can even activate *xyl* operons transcription without xylose binding (SIEVERT *et al.*, 2017).

Based on that, XylR overexpression is believed to be a good strategy to alleviate CCR and enables *E. coli* cells to consume pentoses and hexoses present in the hemicellulosic hydrolysate at the same time. Therefore, the *xylR* gene from *E. coli* was cloned in the pRSM6 plasmid, generating the pRx (Figure 37 – A and B). The *xylR* gene in this plasmid had its Pro363 amino acid residue changed for Ser, generating the pRx* plasmid (Figure 37 - C), to evaluate the performance of cells expressing the P636S mutant gene for co-consuming glucose and xylose.

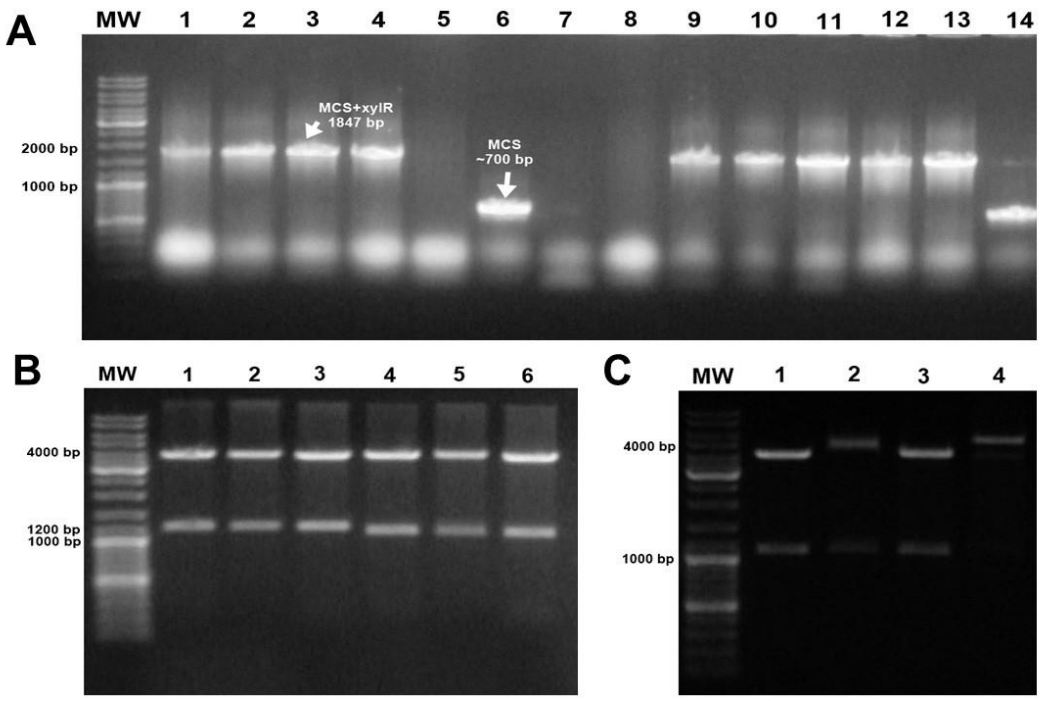


Figure 37 – Cloning confirmation of: (A) *xylR* in pRSM6 by colony PCR with amplification of the MCS region of the plasmid plus the cloned *xylR* gene, (B) pRx assembly by digestion of the extracted plasmids with *Nde*I (lanes 1-3) and *Sal*I (lanes 4-6) since both of these enzymes have cut sites inside the plasmid and inside the gene sequence and (C) *xylR** by digestion of the extracted pRx* plasmids (lanes 1-3) and pRx (lane 4, control) with *Pst*I and *Bgl*II: this double digestion produces a second band of 1103 bp for the mutant plasmid but none for the control,

that does not have a cut site for *Pst*I. All bands have the expected size. MW: molecular weight marker Gene Ruler DNA ladder mix (Thermo Scientific).

The plasmids pRx and pRx* were individually transformed into PS100 strain, thus creating PSX100 and PSXM100 strains, respectively. The PS100 strain was also transformed with an empty pRSM6 vector to assess the metabolic burden caused by harboring this high copy plasmid.

The strains were then grown on glucose, xylose, and a mixture of glucose and xylose (1:1, on a C-mol basis) to evaluate 3-HP production for each carbon source and their capacity to co-consume both sugars in the mixture. The OD_{600nm} growth profile of these strains is shown in Figure 38. The OD_{600nm} reached higher values when the strains were grown on glucose than xylose. Based on PSR100 growth, the metabolic burden of carrying an empty plasmid of high copy number had little effect on cell growth. However, for the PSXM100 strain that carries the mutant P363S *xylR* gene, a metabolic burden effect is evident, possibly because this XylR* transcriptional regulator increases the transcription of *xyl* operon, diverting cell resources and energy. This effect is even more pronounced when cells are grown on glucose.

When cells were grown on a mixture of glucose and xylose, the CCR was present for PS100 and PSR100 (with a new lag phase between 9 and 12 h of growth), whereas no CCR was observed for PSX100 and PSXM100, as expected. Interestingly, only the *xylR* overexpression was enough to alleviate the sugar preference uptake. In the study of Sievert *et al.* (2017), the P363S point mutation was performed in the genomic copy of the *xylR* gene. In both cases, *xylR* overexpression and P363S point mutation is enough to enable *E. coli* K-12 MG1655 cells to co-consume glucose and xylose simultaneously.

Interestingly,PS100 and PSR100 strains continued to grow after the new lag phase (after 12 h) until OD_{600nm} reached almost 12 in the sugar mixture, whereas PSX100 and PSXM100 did not, although the initial substrate concentration was the same.

Figure 39 presents the 3-HP (A) and acetate (B) production of these fermentation trials. With glucose as a carbon source, PSX100 and PSXM100 produced more 3-HP and less acetate than PS100 and PSR100. For these strains, glucose may have been transported into the cells by the xylose low-affinity proton symporters (XylE transporters) instead of being transported exclusively by the PTS system. This alternative glucose transport could alleviate the PEP consumption for glucose uptake, favoring the PEP channeling to OAA formation. The reduction of PEP conversion into pyruvate by the PTS system also results in less acetyl-CoA formation from pyruvate, reducing acetate accumulation. XylE transporters from *E. coli* are

known to be promiscuous (DESAI AND RAO, 2010; JECKELMANN AND ERNI, 2020) and Andreeva, Golubeva, and Katashkina (2013) reported that the constitutive expression of XylE from *Pantoea ananatis* could restore growth on glucose of *E. coli* MG1655 cells that had the PTS inactivated.

3-HP production was higher on glucose than xylose for all strains, since the fermentation with xylose as the sole carbon source produced 2-fold more acetate for PS100 and PSR100 and almost 10-fold more acetate for PSX100 and PSXM100 when compared to the levels observed when glucose was used. We hypothesize that this higher acetate production for PSX100 and PSXM100 may be due to an increase of ATP requirements for phosphorylation of xylulose to xylulose-5-phosphate and xylose transport through XylFGH in these strains. The extra ATP needed could be derived from acetate formation by acetate kinase and phosphate acetyltransferase.

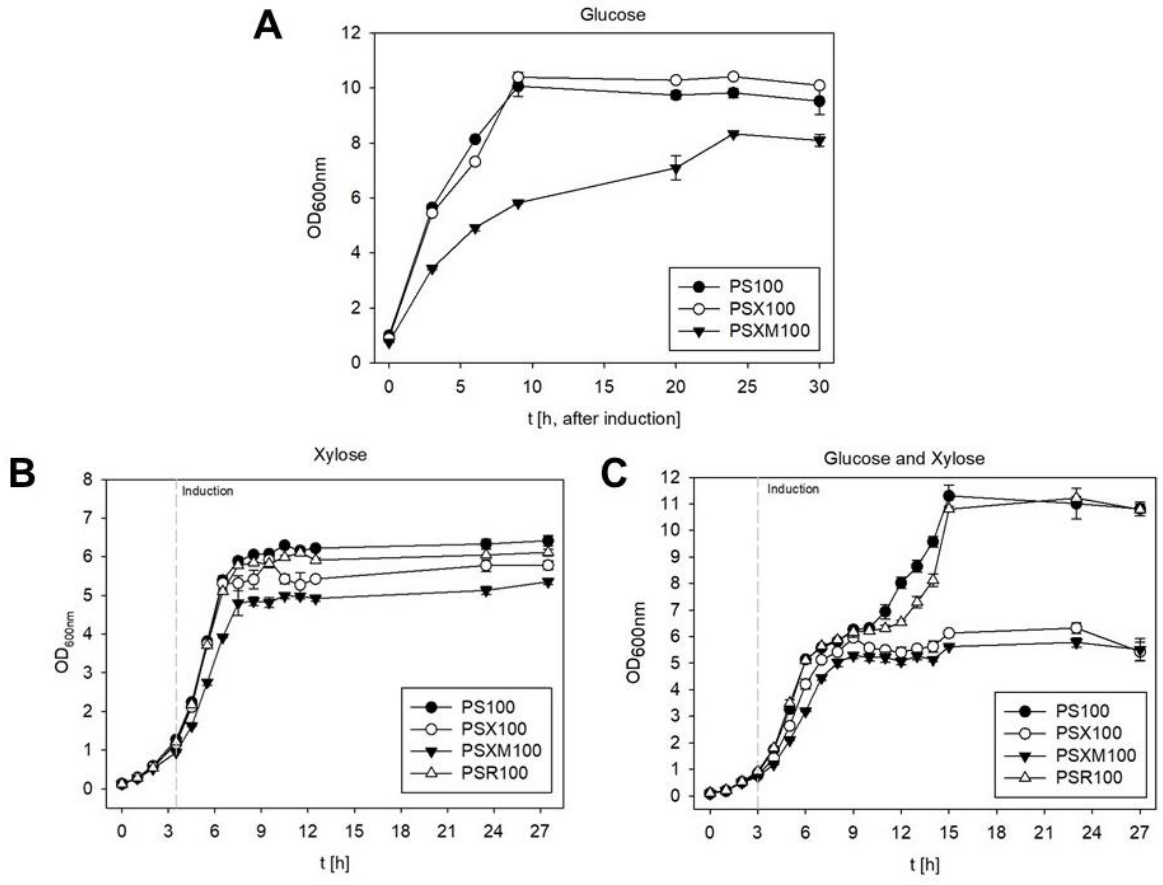


Figure 38 - Growth profile of PS100, PSR100, PSX100, and PSXM100 on (A) glucose, (B) xylose and (C) glucose : xylose (1:1, in C-mol basis). Medium values are shown and error bars represent standard deviations.

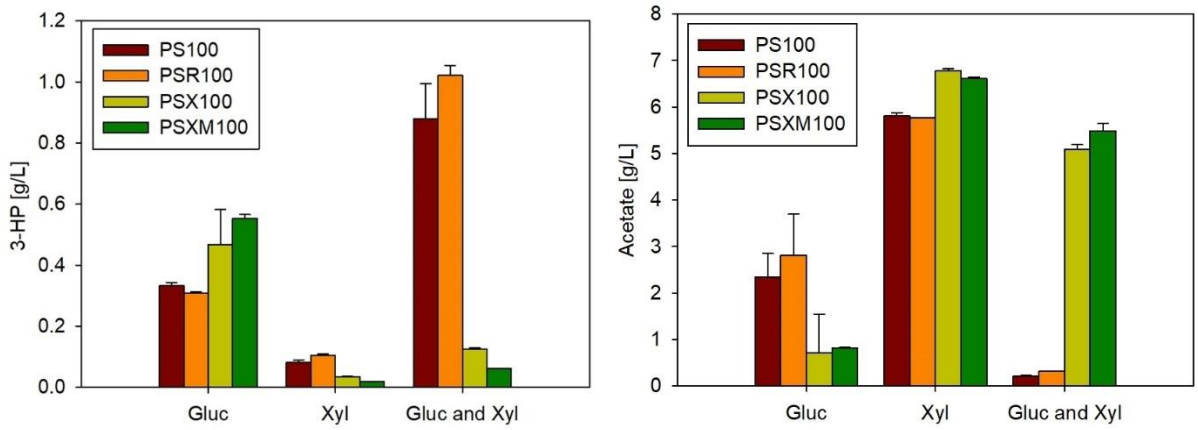


Figure 39 – 3-HP and acetate production by PS100, PSR100, PSX100, and PSXM100 in glucose, xylose, and in a mixture of glucose : xylose (1:1, in C-mol basis). Medium values are shown and error bars represent standard deviations.

In the glucose : xylose mixture, the PS100 and PSR100 strains produced more 3-HP and less acetate than both PSX100 and PSXM100, reaching 0.878 ± 0.110 g/L and 1.021 ± 0.033 g/L for PS100 and PSR100, respectively, after 24 h of induction. The PSX100 and PSXM100 strains produced almost 25-fold more acetate and 7-fold less 3-HP than PS100. The reduced acetate production by PS100 seems to have an important role in this improved 3-HP titer in the glucose : xylose condition.

Figure 40 depicts the results for cellular growth, sugar consumption, and acetate production for each strain in the sugar mixture condition. For PS100 and PSR100, xylose is consumed after glucose depletion due to CCR, as expected. In the case of PSX100 and PSXM100, both sugars are co-consumed during the entire fermentation due to the XylR overproduction and, therefore, a higher expression of *xyl* operon. This result confirms that the CCR can be bypassed by *xylR* overexpression, as described before . Moreover, in the case of PSXM100, xylose is consumed faster than glucose, indicating that the P363S point-mutation indeed leads to a higher activity of XylR* in regulating *xyl* operon expression.

Acetate build-up increases during exponential and stationary growth phases of PSX100 and PSXM100 strains. However, this behavior is not observed for PS100 and PSR100, where acetate is consumed along with xylose after glucose exhaustion. Under glucose growth conditions, and therefore catabolite repression, acetate is accumulated due to the overflow metabolism, as discussed before, caused by a low activity of TCA cycle enzymes and an imbalance in the acetate node (MATSUOKA AND SHIMIZU, 2013). This metabolite was continuously produced till it reached 5.093 ± 0.099 g/L and 5.489 ± 0.156 g/L for PSX100 and

PSXM100, respectively. In this case, a great carbon amount is deviated from central metabolism towards the acetate production pathway, hindering the substrate conversion into the product of interest. Also, this high titer of acetate in the culture medium has harsh effects on cells and may cause growth inhibition.

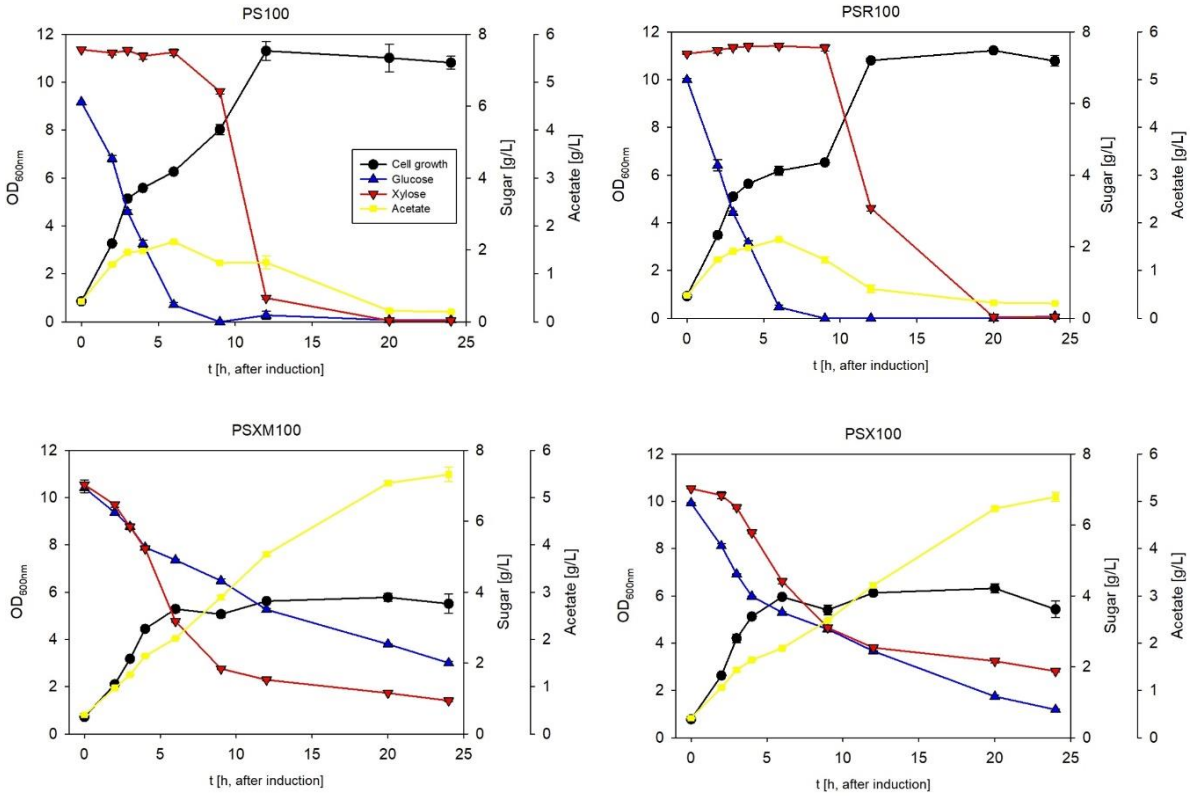


Figure 40 – Growth profile of PS100, PSR100, PSX100, and PSXM100 in a glucose : xylose mixture. Glucose consumption (red line), xylose consumption (blue line), cell growth (black line), and acetate production (yellow line). Medium values are shown and error bars represent standard deviations.

When glucose level diminishes, this catabolite repression, which is mainly mediated by CAP (cAMP-CRP complex) levels, is alleviated. The CAP complex regulates the IclR (isocitrate lyase regulator) expression, besides many other genes and operons in the *E. coli* genome. This DNA-transcriptional binding regulator represses the *aceBAK* operon expression. Under catabolite repression and low levels of CAP, IclR levels are high and *aceBAK* operon is repressed. When cell growth shifts to the xylose consumption phase in PS100 and PSR100, IclR levels decrease, and *aceBAK* operon is activated to enable acetate consumption, as observed after 6 h of growth for these two strains. The *aceBAK* operon is comprised of isocitrate lyase (encoded by *aceA* gene), malate synthase (encoded by *aceB* gene), and isocitrate kinase

(encoded by *aceK* gene). Isocitrate lyase and malate synthase catalyze the glyoxylate bypass reactions (Figure 41), while isocitrate kinase regulates isocitrate flux to α-ketoglutarate by phosphorylating isocitrate dehydrogenase (CORTAY *et al.*, 1991). Glyoxylate bypass activation diverts carbon flux from the TCA cycle to glyoxylate, preventing carbon loss as CO_2, thus increasing biomass yield. It can also increase the OAA pool from glyoxylate, besides reducing acetate formation (WAEGEMAN *et al.*, 2011). Because of this flux redistribution, α-ketoglutarate formation in the TCA cycle is diminished and may be replenished by other reactions like OAA and L-glutamate transamination into L-aspartate and α-KG by aspartate transaminase (*aspC* gene). In PSX100 and PSXM100 cultivations, glucose was not depleted after 24 h of induction, and acetate was continuously produced by overflow metabolism and catabolite repression.

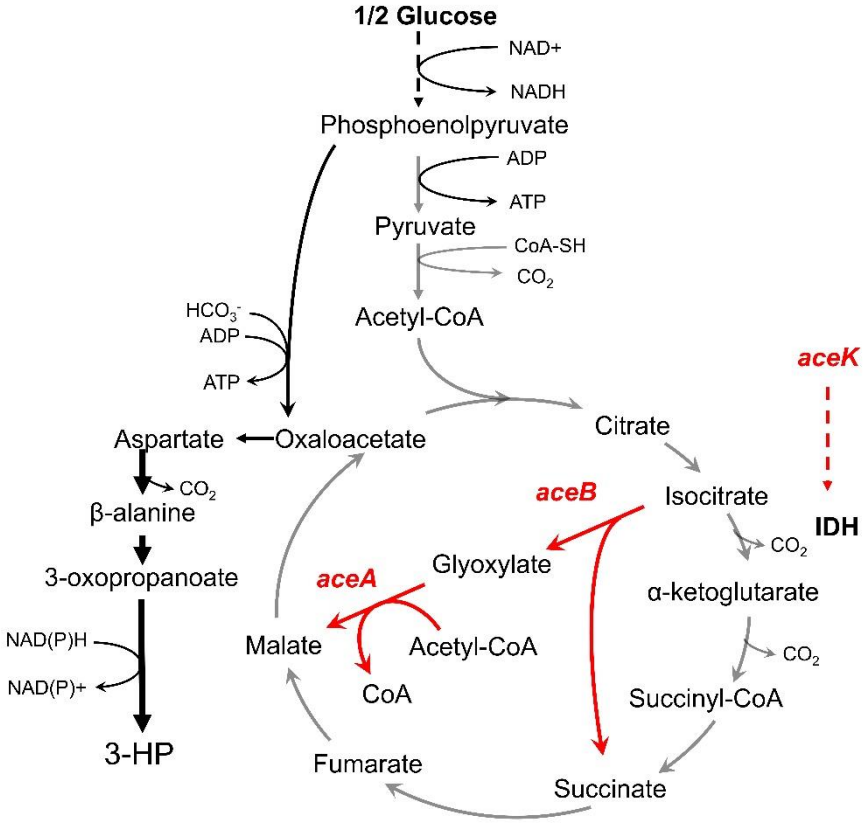


Figure 41 - The glyoxylate shunt in TCA cycle encoded by the *aceBAK* operon: *aceA* - isocitrate lyase, *aceB* - malate synthase and *aceK* - isocitrate kinase/phosphorylase, which regulates isocitrate dehydrogenase (IDH) activity. Red arrows represent the glyoxylate shunt and the regulation of IDH by *aceK*.

The best-engineered strain for 3-HP production, obtained in section 4.4, was also tested in this mixture, and its growth profile is shown in Figure 42. The final 3-HP production of PSO107 in glucose : xylose medium was 1.147 ± 0.015 g/L, slightly higher (10 %) than the production of PS100 cultivated at the same time of PSO107 (1.040 ± 0.050 g/L), for comparison purposes. The growth profile, substrate consumption, and acetate production are similar to that displayed by PS100 in the same conditions.

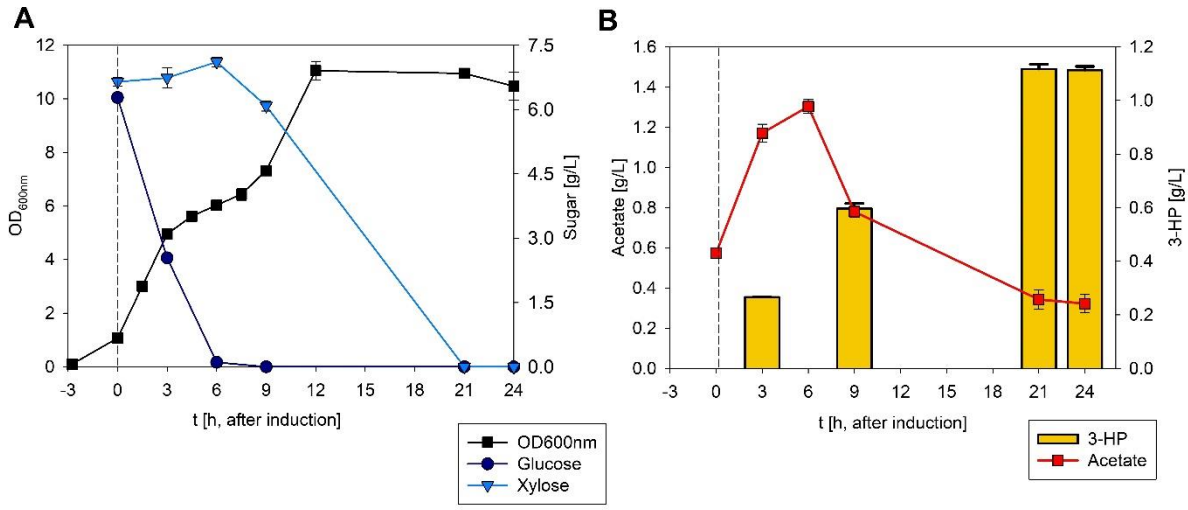


Figure 42 - Growth profile of PSO107 (*E. coli* K-12 MG1655 Δ*ilvE* harboring pRAr plasmid) in a glucose : xylose mixture. Dashed line represents the induction moment (t = 0 h). Medium values are shown and error bars represent standard deviations.

Finally, this work obtained an optimized strain for 3-HP production and reported a novel cultivation condition that improved the 3-HP final titer. Although many works have searched ways of disrupting the CCR mechanism to allow the co-consumption of glucose and xylose, this work shows that the diauxic growth displayed by our engineered strain is better for the 3-HP production than the condition with co-consumption, performed by PSX100 and PSXM100.

There is a general assumption that engineering strains to co-consume various sugars (as those present in lignocellulosic hydrolysates) would favor fermentative processes in terms of their productivity. In other words, it is believed that by having a hierarchical order of sugar consumption, the process productivity is diminished by the increase in the fermentation time required for both sugars depletion. However, in this work, we observed that this assumption does not hold under the conditions evaluated here. Surprisingly, the strains that were not engineered for CCR relief produced higher amounts of 3-HP from a mixture of glucose and xylose, indicating that the observed diauxic growth was not detrimental to the process

productivity. Actually, the co-consuming strains took a longer time to consume the carbon sources than the PS100/PSR100/PSO107 strains. Moreover, for PS100 and PSO107 strains, the time for total sugar consumption was similar (~20 h) both for glucose or glucose : xylose mixture, but the final 3-HP titers were higher when the sugar mixture was used. These results led to higher volumetric productivities for both strains grown on the sugar mixture.

The productivity and overall yields of the strains tested for 3-HP production in glucose : xylose is presented in Table 10. The biomass yield of PS100 and PSO107 was higher than observed for PSX100 and PSXM100 because PS100 and PSO107 continued growing after glucose depletion, whereas PSX100 and PSXM100 growth ceased after 9 h of induction. The 3-HP productivity of PS100 and PSO107 was also higher than for PSX100 and PSXM100, reflecting their final titers. Furthermore, the productivity of PS100 was almost 176 % higher in the glucose : xylose condition than for the glucose condition. The same behavior was observed for PSO107, with a productivity increment of 54 % in the glucose : xylose mixture.

Table 10 – Global yields and productivity of PS100, PSX100, PSXM100, and PSO107 strains in glucose : xylose medium, after 24 h of induction with IPTG 1 mM.

Strain	**$Y_{X/S}$ [(g/L)$^{-1}$]**	**$Y_{P/S}$ [g 3-HP/g sugar]**	**Productivity**
PS100	0.72	0.06	0.036 g/L-h
PSX100	0.36	0.008	0.005 g/L-h
PSXM100	0.37	0.004	0.003 g/L-h
PSO107	0.72	0.08	0.048 g/L-h

These results show that the best condition for 3-HP production found in this work was reached using a mixture of glucose and xylose as carbon sources, consumed hierarchically, which favored the final titer, yield, and volumetric productivity.

5 CONCLUSIONS

This work's primary objective was to design and construct an *E. coli* strain capable of producing 3-HP. First, this objective was accomplished by constructing the engineered strain PS100, in which BAPAT from *P. aeruginosa* PA01 (encoded by its *pa0132* gene), ADC from *C. glutamicum* ATCC13032 (encoded by its *panD* gene), and HADH from *E. coli* K-12 MG1655 (encoded by its *ydfG* gene) were overexpressed in pETM9 backbone in the pseudo-operon configuration. Shake flasks cultivation with this strain produced 3-HP up to 0.338 ± 0.044 g/L after 24 h of induction.

To further increase the 3-HP titer in the production strain, other metabolic targets were identified to redirect carbon flux towards product formation. Two of these targets were phosphoenolpyruvate carboxykinase (PCK) from *A. succinogenes* and L-aspartate aminotransferase (AAT) from *E. coli*. However, these enzymes overexpression did not improve the 3-HP final titer of the engineered strains compared to PS100.

Another three targets were found using *in silico* analysis of the genome-scale metabolic model *i*ML1515, modified to include the BAPAT reaction. The *in silico* optimization results found three non-intuitive targets that could potentially favor 3-HP production: L-alanine aminotransferase and alanine racemase overexpression, ALR and ALAT, respectively, and L-valine transaminase (BCAT) knock-out. The strain PSO107 carrying these three modifications (*ilvE* knock-out and *ALT1* and *alr* overexpression, besides *panD, pa0132*, and *ydfG*) performed better than the strain PS100 producing up to 0.743 ± 0.016 g/L of 3-HP from glucose as carbon source and 1.147 ± 0.015 g/L from glucose and xylose mixture. Although the final titers reached are lower than the optimal values presented in the literature for 3-HP production in *E. coli* through the β-alanine pathway, of about 3.69 g/L in shake flasks cultivations, it is worth mentioning that combining the targets described in this work with some classical modifications described in the literature can configure a potential route to improve the PSO107 strain even further.

6 FUTURE PROSPECTS

To further improve the 3-HP production in the best *E. coli* strain obtained in this work, the next steps could be:

i. Choose another ADC or improve the catalytic activity of ADC: the reaction catalyzed by this enzyme was reported as the bottleneck for 3-HP production. Therefore, efforts could focus on screening other options like ADC from *B. subtilis* and its mutagenized versions. Some works also claim that the cysteine sulfinic acid decarboxylase from *Tribolium castaneum* is a good choice, maybe better than those from bacteria.

ii. Overexpress new enzymes for the PEP carboxylation reaction: the overexpression of this reaction is an important step to channel carbon flux from glycolysis towards product formation and should be targetet for future strain engineering steps. The PCK from *B. subtilis*, the native *E. coli* PPC and the same enzyme from *C. glutamicum* were reported as suitable candidates, for instance.

iii. Clone *aspC* in the pRAr plasmid to assess the combination of AAT overexpression together with the targets from evolutionary optimizations, since the conversion of OAA into L-aspartate is also a transamination step that employs L-glutamate.

iv. Screen for new candidates of alanine racemases and L-alanine aminotransferases and compare their performance with those applied in PSO107.

v. Find new targets for reactions knock-out to further block carbon flux to competitive pathways.

vi. Scale-up the fermentation trials, cultivating PSO107 and further strains derived from it in bench bioreactors, also evaluating new cultivation strategies and process parameters.

In summary, there is still much work to be done to further optimize the strain obtained in this Master's thesis, but the results obtained here are definitely a starting point in this endeavor and show that valuable metabolic engineering strategies can come from *in silico* models and data analysis.

7 REFERENCES

ADOM, F. *et al.* Life-cycle fossil energy consumption and greenhouse gas emissions of bioderived chemicals and their conventional counterparts. **Environmental Science and Technology**, v. 48, n. 24, p. 14624–14631, 2014.

ANDREEVA, I. G.; GOLUBEVA, L. I.; KATASHKINA, J. I. Putative GalP, XylE and FucP H^+ symporters from *Pantoea anantis* are capable of transporting glucose into *Escherichia coli* cells. **Applied Biochemistry and Microbiology**, v. 49, n. 7, p. 638-645, 2013.

BECKER, J. *et al.* Top value platform chemicals: Bio-based production of organic acids. **Current Opinion in Biotechnology**, v. 36, n. Figure 1, p. 168–175, 2015.

BORODINA, I. *et al.* Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in *Saccharomyces cerevisiae* via β-alanine. **Metabolic Engineering**, v. 27, p. 57–64, 2015.

BRUINENBERG, P. M. *et al.* NADH-linked aldose reductase: the key to anaerobic alcoholic fermentation of xylose by yeasts. **Applied Microbiology and Biotechnology**, v. 19, n. 4, p. 256–260, 1984.

BÜCKLE-VALLANT, V. *et al.* Metabolic engineering of *Corynebacterium glutamicum* for 2-ketoisocaproate production. **Applied Microbiology and Biotechnology**, v. 98, n. 1, p. 297–311, 2014.

CHEN, Y. *et al.* Coupled incremental precursor and co-factor supply improves 3-hydroxypropionic acid production in *Saccharomyces cerevisiae*. **Metabolic Engineering**, v. 22, p. 104–109, 2014.

CHEN, Y.; NIELSEN, J. Biobased organic acids production by metabolically engineered microorganisms. **Current Opinion in Biotechnology**, v. 37, p. 165–172, 2016.

CHEN, Z. *et al.* Metabolic engineering of *Corynebacterium glutamicum* for the production of 3-hydroxypropionic acid from glucose and xylose. **Metabolic Engineering**, v. 39, n. September 2016, p. 151–158, 2017b.

CHOI, S. *et al.* Biorefineries for the production of top building block chemicals and their

derivatives. **Metabolic Engineering**, v. 28, p. 223–239, 2015.

CONCORDET, J.; HAEUSSLER, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. **Nucleic Acids Research**, v. 46, p. W242-W245, 2018.

CORMA CANOS, A.; IBORRA, S.; VELTY, A. Chemical routes for the transformation of biomass into chemicals. **Chemical Reviews**, v. 107, n. 6, p. 2411–2502, 2007.

CORTAY, J. *et al.* Regulation of the acetate operon in *Escherichia coli*: purification and functional characterization of the IclR repressor. **The EMBO Journal**, v. 10, n. 3, p. 675-679, 1991.

DESAI, T. A.; RAO, C. V. Regulation of arabinose and xylose metabolism in *Escherichia coli*. **Applied and Environmental Microbiology**, v. 76, n. 5, p. 1524-1532, 2010.

DIEN, S. VAN. From the first drop to the first truckload: Commercialization of microbial processes for renewable chemicals. **Current Opinion in Biotechnology**, v. 24, n. 6, p. 1061–1068, 2013.

DUFF, S. M. G. *et al.* The Enzymology of alanine aminotransferase (AlaAT) isoforms from *Hordeum vulgare* and other organisms, and the HvAlaAT crystal structure. **Archives of Biochemistry and Biophysics**, v. 528, n. 1, p. 90–101, 2012.

FOUCHÉCOUR, F. DE *et al.* Process engineering for microbial production of 3-hydroxypropionic acid. **Biotechnology Advances**, v. 36, n. 4, p. 1207–1222, 2018.

GARCÍA-CAMPUSANO, F. *et al. ALT1* -encoded alanine aminotransferase plays a central role in the metabolism of alanine in *Saccharomyces cerevisiae*. **Canadian Journal of Microbiology**, v. 55, n. 4, p. 368–374, 2009.

GONZALEZ, J. E.; LONG, C. P.; ANTONIEWICZ, M. R. Comprehensive analysis of glucose and xylose metabolism in *Escherichia coli* under aerobic and anaerobic conditions by ^{13}C metabolic flux analysis. **Metabolic Engineering**, v. 39, n. July 2016, p. 9–18, 2017.

GÖRKE, B.; STÜLKE, J. Carbon catabolite repression in bacteria: Many ways to make the most out of nutrients. **Nature Reviews Microbiology**, v. 6, n. 8, p. 613–624, 2008.

GROFF, D. *et al.* Supplementation of intracellular XylR leads to coutilization of hemicellulose sugars. **Applied and Environmental Microbiology**, v. 78, n. 7, p. 2221–2229, 2012.

GUNUKULA, S.; RUNGE, T.; ANEX, R. Assessment of Biocatalytic Production Parameters to Determine Economic and Environmental Viability. **ACS Sustainable Chemistry and Engineering**, v. 5, n. 9, p. 8119–8126, 2017.

HAMMER, S. K.; AVALOS, J. L. Uncovering the role of branched-chain amino acid transaminases in *Saccharomyces cerevisiae* isobutanol biosynthesis. **Metabolic Engineering**, v. 44, n. September, p. 302–312, 2017.

HEO, W. *et al.* Enhanced production of 3-hydroxypropionic acid from glucose and xylose by alleviation of metabolic congestion due to glycerol flux in engineered *Escherichia coli*. **Bioresource Technology**, v. 285, n. February, p. 121320, 2019.

HÖFER, M.; BETZ, A.; KOTYK, A. Metabolism of the obligatory aerobic yeast *Rhodotorula gracilis* IV. Induction of an enzyme necessary for D-xylose catabolism. **BBA - General Subjects**, v. 252, n. 1, p. 1–12, 1971.

HOLÁTKO, J. *et al.* Metabolic engineering of the L-valine biosynthesis pathway in *Corynebacterium glutamicum* using promoter activity modulation. **Journal of Biotechnology**, v. 139, n. 3, p. 203–210, 2009.

HÜSER, A. T. *et al.* Rational design of a *Corynebacterium glutamicum* pantothenate production strain and its characterization by metabolic flux analysis and genome-wide transcriptional profiling. **Applied and Environmental Microbiology**, v. 71, n. 6, p. 3255–3268, 2005.

JECKELMANN, J.; ERNI, B. Transporters of glucose and other carbohydrates in bacteria. **European Journal of Physiology**, v. 472, p. 1129-1153, 2020.

JIANG, M. *et al.* Co-expression of phosphoenolpyruvate carboxykinase and nicotinic acid phosphoribosyltransferase for succinate production in engineered *Escherichia coli*. **Enzyme and Microbial Technology**, v. 56, p. 8-14, 2014.

JIANG, X.; MENG, X.; XIAN, M. Biosynthetic pathways for 3-hydroxypropionic acid production. **Applied Microbiology and Biotechnology**, v. 82, n. 6, p. 995–1003, 2009.

JU, J. *et al.* Correlation between catalytic activity and monomer-dimer equilibrium of bacterial alanine racemases. **Journal of Biochemistry**, v. 149, n. 1, p. 83–89, 2011.

JUNG, I. Y. *et al.* Simultaneous conversion of glucose and xylose to 3-hydroxypropionic acid in engineered coli by modulation of sugar transport and glycerol synthesis. **Bioresource Technology**, v. 198, p. 709–716, 2015.

KATSUBE, S. *et al.* Secretion of D-alanine by *Escherichia coli*. **Microbiology (United Kingdom)**, v. 162, n. 7, p. 1243–1252, 2016.

KILDEGAARD, K. R *et al*. Production of 3-hydroxypropionic acid from glucose and xylose by metabolically engineered *Saccharomyces cerevisiae*. **Metabolic Engineering Communications**, v. 2, p. 132–136, 2015.

KILDEGAARD, K. R. *et al.*. Engineering and systems-level analysis of *Saccharomyces cerevisiae* for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway. **Microbial Cell Factories**, v. 15, n. 1, p. 1–13, 2016.

KIM, S. H.; SCHNEIDER, B. L.; REITZER, L. Genetics and regulation of the major enzymes of alanine synthesis in *Escherichia coli*. **Journal of Bacteriology**, v. 192, n. 20, p. 5304–5311, 2010.

KIM, P. *et al.* Effect of overexpression of *Actinobacillus succinogenes* phosphoenolpyruvate carboxykinase on succinate production in *Escherichia coli*. **Applied and Environmental Microbiology**, v. 70, n. 2, p. 1238-1241, 2004.

KLAMT, S.; HÄDICKE, O.; KAMP, A. VON. Stoichiometric and constraint-based analysis of biochemical reaction networks. *In*: BENNER, P. *et al.* **Large-Scale Networks in Engineering and Life Sciences**. Birkhäuser, 2016. p. 263-316.

KO, Y. S. *et al.* Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. **Chemical Society Reviews**, v. 49, n. 14, p. 4615–4636, 2020.

KRAUSE, F. S.; BLOMBACH, B.; EIKMANNS, B. J. Metabolic engineering of *Corynebacterium glutamicum* for 2-Ketoisovalerate production. **Applied and Environmental Microbiology**, v. 76, n. 24, p. 8053–8061, 2010.

KUMAR, V.; ASHOK, S.; PARK, S. Recent advances in biological production of 3-hydroxypropionic acid. **Biotechnology Advances**, v. 31, n. 6, p. 945–961, 2013.

KUMARI, S. *et al.* Cloning, characterization, and functional expression of acs, the gene which encodes acetyl coenzyme A synthetase in *Escherichia coli*. **Journal of Bacteriology**, v. 177, n. 10, p. 2878–2886, 1995.

KUMARI, S. *et al.* Regulation of acetyl coenzyme A synthetase in *Escherichia coli*. **Journal of Bacteriology**, v. 182, n. 15, p. 4173–4179, 2000.

KURAMITSU, S. *et al.* Nucleotide Amino Sequence of the *ilvE* Gene and the Deduced Acid Sequence1. **J. Biochem.**, v. 97, n. 4, p. 993–999, 1985.

KWAK, S. *et al.* Production of biofuels and chemicals from xylose using native and engineered yeast strains. **Biotechnology Advances**, v. 37, n. 2, p. 271–283, 2019.

LACMATA, S. T. *et al.* Enhanced poly(3-hydroxypropionate) production via β-alanine pathway in recombinant *Escherichia coli*. **PLoS ONE**, v. 12, n. 3, p. 1–11, 2017.

LAN, E. I. *et al.* Metabolic engineering of cyanobacteria for photosynthetic 3-hydroxypropionic acid production from CO_2 using *Synechococcus elongatus* PCC 7942. **Metabolic Engineering**, v. 31, p. 163–170, 2015.

LEE, S. Y. *et al.* A comprehensive metabolic map for production of bio-based chemicals. **Nature Catalysis**, v. 2, p. 18-33, 2019.

LEWIS, N. E. *et al.* Omic data from evolved *E . coli* are consistent with computed optimal growth from genome-scale models. **Molecular Systems** Biology, v. 6, n. 390, p. 1-13, 2010.

LEWIS, N. E.; NAGARAJAN, H.; PALSSON, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. **Nature Reviews Microbiology**, v. 10, n. 4, p. 291–305, 2012.

LIAO, H. H. *et al.* **Alanine 2,3 aminomutase** US Patent 20080124785A1. Deposit: 9 nov. 2007. Concession: 29 may, 2008.

LIAO, H. H. *et al.* **Production of 3-hydroxypropionic acid using beta-alanine/pyruvate**

aminotransferase US Patent 009124388B2. Deposit: 29 jun. 2010. Concession: 28 fev. 2012.

LIN, H. *et al.* Acetyl-CoA synthetase overexpression in *Escherichia coli* demonstrates more efficient acetate assimilation and lower acetate accumulation: A potential tool in metabolic engineering. **Applied Microbiology and Biotechnology**, v. 71, n. 6, p. 870–874, 2006.

LIS, A. V. *et al.* Exploring small-scale chemostats to scale up microbial processes: 3-hydroxypropionic acid production in *S. cerevisiae.* **Microbial Cell Factories**, v. 18, n. 1, p. 1-11, 2019.

LIU, C. *et al.* Dissection of Malonyl-Coenzyme A Reductase of *Chloroflexus aurantiacus* Results in Enzyme Activity Improvement. **PLoS ONE**, v. 8, n. 9, p. 1–8, 2013.

LIU, C. *et al*. Functional balance between enzymes in malonyl-CoA pathway for 3-hydroxypropionate biosynthesis. **Metabolic Engineering**, v. 34, p. 104–111, 2016.

LIU, R. *et al.* Directed combinatorial mutagenesis of *Escherichia coli* for complex phenotype engineering. **Metabolic Engineering**, v. 47, n. March, p. 10–20, 2018b.

LIU, R. *et al.* Fermentation of xylose to succinate by enhancement of ATP supply in metabolically engineered *Escherichia coli*. **Applied Microbiology and Biotechnology**, v. 94, n. 4, p. 959-968, 2012.

LIU, R. *et al.* Iterative genome editing of *Escherichia coli* for 3-hydroxypropionic acid production. **Metabolic Engineering**, v. 47, n. April, p. 303–313, 2018a.

LOH, K. D. *et al.* A previously underscribed pathway for pyrimidine catabolism. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 13, p. 5114–5119, 2006.

LYNCH, M. D.; GILL, R. T.; LIPSOCOMB, T. E. W. **Methods for producing 3-hydroxypropionic acid and other products** US Patent 20130071893A1. Deposit: 27 nov. 2010. Concession: 21 mar. 2013.

MAIA, P.; ROCHA, M.; ROCHA, I. *In Silico* Constraint-Based Strain Optimization Methods: the Quest for Optimal Cell Factories . **Microbiology and Molecular Biology Reviews**,

v. 80, n. 1, p. 45–67, 2016.

MARTINEZ, R. *et al.* The XylR variant (R121C and P363S) releases arabinose-induced catabolite repression on xylose fermentation and enhances coutilization of lignocellulosic sugar mixtures. **Biotechnology and Bioengineering**, v. 116, n. 12, p. 3476–3481, 2019.

MATSUOKA, Y.; SHIMIZU, K. Catabolite regulation analysis of *Escherichia coli* for acetate overflow mechanism and co-consumption of multiple sugars based on systems biology approach using computer simulation. **Journal of Biotechnology**, v. 168, p. 155-173, 2013.

MCALLISTER, C. H. *et al.* Analysis of the Enzymatic Properties of a Broad Family of Alanine Aminotransferases. **PLoS ONE**, v. 8, n. 2, 2013.

MENG, J. *et al.* High-yield anaerobic succinate production by strategically regulating multiple metabolic pathways based on stoichiometric maximum in *Escherichia coli*. **Microbial Cell Factories**, v. 15, n. 1, p. 1-13, 2016.

MO, Q. *et al.* Identification of mutations restricting autocatalytic activation of bacterial L-aspartate α-decarboxylase. **Amino Acids**, v. 50, n. 10, p. 1433-1440, 2018.

MONK, J. M. *et al. i*ML1515, a knowledgebase that computes *Escherichia coli* traits. **Nature Biotechnology**, v. 35, n. 10, p. 904-908, 2017.

NI, L. *et al.* Structures of the *Escherichia coli* transcription activator and regulator of diauxie, XylR: An AraC DNA-binding family member with a LacI/GalR ligand-binding domain. **Nucleic Acids Research**, v. 41, n. 3, p. 1998–2008, 2013.

NOVAK, K. *et al.* Characterizing the effect of expression of an acetyl-CoA synthetase insensitive to acetylation on co-utilization of glucose and acetate in batch and continuous cultures of *E. coli* W. **Microbial Cell Factories**, v. 17, n. 1, p. 1–15, 2018.

OCHMAN, H.; GERBER, A. S.; HARTL, D. L. Genetic Applications of an Inverse Polymerase Chain Reaction. **Genetics Society of America**, v. 120, p. 621-623, 1988.

PARK, Y. S. *et al.* Engineering an aldehyde dehydrogenase torwards its substrates, 3-hydroxypropanal and NAD+, for enhancing the production of 3-hydroxypropionic acid. **Scientific Reports,** v. 7, n. 17155, p. 1-12, 2017.

PIAO, X. *et al*. Metabolic engineering of *Escherichia coli* for production of L-aspartate and is derivative β-alanine with high stoichiometric yield. **Metabolic Engineering**, v. 54, p. 244-254, 2019.

PINA, C. DELLA; FALLETTA, E.; ROSSI, M. A green approach to chemical building blocks. the case of 3-hydroxypropanoic acid. **Green Chemistry**, v. 13, n. 7, p. 1624–1632, 2011.

QIAN, Y. *et al.* Production of β-alanine from fumaric acid using a dual-enzyme cascade. **ChemCatChem**, v. 10, n. 21, p. 4998-5005, 2018.

QUAN, J.; TIAN, J. Circular polymerase extension cloning of complex gene libraries and pathways. **PLoS ONE**, v. 4, n. 7, 2009.

RATHNASINGH, C. *et al.* Production of 3-hydroxypropionic acid via malonyl-CoA pathway using recombinant *Escherichia coli* strains. **Journal of Biotechnology**, v. 157, n. 4, p. 633–640, 2012.

REISCH, C. R.; PRATHER, K. L. J. Scarless Cas9 assisted recombineering (no-SCAR) in *Escherichia coli*, an easy-to-use system for genome editing. **Current Protocols in Molecular Biology**, v. 2017, n. January, p. 1–20, 2017.

REISCH, C. R.; PRATHER, K. L. J. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. **Scientific Reports**, v. 5, n. June, p. 1–12, 2015.

ROCHA, I. *et al. OptFlux*: An open-source software platform for in silico metabolic engineering. **BMC Systems Biology**, v. 4, 2010.

SAMBROOK, J.; RUSSEL, D. W. **Molecular Cloning:** a laboratory manual. 3RD edition. New York: Cold Spring Harbor Press, 2001.

SANKARANARAYANAN, M. *et al.* Production of 3-hydroxypropionic acid by balancing the pathway enzymes using synthetic cassette architecture. **Journal of Biotechnology**, v. 259, p. 140-147, 2017.

SHEN, Y. *et al.* Synthesis of β-alanine from L-aspartate using L-aspartate α-decarboxylase from *Corynebacterium glutamicum.* **Biotechnology Letters**, v. 36, n. 8, p. 1681-1686, 2014.

SIEVERT, C. *et al.* Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. **Proceedings of the National Academy of Sciences of the United States of America**, v. 114, n. 28, p. 7349–7354, 2017.

SILVA, A. J. da, *et al.* Metabolic engineering of *E. coli* for pyocyanin production. **Metabolic Engineering**, v. 64, p. 15-25, 2021.

SONG, C. W. *et al*. Metabolic engineering of *Escherichia coli* for the production of 3-aminopropionic acid. **Metabolic Engineering**, v. 30, p. 121–129, 2015.

SONG, C. W. *et al*. Metabolic Engineering of *Escherichia coli* for the Production of 3-Hydroxypropionic Acid and Malonic Acid through β-Alanine Route. **ACS Synthetic Biology**, v. 5, n. 11, p. 1256–1263, 2016.

SONG, C. W. *et al.* Metabolic engineering of Escherichia coli for the production of fumaric acid. **Biotechnology and Bioengineering**, v. 110, n. 7, p. 2025–2034, 2013.

SONG, S.; PARK, C. Organization and regulation of the D-xylose operons in Escherichia coli K-12: XylR acts as a transcriptional activator. **Journal of Bacteriology**, v. 179, n. 22, p. 7025–7032, 1997.

SUYAMA, A. *et al.* Production of 3-hydroxypropionic acid via the malonyl-CoA pathway using recombinan fission yeast strains. **Journal of Bioscience and Bioengineering**, v. 124, n. 4, p. 392-399, 2017.

TAKAYAMA, S. *et al.* Enhancing 3-hydroxypropionic acid production in combination with sugar supply engineering by cell surface-display and metabolic engineering of *Schizosaccharomyces pombe*. **Microbial cell factories**, v. 17, n. 1, p. 176, 2018.

TAMURA, T.; LU, W.; AKUTSU, T. Computational Methods for Modification of Metabolic Networks. **Computational and Structural Biotechnology Journal**, v. 13, p. 376–381, 2015.

TARASAVA, K. *et al.* Combinatorial pathway engineering using type I-E CRISPR interference. **Biotechnology and Bioengineering**, v. 115, n. 7, p. 1878–1883, 2018.

VALDEHUESA, K. N. G. *et al.* Recent advances in the metabolic engineering of microorganisms for the production of 3-hydroxypropionic acid as C3 platform chemical.

Applied Microbiology and Biotechnology, v. 97, n. 8, p. 3309–3321, 2013.

VALGEPEA, K. *et al.* Systems biology approach reveals that overflow metabolism of acetate in Escherichia coli is triggered by carbon catabolite repression of acetyl-CoA synthetase. **BMC Systems Biology**, v. 4, n. 1, p. 166, 2010.

WAEGEMAN, H. *et al.* Effect of *iclR* and *arcA* knockouts on biomass formation and metabolic fluxes in *Escherichia coli* K12 and its implications on understanding the metabolism of *Escherichia coli* BL21 (DE3). **BMC Microbiology**, v. 11, n. 70, p. 1-17, 2011.

WANG, Q. *et al.* Metabolic engineering of *Escherichia coli* for poly(3-hydroxypropionate) production from glycerol and glucose. **Biotechnology Letters**, v. 36, n. 11, p. 2257-2261, 2014.

WERPY, T.; PETERSEN, G. Top Value Added Chemicals from Biomass: Volume I -- Results of Screening for Potential Candidates from Sugars and Synthesis Gas. Office of Scientific and Technical Information (OSTI). **Office of Scientific and Technical Information**, p. 69, 2004.

WESTBROOK, A. W. *et al.* Strain engineering for microbial production of value-added chemicals and fuels from glycerol. **Biotechnology Advances**, v. 37, n. 4, p. 538–568, 2019.

WILD, J. *et al.* Identification of the *dadX* gene coding for the predominant isozyme of alanine racemase in *Escherichia coli* K12. **MGG Molecular & General Genetics**, v. 198, n. 2, p. 315–322, 1985.

XU, P. *et al. EPathBrick:* A synthetic biology platform for engineering metabolic pathways in E. coli. **ACS Synthetic Biology**, v. 1, n. 7, p. 256–266, 2012.

YAO, R.; SHIMIZU, K. Recent progress in metabolic engineering for the production of biofuels and biochemicals from renewable sources with particular emphasis on catabolite regulation and its modulation. **Process Biochemistry**, v. 48, p. 1409-1417, 2013.

YONEYAMA, H. *et al.* Isolation of a mutant auxotrophic for L-Alanine and identification of three major aminotransferases that synthesize L-Alanine In *Escherichia Coli*. **Bioscience, Biotechnology and Biochemistry**, v. 75, n. 5, p. 930–938, 2011.

ZABED, H. M. *et al.* Co-biosynthesis of 3-hydroxypropionic acid and 1,3-propanediol by a newly isolated *Lactobacillus reuteri* strain during whole celll biotransformation of glycerol. **Journal of Cleaner Production**, v. 226, p. 432-442, 2019.

ZHANG, T. *et al.* Glu56Ser mutation improves the enzymatic activity and catalytic stability of *Bacillus subtilis* L-aspartate α-decarboxylase for an efficient β-alanine production. **Process Biochemistry**, v. 70, n. April, p. 117–123, 2018.

ZHAO, P. *et al.* Exploiting tandem repetitive promoters for high-level production of 3-hydroxypropionic acid. **Applied Microbiology and Biotechnology**, v. 103, p. 4017-4031, 2019.

ZITZLER, E.; LAUMANNS, M.; THIELE, L. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. **TIK-Report 103**, may 2001.

APPENDICES

APPENDIX A – *Optflux* information

Table A1 – Optflux reactions and their respective enzymes.

Enzyme	Abbr.	Optflux reaction name	Reaction
Pyruvate dehydrogenase	PDH	R_PDH	$CoA + NAD^+ + pyruvate \leftrightarrow acetyl\text{-}CoA + CO_2 + NADH$
Phosphoenolpryuvate carboxylase	PPC	R_PPC	$CO_2 + H_2O + phosphoenolpyruvate \leftrightarrow H^+ + oxaloacetate + Pi$
Citrate synthase	CS	R_CS	$acetyl\text{-}CoA + H_2O + oxaloacetate \leftrightarrow citrate + CoA + H^+$
Aconitase	ACONT	R_ACONTa and R_ACONTb	citrate ↔ isocitrate
Isocitrate dehydrogenase	ICDH	R_ICDHyr	$isocitrate + NADP^+ \leftrightarrow \alpha\text{-}KG + CO_2 + NADPH$
α-ketoglutarate dehydrogenase	AKGDH	R_AKGDH	$\alpha\text{-}KG + CoA + NAD^+ \leftrightarrow CO_2 + NADH + succinyl\text{-}CoA$
Succinyl-CoA synthetase	SUCOAS	R_SUCOAS	$ATP + CoA + succinate \leftrightarrow ADP + Pi + succinyl\text{-}CoA$
Succinate dehydrogenase	SUCD	R_SUCDi	ubiquinone-8 + succinate ↔ fumarate + ubiquinol-8
Fumarase	FUM	R_FUM	$fumarate + H_2O \leftrightarrow$ L-malate
Malate dehydrogenase	MDH	R_MDH	$L\text{-}malate + NAD^+ \leftrightarrow H^+ + NADH + oxaloacetate$
L-aspartate aminotransferase	AAT	R_ASPTA	α-KG + L-aspartate ↔ L-glutamate + oxaloacetate
Branched-chain amino acid dehydrogenase	BCAT	R_VALTA_L	α-KG + L-valine ↔ L-glutamate + 3-methyl-2-oxobutanoate
L-alanine aminotransferase	ALAT	R_ALATA_L	α-KG + L-alanine ↔ L-glutamate + pyruvate

L-aspartate 1-decarboxylase	ADC	R_ASP1DC	L-aspartate + H^+ ↔ β-alanine + CO_2
β-alanine:pyruvate aminotransferase	BAPAT	R_APATr	β-alanine + pyruvate ↔ malonic semialdehyde + L-alanine
Alanine racemase	ALAR	R_ALAR	L-alanine ↔ D-alanine
D-amino acid dehydrogenase	DAD	R_DAAD	D-alanine + FAD^+ + H_2O ↔ $FADH_2$ + NH_4 + pyruvate
Pyrimidine oxygenase	RUTA	R_PYROX[a]	H^+ + NADH + O_2 → NAD^+ + ureidoacrylate peracid
Peroxyureidoacrylate	RUTB	R_URACPAH[a]	H_2O + ureidoacrylate peracid → H^+ + peroxyaminoacrylate + carbamate
Peroxyaminoacrylate reductase		R_POAACR[a]	NADH + peroxyaminoacrylate → H_2O + NAD^+ + 3-aminoacrylate
3-aminoacrylate hydrolase	RUTD	R_3AMACHYD[a]	H_2O + H^+ + 3-aminoacrylate → NH_4 + malonic semialdehyde
3-hdyroxyacid dehydrogenase	HADH	R_MSAR[a]	H^+ + NADPH + malonic semialdehyde → M_nadp_c + 3-hydroxypropanoate
Glutamine synthetase	GLNS	R_GLNS	ATP + L-glutamate +NH_4 ↔ ADP + L-glutamine + Pi + H^+
Glutamate dehydrogenase	GLD	R_GLUDy	L-glutamate + H_2O + $NADP^+$ ↔ α-KG + H^+ + NADPH + NH_4
Valine:pyruvate aminotransferase	VPAT	R_VPAMTr	L-alanine + 3-methyl-2-oxobutanoate ↔ L-valine + pyruvate

[a]Uracil degradation pathway

Figure A1 depicts the information contained in Figure 27 from the layout tool of *Optflux.*

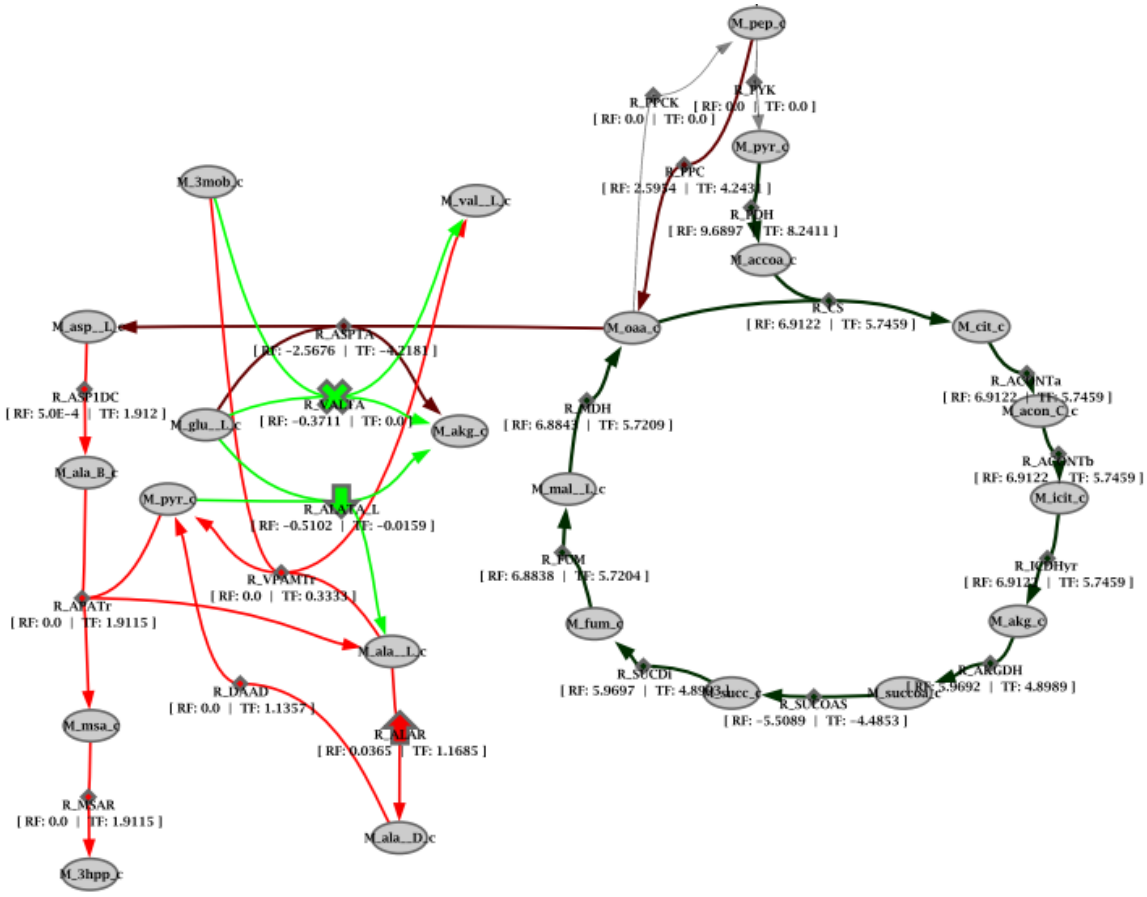


Figure A1 – Comparison of wild-type (RF) and mutant strain (TF) simulations in the β-alanine pathway layout from *Optflux*. Red arrows stand for reactions that had higherabsolute values in the TF simulation, whereas green arrows stand for the ones that had higher absolute values in the RF simulation (in module). Cell growth maximization was set as OF for both strains. The mutant strain (TF) was run with the following genetic conditions: R_ALATA_L value x 0.03125, R_ALAR x 32 and R_VALTA x 0.

Figure A2 simulates the comparison of maximum 3-HP production through the uracil degradation pathway (TF flux values) and the β-alanine pathway (RF flux values) with glucose as carbon source (simulations 3 and 5 from Table 9). It can be seen that the uracil degradation pathway requires more carbon flux through PPP and through the TCA cycle to synthesize the precursors for uracil formation, namely orotate (M_orot_c) and phosphoribosyl phosphate (M_prpp_c).

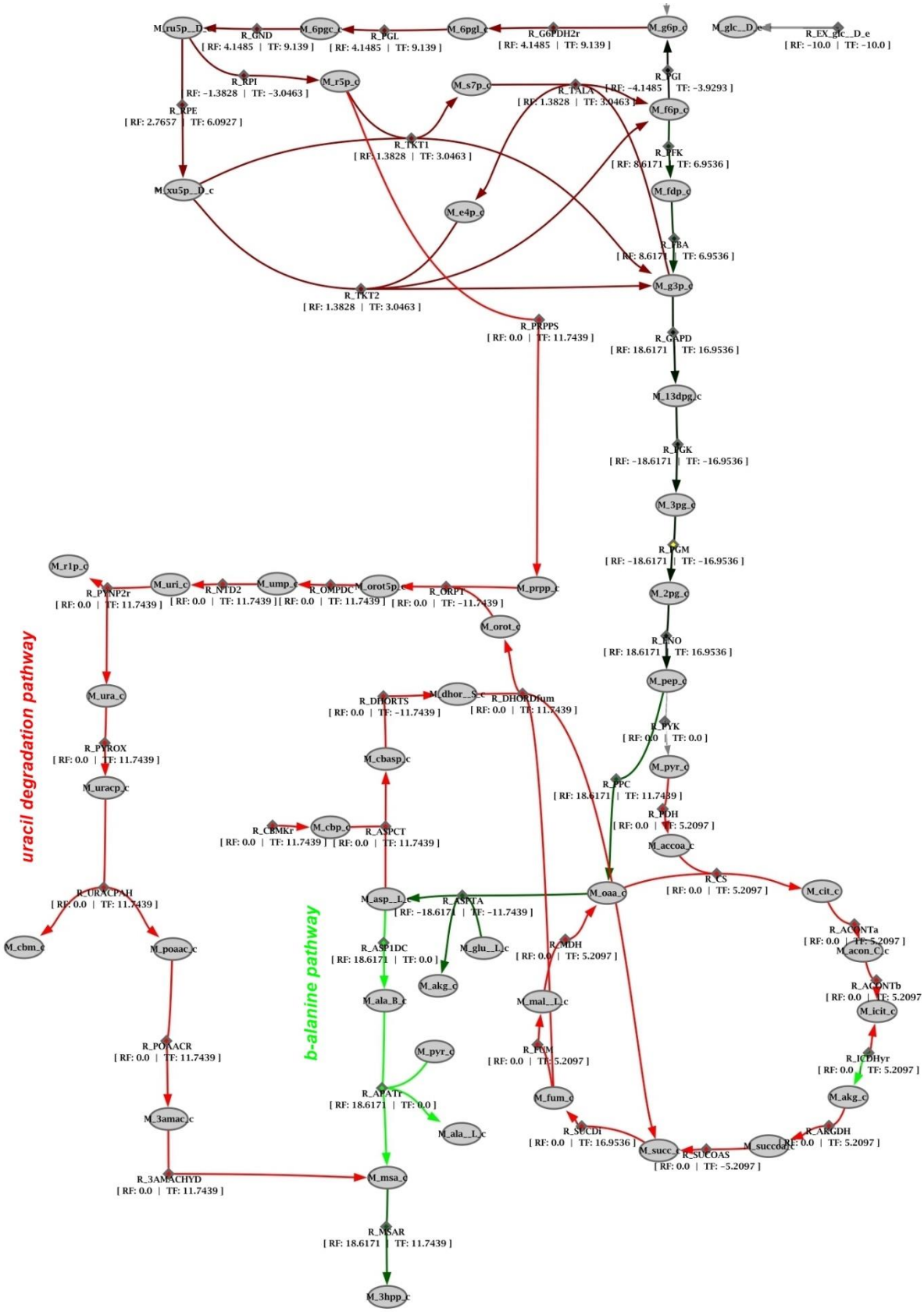


Figure A2 – Simulation comparison of 3-HP maximum production through uracil degradation pathway and β-alanine pathway. Red arrows stand for reactions that had higher absolute values in the TF simulation, whereas green arrows stand for reactions with higher absolute values in

the RF simulation (in module). TF: maximum 3-HP production as OF in the original *i*ML1515 model. RF: maximum 3-HP production as OF in the modified *i*ML1515 model, upon the addiction of R_APATr.

Code for R_APATr inclusion in the *i*ML1515 model:

```
<reaction id="R_APATr" fast="false" reversible="true" name="b-alanine:pyruvate transaminase"
metaid="R_APATr"         sboTerm="SBO:0000176"         fbc:upperFluxBound="cobra_default_ub"
fbc:lowerFluxBound="cobra_default_lb">
        <sbml:annotation xmlns:sbml="http://www.sbml.org/sbml/level3/version1/core">
          <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
            <rdf:Description rdf:about="#R_APATr">
              <bqbiol:is xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
                <rdf:Bag>
                  <rdf:li rdf:resource="http://bigg.ucsd.edu/universal/reactions/APATr"/>
                  <rdf:li
rdf:resource="http://identifiers.org/metanetx.reaction/MNXR95865"/>
                </rdf:Bag>
              </bqbiol:is>
            </rdf:Description>
          </rdf:RDF>
        </sbml:annotation>
        <listOfReactants>
          <speciesReference species="M_ala_B_c" stoichiometry="1" constant="true"/>
          <speciesReference species="M_pyr_c" stoichiometry="1" constant="true"/>
        </listOfReactants>
        <listOfProducts>
          <speciesReference species="M_ala__L_c" stoichiometry="1" constant="true"/>
          <speciesReference species="M_msa_c" stoichiometry="1" constant="true"/>
        </listOfProducts>
</reaction>
```

APPENDIX B – Construction of pETM9 and pRSM6 plasmids

The *XmaJ*I/*Xba*I were swapped in the pETM9 and pRSM6 plasmids to allow the construction of monocistronic and pseudo-operon configurations using *Xba*I instead of *XmaJ*I, which is a much more expensive enzyme than the former. The site swapping was carried out using site-directed mutagenesis.

Primers annealing the target restriction sequences were designed with the desired nucleotide mismatches to generate the site mutagenesis (Table B1). PCR reactions were carried out using these primers and following the QuickChange II Site-Directed Mutagenesis kit (Agilent) protocol. The restriction sites exchange was performed in two rounds of mutagenesis. In the first, the *Xba*I site was replaced by a *XmaJ*I site, producing a plasmid that contained two sites of *XmaJ*I and none of *Xba*I. After, the *XmaJ*I site was replaced by a *Xba*I site, completing the site's exchange.

Table B1 – Primers used for the two-round site-directed mutagenesis of *ePathBrick Vectors* pETM5 and pRSM3 to swap *Xba*I (TCTAGA) and *XmaJ*I (CCTAGG) restriction sites (underlined). Changed bases in bold.

Primer	**5-3' Sequence**
1 mut_*Xba*I_p_*XmaJ*I_F	GAGCGGATAACAATTCCCC**C**CTAG**G**AATAATT
2 mut_*Xba*I_p_*XmaJ*I_R	AAATTATT**C**CTAG**G**GGGGAATTGTTATCCGCT
3 mut_*XmaJ*I_p_*Xba*I_F	CGTCCGGCGTAG**T**CTAG**A**ATCAAGATCGA
4 mut_*XmaJ*I_p_*Xba*I_R	TCGATCTTGAT**T**CTAG**A**CTACGCCGGACG

For each round of site-directed mutagenesis, a 50 µL of a PCR mix containing 25 µL of Platinum™ SuperFi II PCR Master Mix (Invitrogen), 10 ng of the template plasmid, 2.5 µL of primer mix (10 pmol/µL of each primer), and water, was placed to react in a thermal cycler with the following steps: *i.*DNA denaturation at 98 °C for 30 s, *ii.* 18 cycles of DNA denaturation at 98 °C for 10 s, primer annealing at 60 °C for 30 s and DNA extension at 72 °C for 4 min and, finally, *iii.* one more cycle of DNA extension at 72 °C for 5 min. The PCR product was then cooled at 4 °C. After the PCR, 1 µL of *Dpn*I was added to degrade the template

DNA, and the reaction was incubated at 37 °C for 1 h. Then, 1 µL of this reaction was used to transform *E. coli* DH5α accordingly to section 3.3.2.

After the first round, plasmids were extracted and checked by digestion with *Xba*I and *Apa*I restriction enzymes. As the mutated plasmid does not have the site for *Xba*I anymore, only one band is visible in the agarose gel corresponding to the linear plasmid cut with *Apa*I, differently from the pETM7 and pRSM4[3] in which a 1013 bp lane appeared in their respective lanes (Figure B1 – A). The mutated plasmid was submitted to the second round of mutagenesis followed by the restriction digestion confirmation using the same enzymes (Figure B1 – C and D).

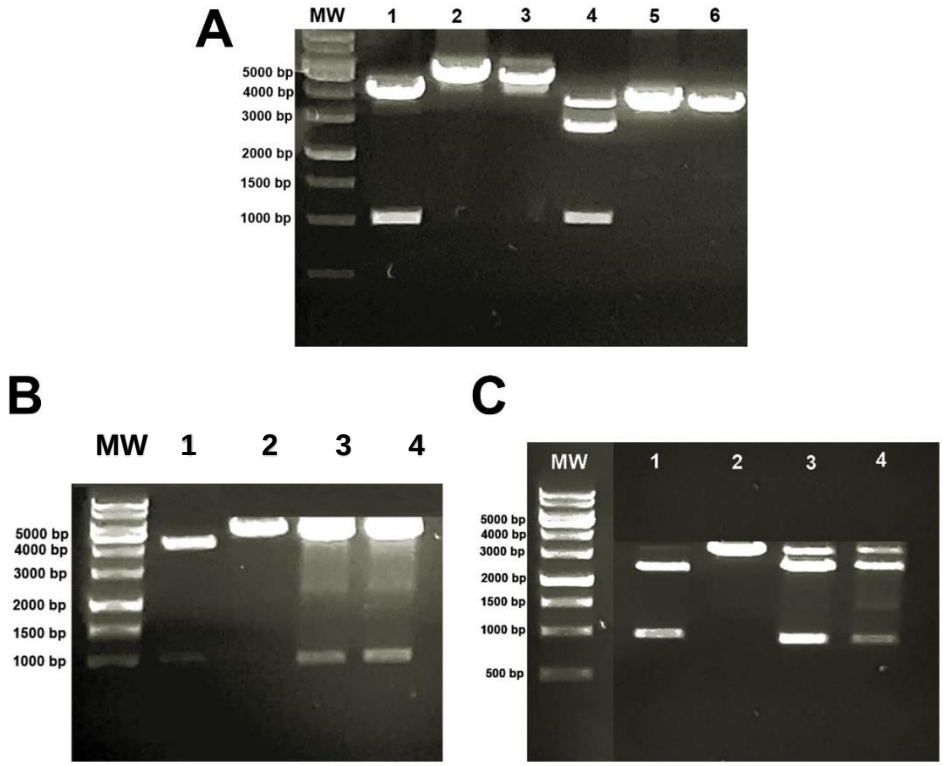


Figure B1 – Confirmation of *XmaJ*I and *Xba*I sites exchange in the plasmids pETM5 and pRSM3. (A) First round of mutagenesis, replacing *Xba*I site for *XmaJ*I. Lanes: 1 pETM6, 2 and 3 mutated pETM6 plasmids, 4 pRSM3, 5 and 6 mutated pRSM3 plasmids. (B) Second round of mutagenesis for pETM6, replacing *XmaJ*I site for *Xba*I. Lanes: 1 pETM6, 2 mutated plasmid from first-round (negative control), 3 and 4 pETM9. (C) Second round of mutagenesis for pRSM3, replacing *XmaJ*I site for *Xba*I. Lanes: 1 pRSM3, 2 mutated plasmid from first-round (negative control), 3 and 4 pRSM6. MW: 1 kb DNA ladder (Sinapse).

[3] See appendix C.

APPENDIX C – Plasmid maps and cloning histories, from SnapGene® v.4.2.11

The pETM9 and pRSM6 plasmids used in this study were derived from the plasmids pETM6 and pRSM3 from the *ePathBrick Vectors* (Xu *et al.*, 2012). These two plasmids are derived from the pET-Duet and pRSF-Duet plasmids from Novagen. In pETM6 and pRSM3, the *T7* promoter was replaced by the *lacUV5* promoter, forming pETM7 and pRSM4 brought to our lab by Prof. Adilson Silva after his post-doctoral studies in the Rensselear Polytechnic Institute – RPI, NY/USA, where these plasmids were first constructed.

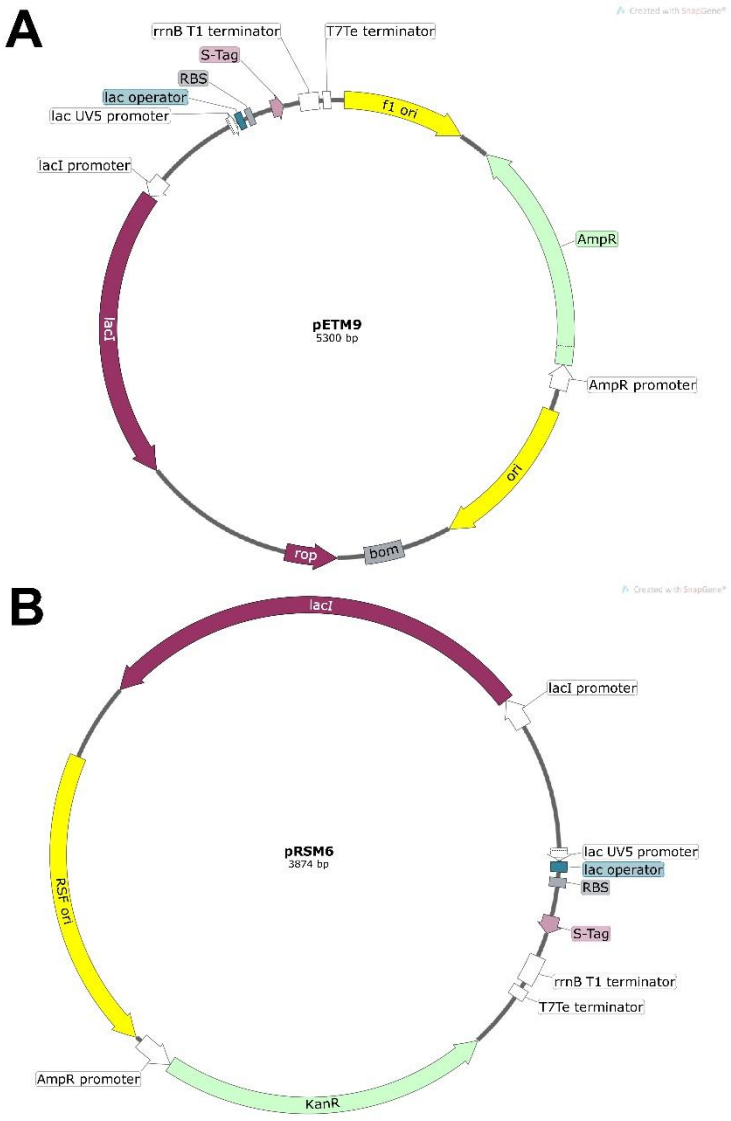


Figure C1 - *ePathBrick vectors* pETM9 (A) and pRSM6 (B).

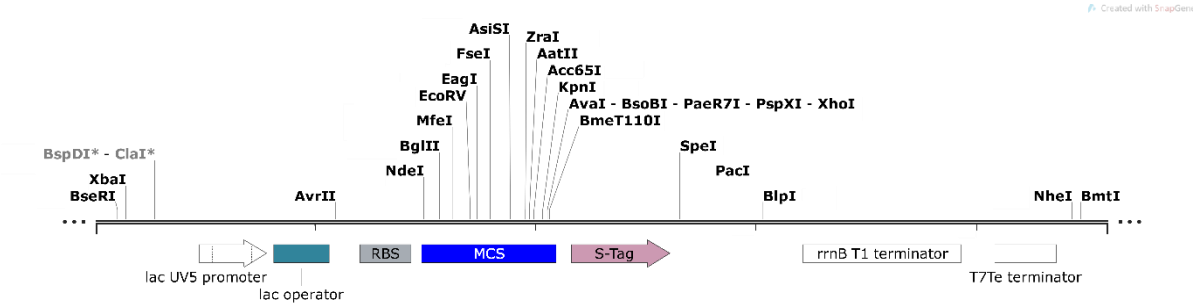


Figure C2 - Promotor – terminator region of *ePathBrick Vectors* pETM9 and pRSM6 in detail. *Avr*II = *XmaJ*I, and *Spe*I = *Bcu*I.

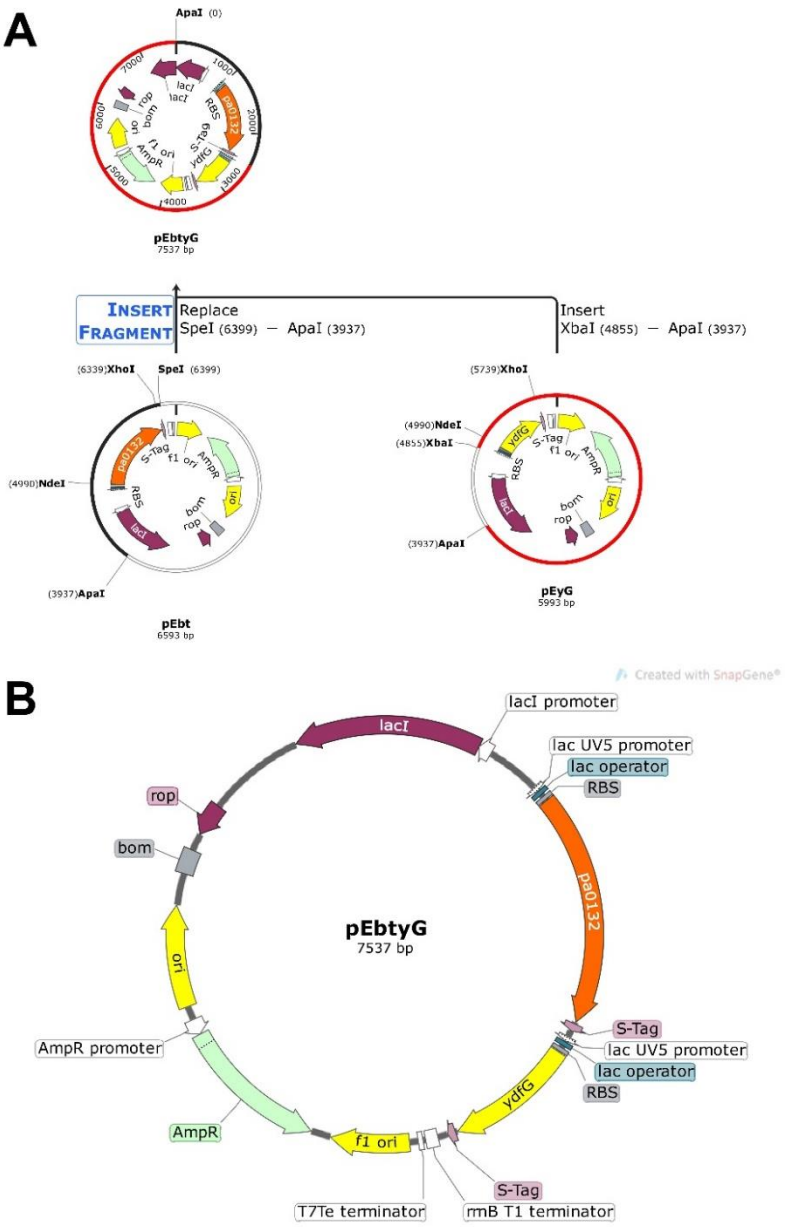


Figure C3 - pEbtyG cloning history (A) and the plasmid map (B)

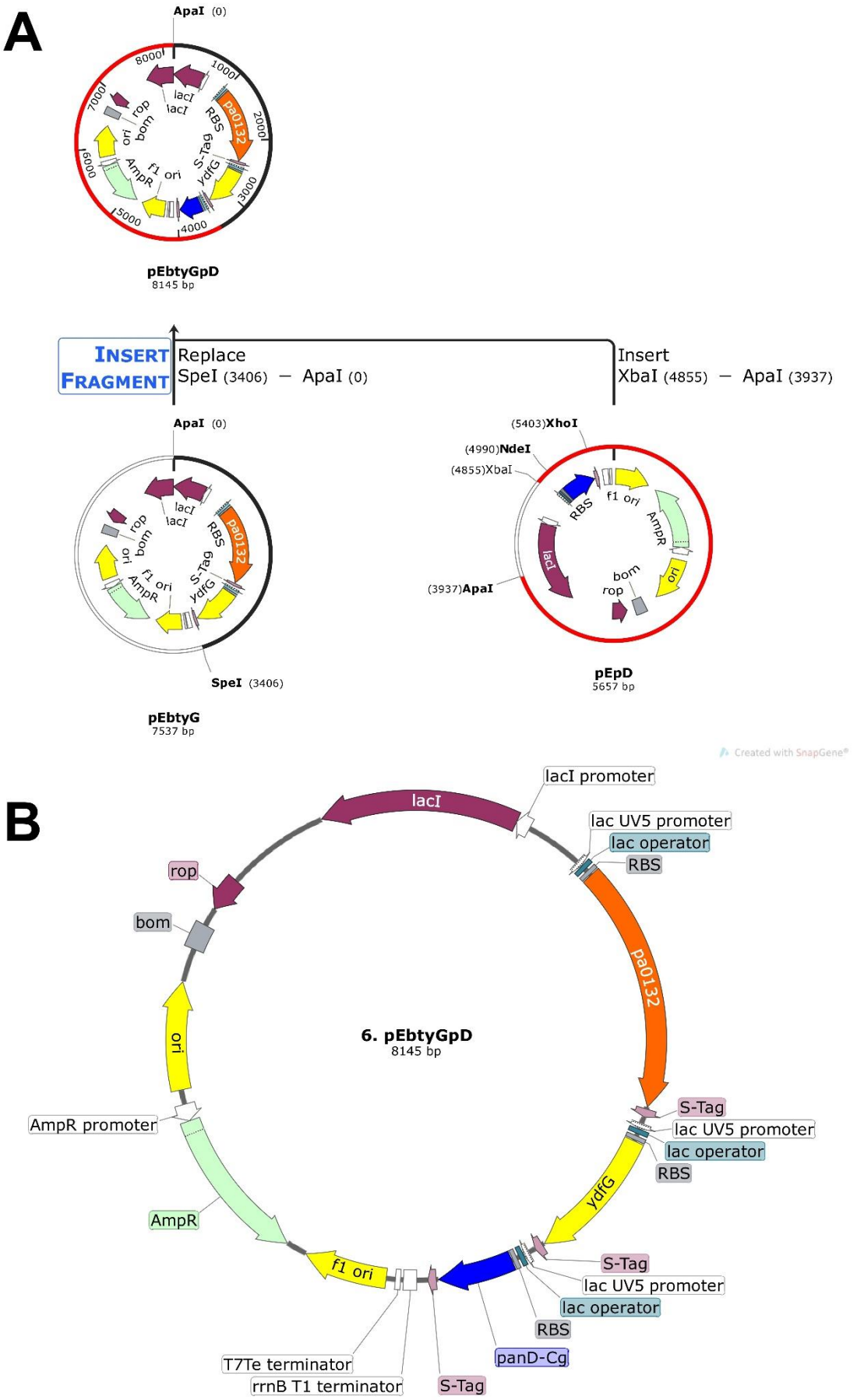


Figure C4 - pEbtyGpD cloning history (A) and the plasmid map (B)

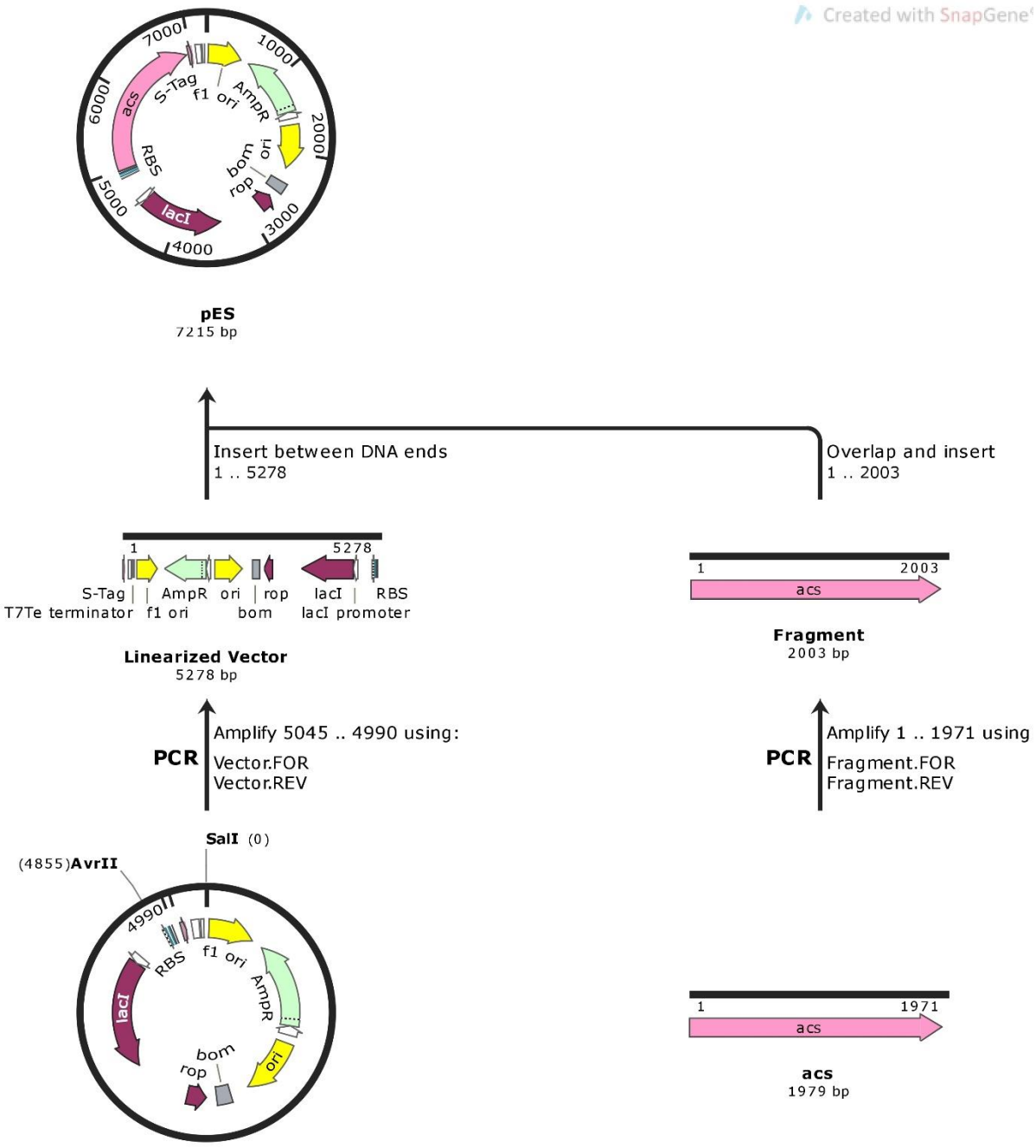


Figure C5 - pES cloning by CPEC

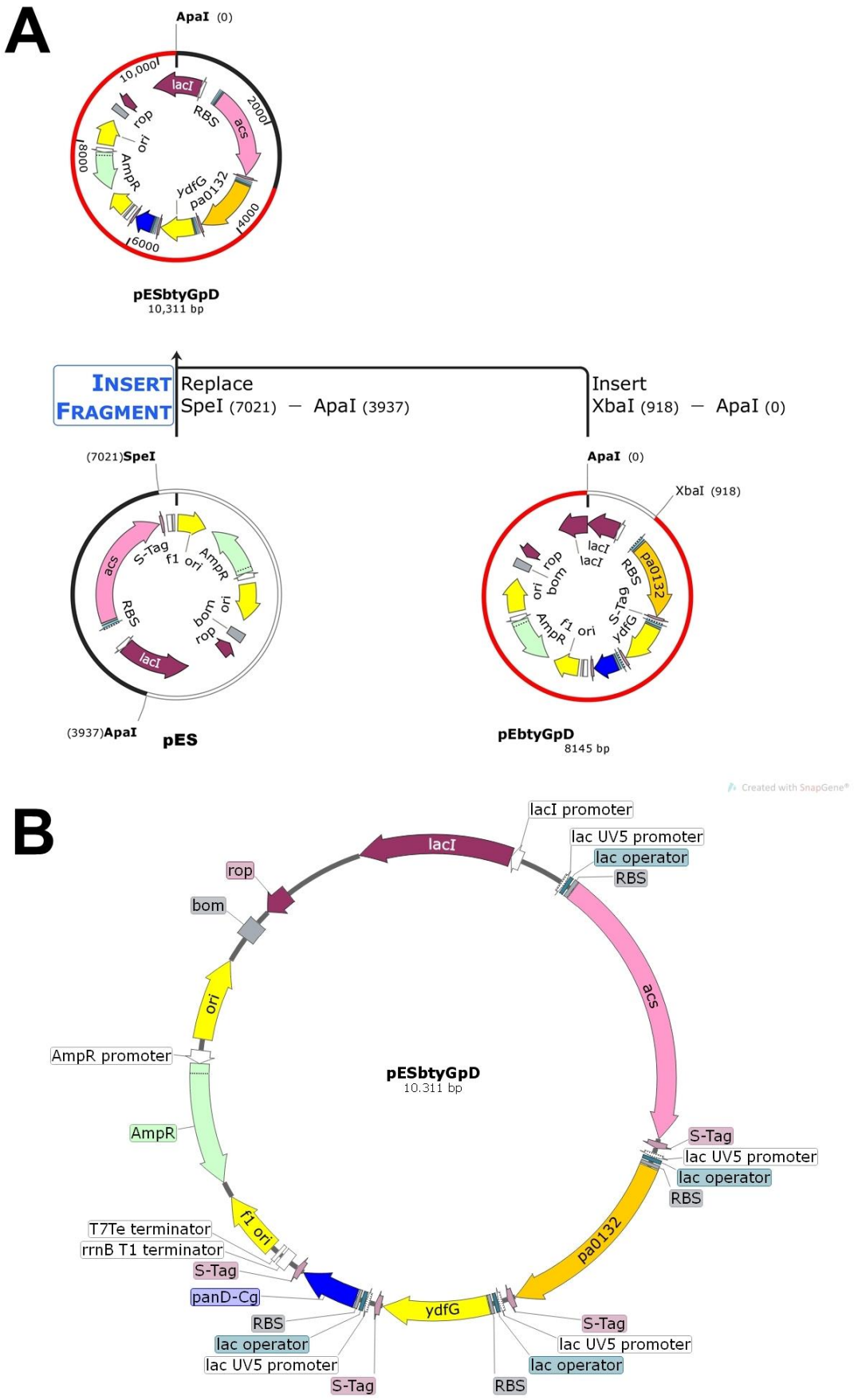


Figure C6 - pESbtyGpD plasmid cloning history (A) and map (B).

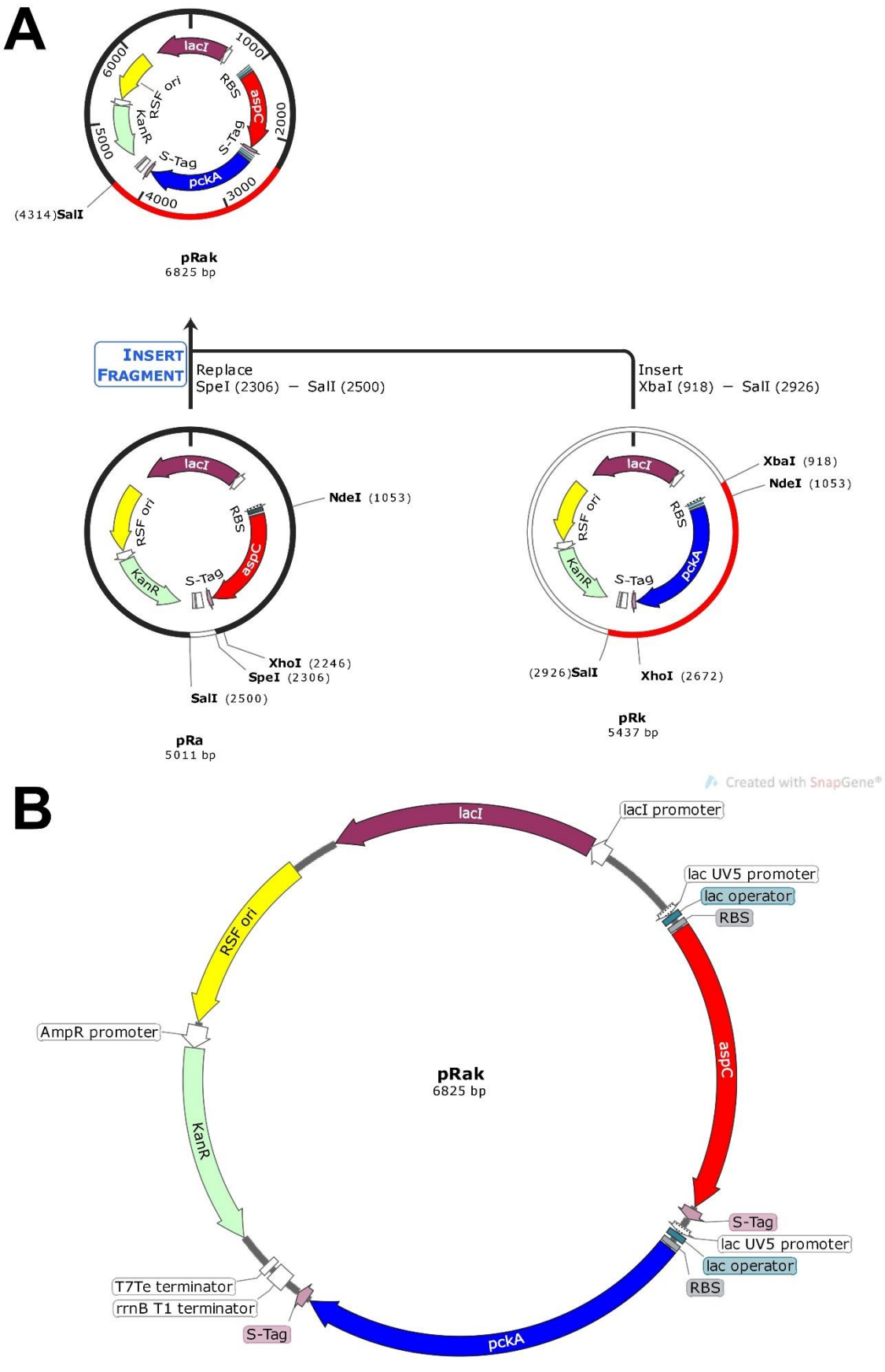


Figure C7 - pRak plasmid cloning history (A) and map (B).

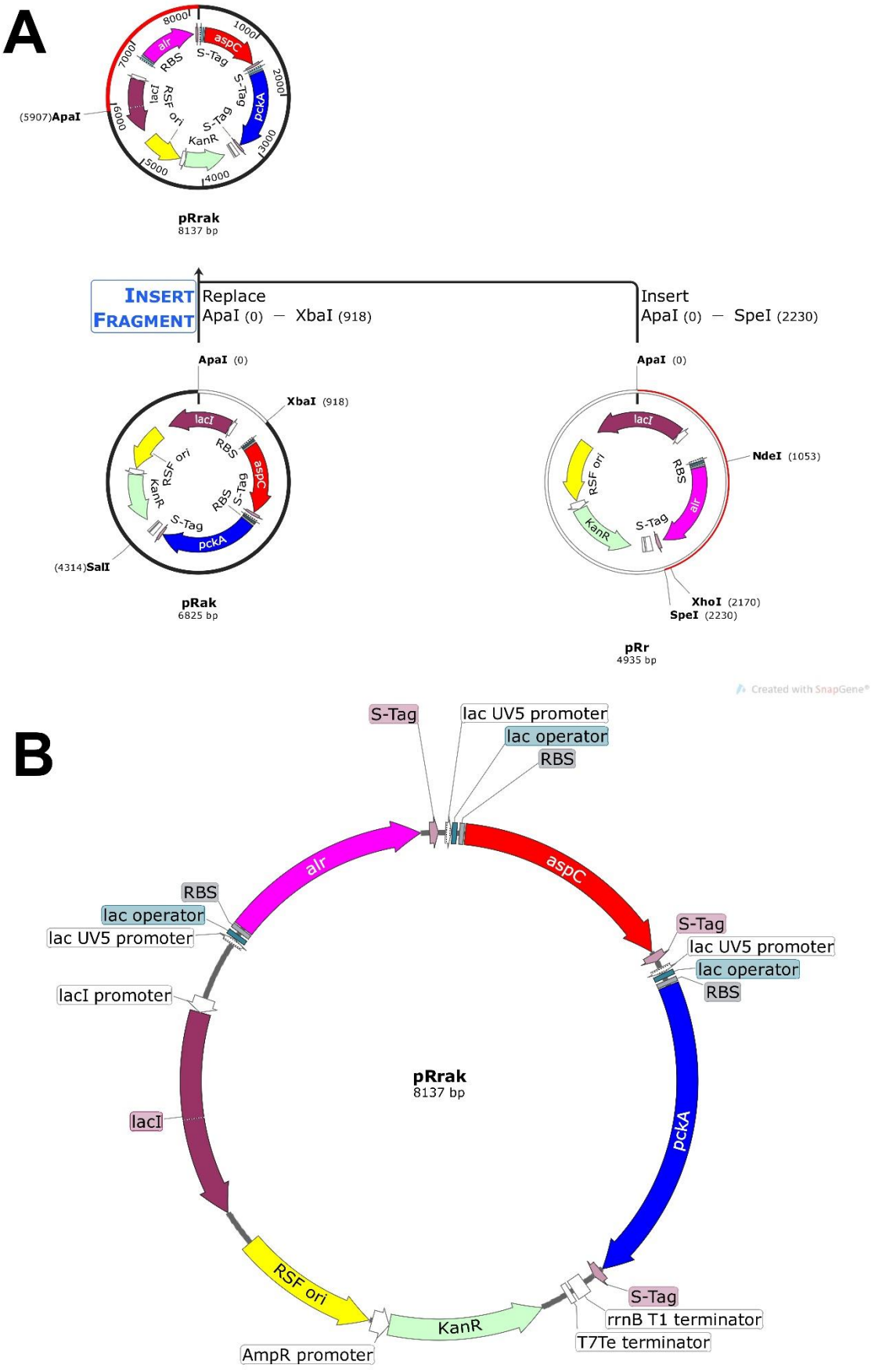


Figure C8 - pRrak plasmid cloning history (A) and map (B).

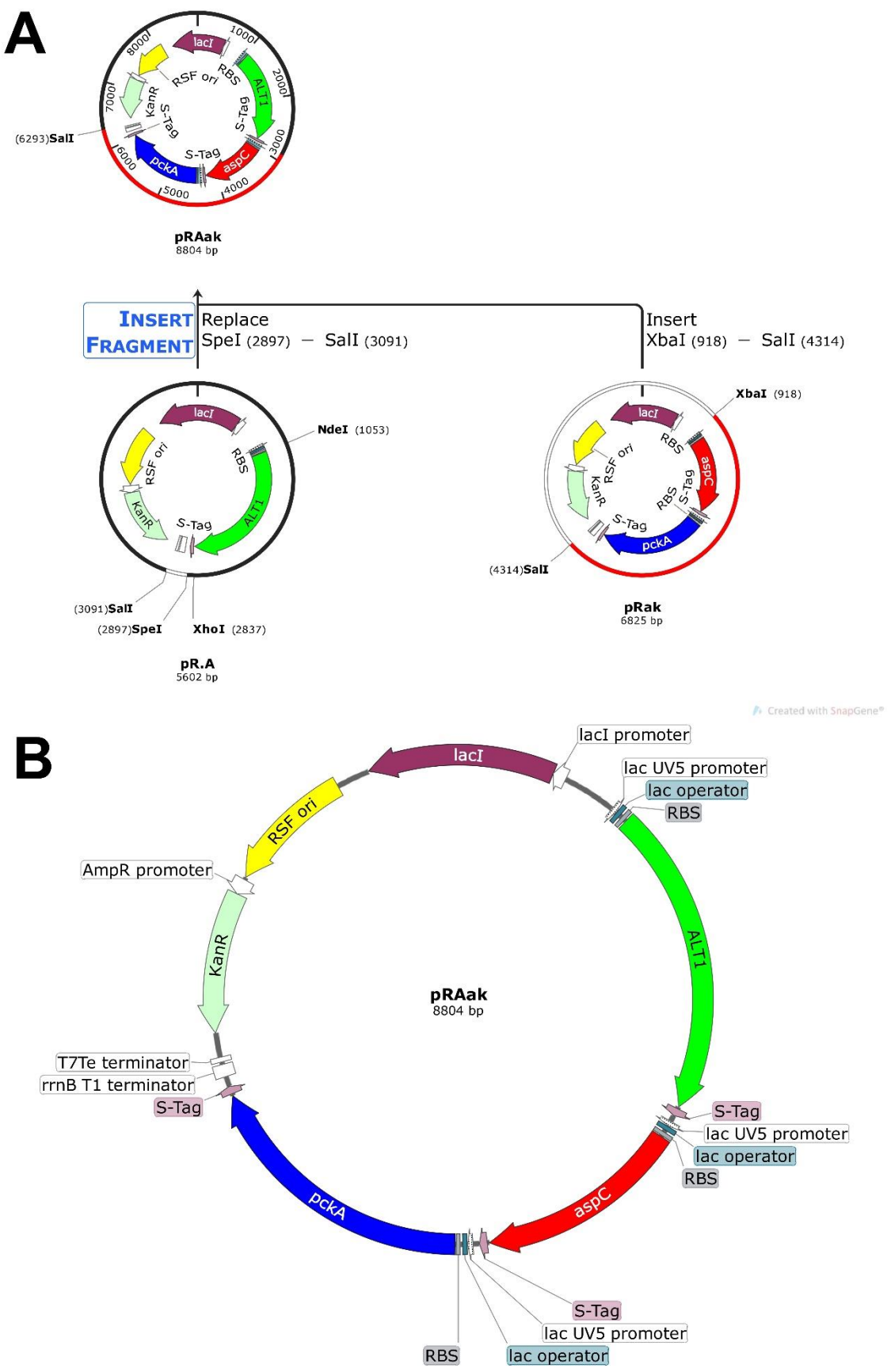


Figure C9 - pRAak plasmid cloning history (A) and map (B).

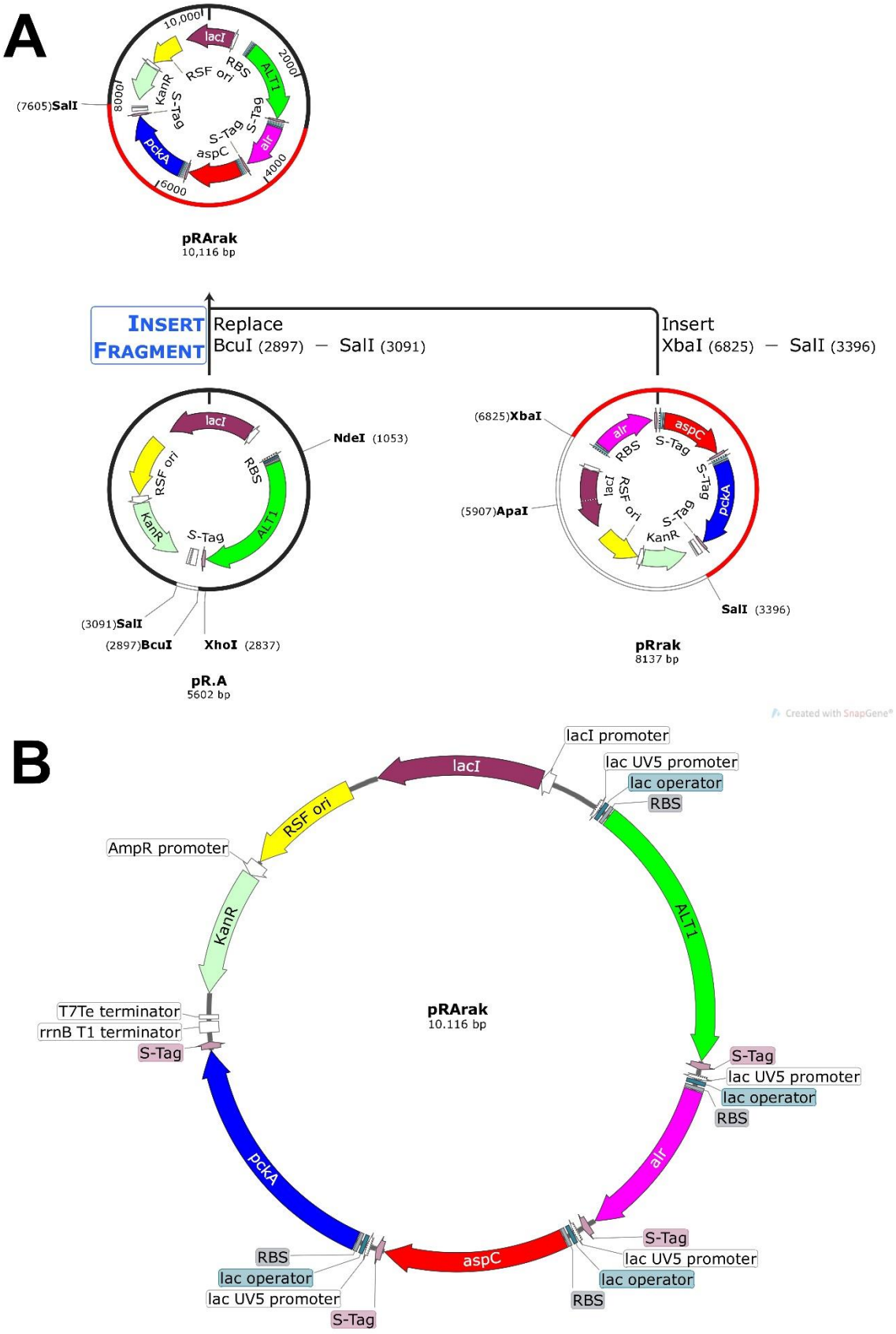
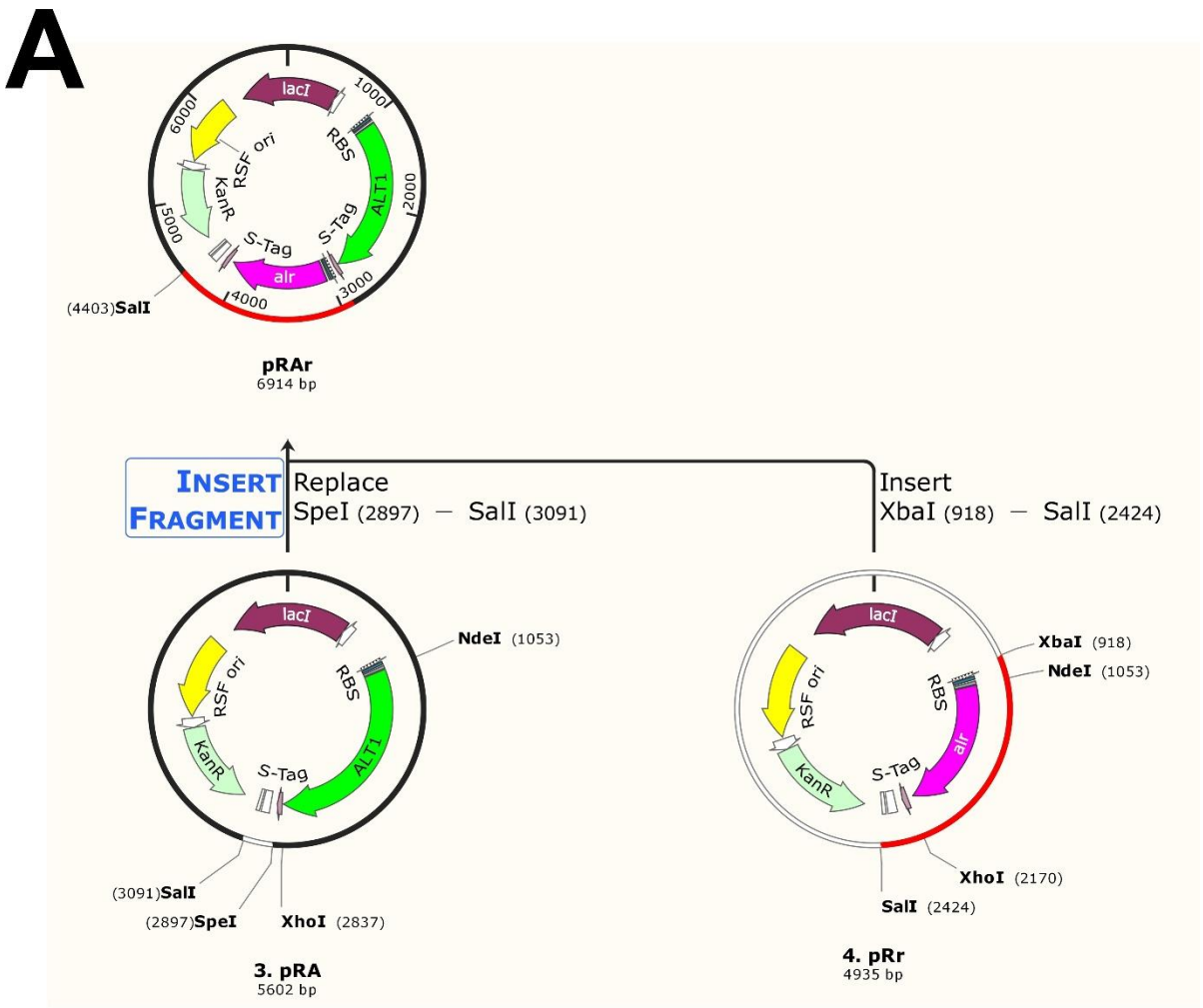


Figure C10 - pRArak plasmid cloning history (A) and map (B).



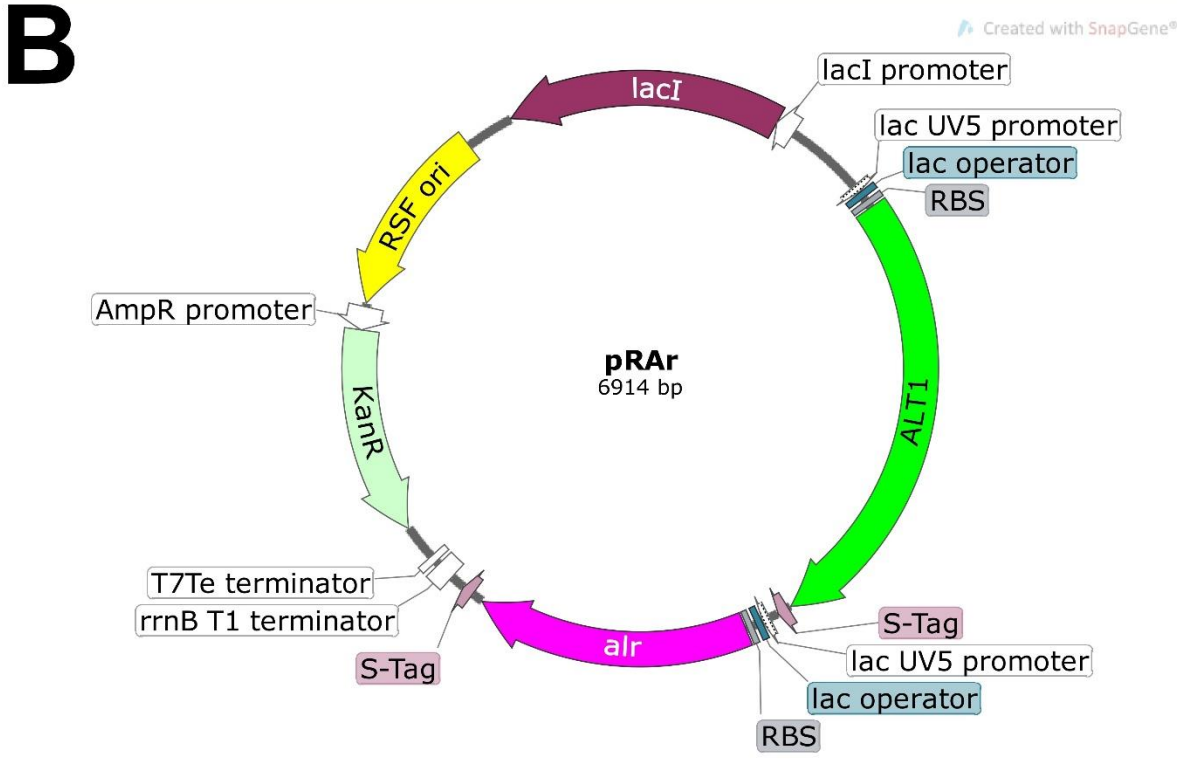


Figure C11 - pRAr plasmid cloning history (A) and map (B).

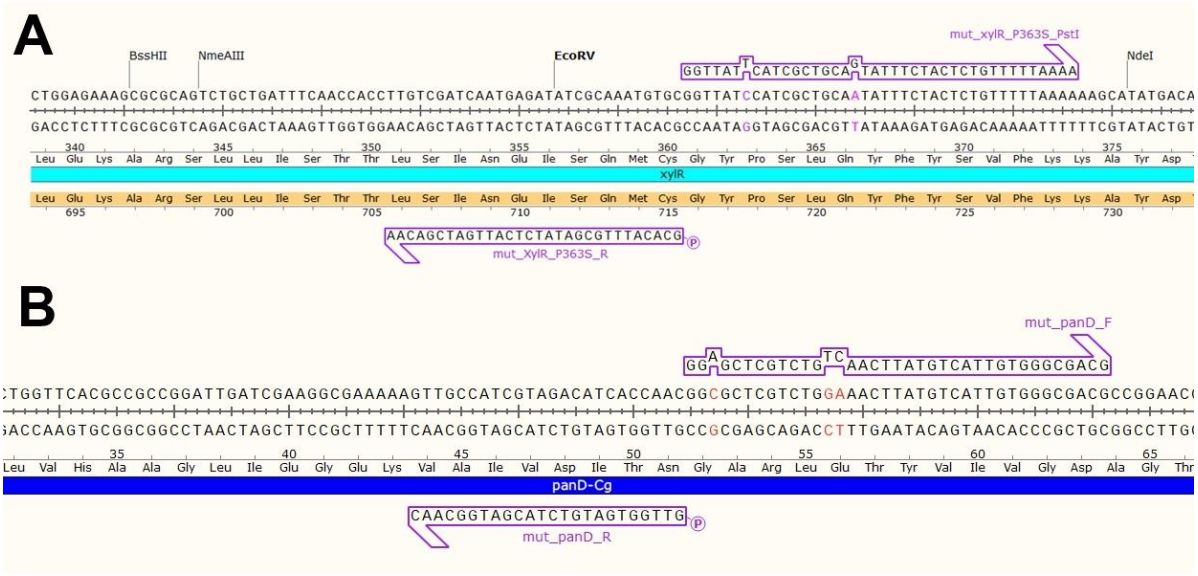


Figure C12 - Primer design for site-directed mutagenesis using round the horn PCR amplification. (A) *xylR* Pro363Ser and (B) *panD* Glu56Ser.

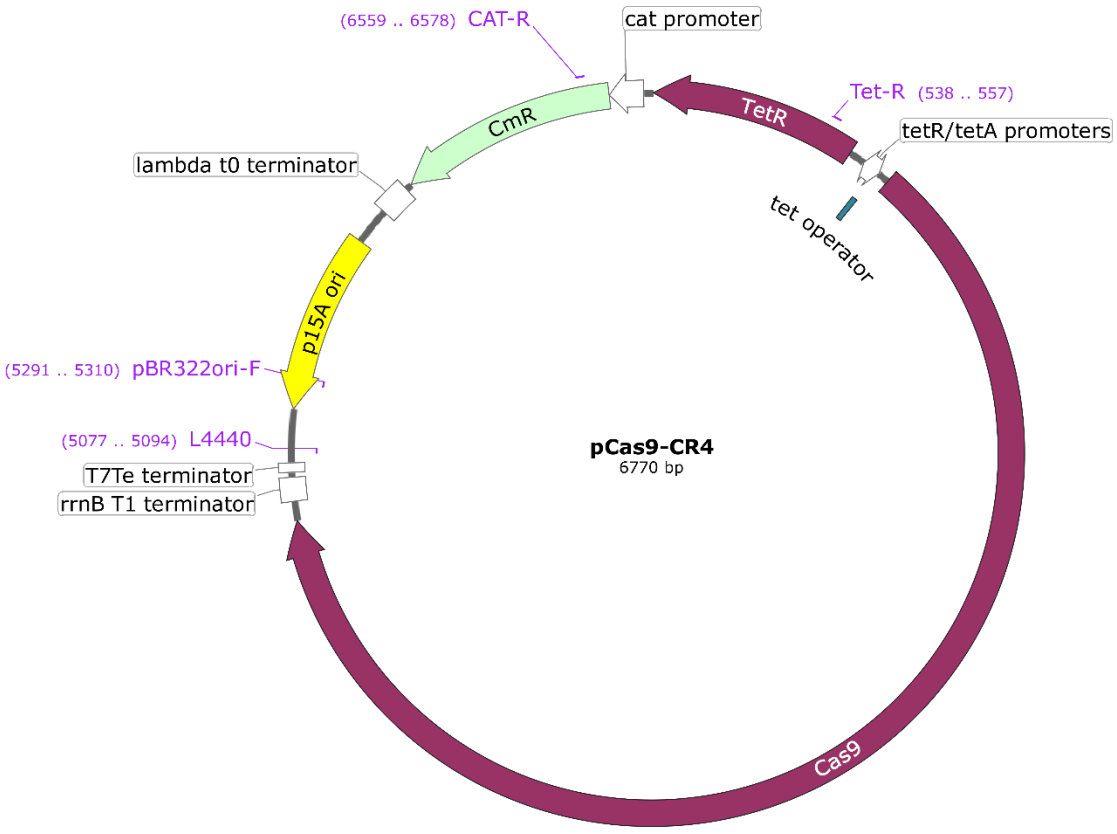


Figure C13 - pCas9cr4 plasmid map.

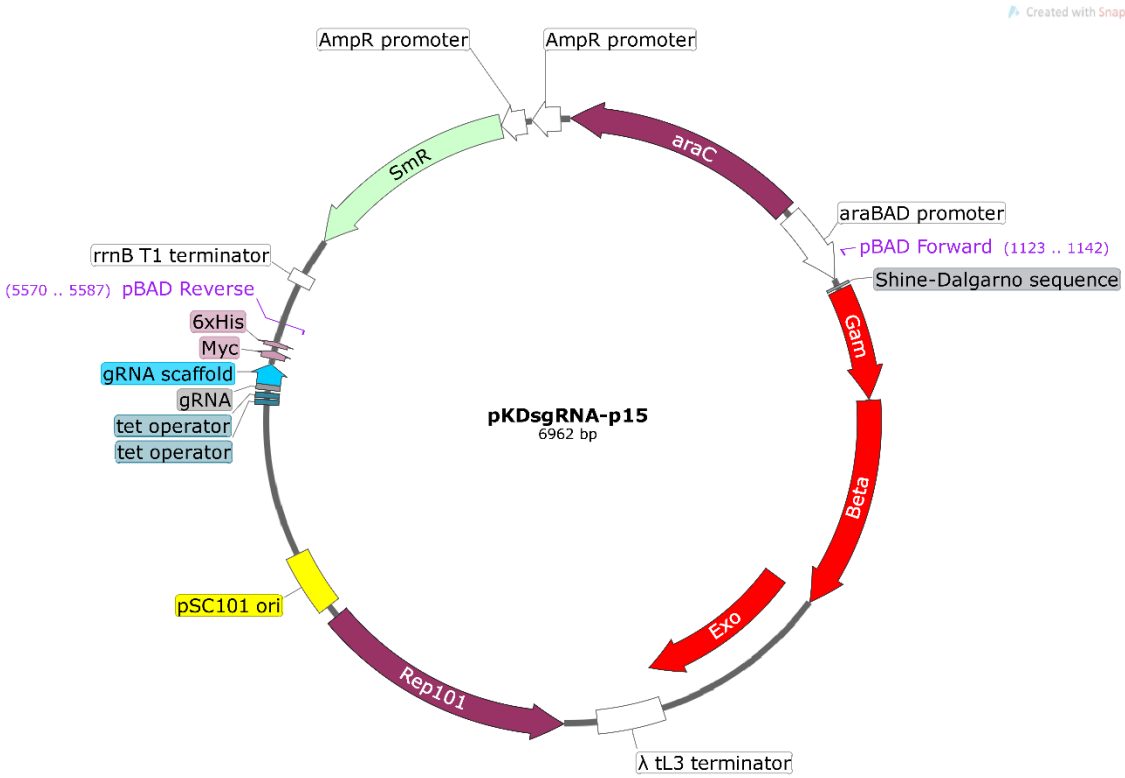


Figure C14 - pKDsgRNAp15 plasmid map. This plasmid was used as template for pKDsgRNA*xxx* cloning and pCas9cr4 cure, where *xxx* represents the target gene.

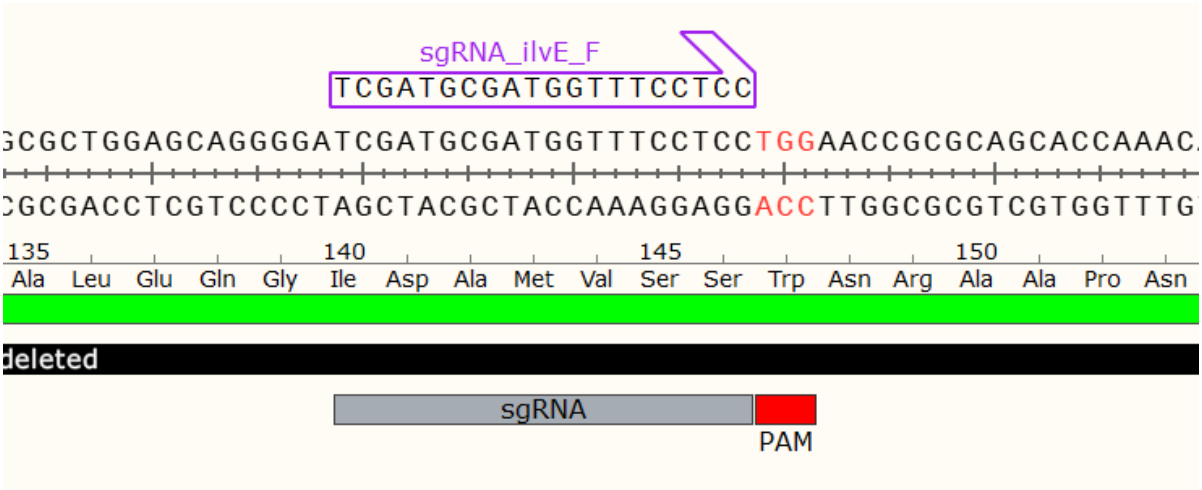


Figure C15 – Design of sgRNA_F primer for pKDsgRNAilvE cloning for *ilvE* deletion in *E. coli*.

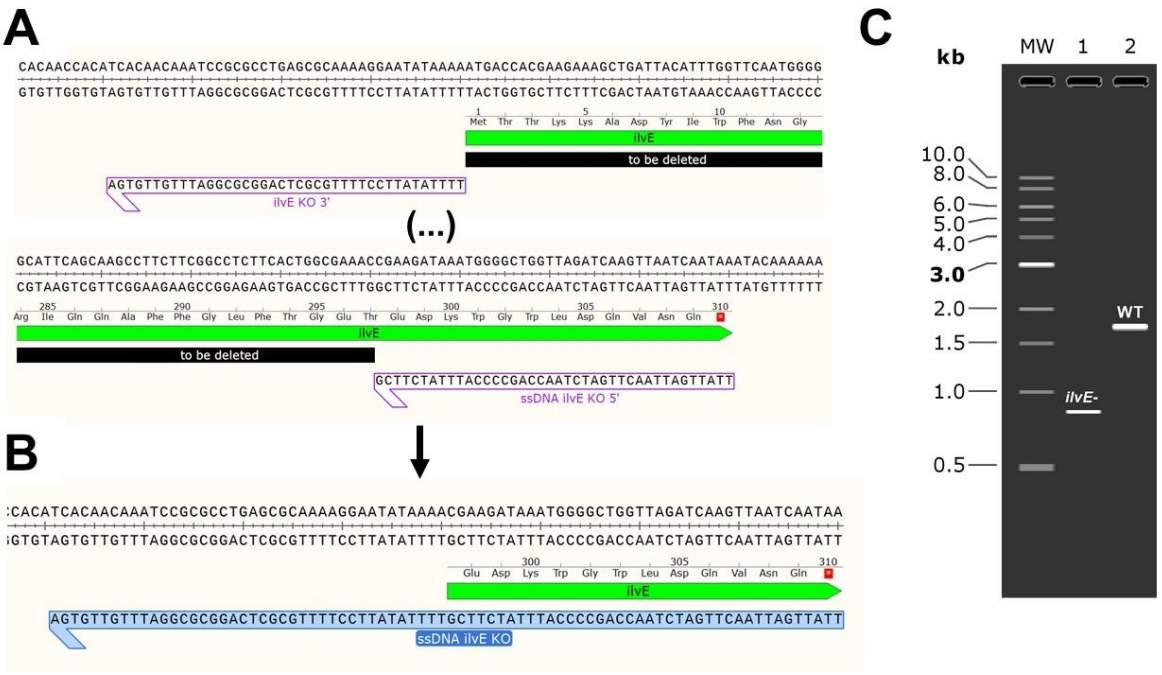


Figure C16 – ssDNA for recombineering for *ilvE* knock-out. (A) 3' and 5' homology arms, (B) ssDNA annealing to the mutant DNA, and (C) agarose gel simulation of PCR reactions of the target region in Δ*ilvE* cells (lane 1) and the wild-type strain(lane 2) with 890 bp of difference from each other.

APPENDIX D – Evaluation of the moment of induction

To evaluate the best moment for induction with IPTG, three cultivations were carried out with PS100 and, for each of these cultures, the inducer was added in different moments of cell growth: *i.* during the early exponential phase (OD_{600nm} = 0.8 – 1.0, Figure D1 - A), *ii.* during the mid-exponential phase (OD_{600nm} ~ 5, Figure D1 - B) and *iii.* during the final exponential phase (OD_{600nm} = 8.0 – 10.0, Figure D1 - C).

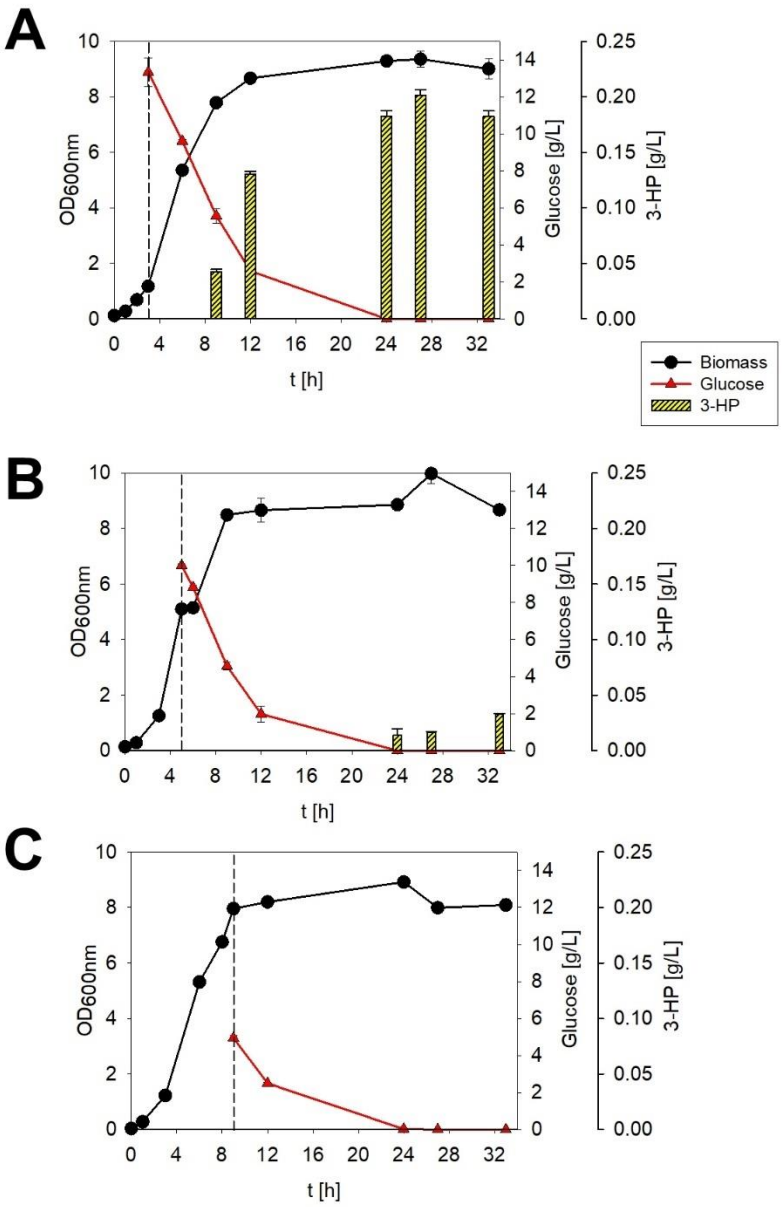


Figure D1 – Comparison of different moments of induction for PS100 growth in glucose. Induction with IPTG 1 mM (dashed vertical line). Medium values are shown and error bars represent standard deviations.

The best moment for induction was found to be in the early exponential phase. In this case, the genes carried by the plasmid are expressed during the entire exponential phase of growth, yielding higher amounts of product formation. This early expression of exogenous genes did not cause an overload for cell metabolism since the culture in the three cases had similar growth curves. The induction in the mid-exponential phase does not produce as much 3-HP as in the first case. The worst-scenario was the induction right after the shift from exponential to stationary phase, in which the PS100 cells did not produce 3-HP.