


# A REMIX IN MICROBIOLOGY

CONSTANTINOS PATINIOS



enabling microbial ester production through  
the engineering of enzymes,  
pathways and genomes

## **Propositions**

1. Microbial production of esters is currently not a viable business model, unless governmental subsidies are provided.  
(this thesis)
2. SIBR-Cas is the first-to-try genome engineering approach in non-model mesophilic prokaryotes.  
(this thesis)
3. All scientists should approach their experiments like children approach the marshmallow-spaghetti tower.
4. Citation indices should be assigned separately for reviews and experimental articles.
5. Going to Mars is simply a curiosity-based task instead of an escape route.
6. The COVID-19 pandemic was just not significant enough to stop the anti-vax movement and to halt global warming.

Propositions belonging to the thesis entitled:

A REMIX IN MICROBIOLOGY: enabling microbial ester production through the engineering of enzymes, pathways and genomes

Constantinos Patinios

Wageningen, 22<sup>nd</sup> April 2022





## **A REMIX IN MICROBIOLOGY:**

enabling microbial ester production through the  
engineering of enzymes, pathways and genomes

**Constantinos Patinios**

## **Thesis committee**

### **Promotors**

Prof. Dr John van der Oost  
Professor of Microbial Genetics  
Wageningen University & Research

Prof. Dr Ruud A. Weusthuis  
Personal chair, Bioprocess Engineering  
Wageningen University & Research

### **Co-promotor**

Dr Servé W. M. Kengen  
Assistant professor, Laboratory of Microbiology  
Wageningen University & Research

### **Other members**

Prof. Dr Eddy J. Smid – Wageningen University & Research  
Prof. Dr Yi Wang – Auburn University, USA  
Dr Jules Beekwilder – Isobionics, Geleen  
Dr Stan J. J. Brouns – Delft University of Technology

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences)



# **A REMIX IN MICROBIOLOGY:**

enabling microbial ester production through the  
engineering of enzymes, pathways and genomes

**Constantinos Patinios**

## **Thesis**

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Friday 22 April 2022

at 4 p.m. in the Aula

**Constantinos Patinios**

A REMIX IN MICROBIOLOGY: enabling microbial ester production through the engineering of enzymes, pathways and genomes, 326 pages.

PhD thesis, Wageningen University & Research, Wageningen, the Netherlands (2022)  
With references, with summary in English

ISBN 978-94-6447-079-6

DOI 10.18174/561766

Για τον παπά μου

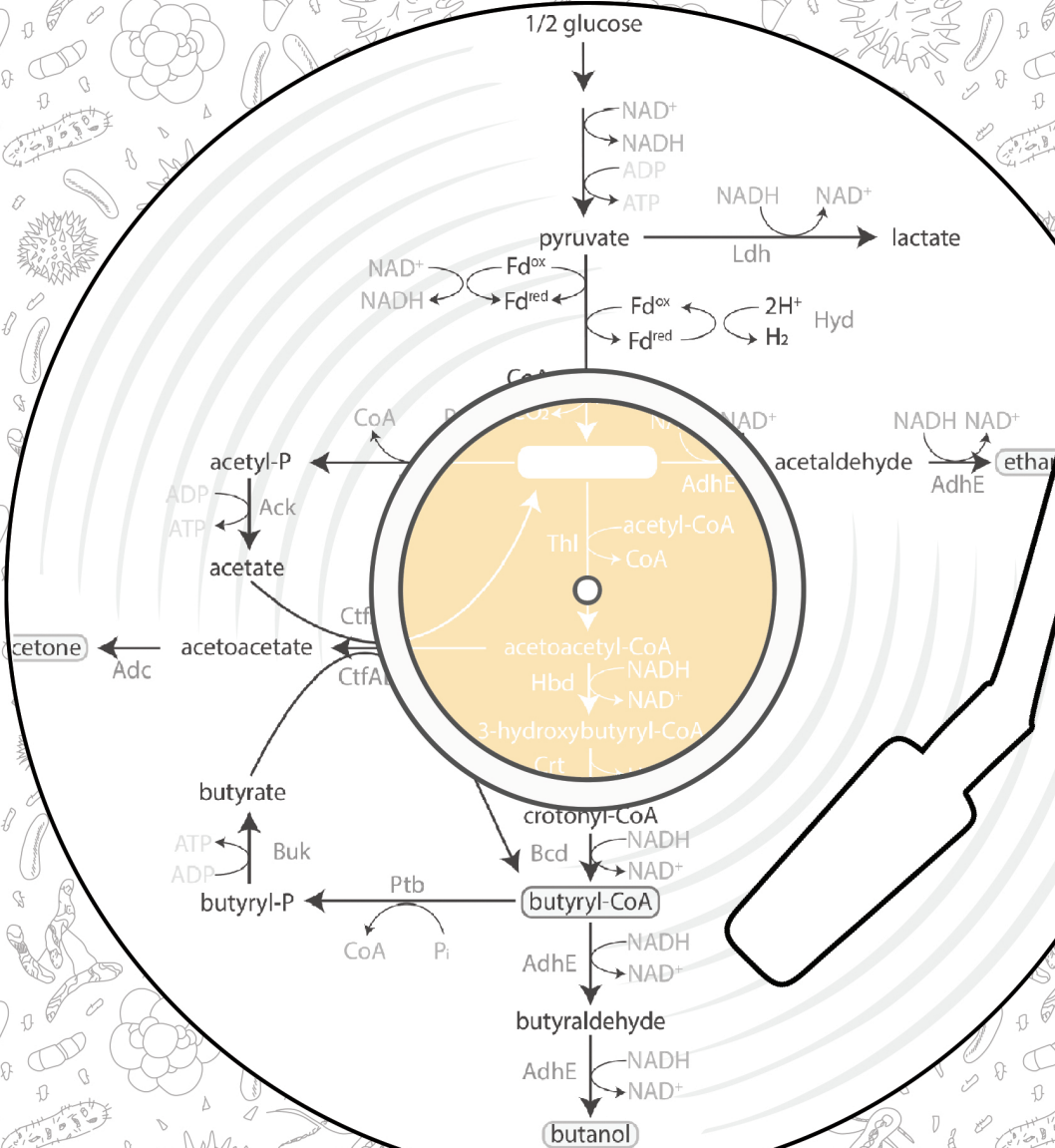




# Table of contents

<b>CHAPTER 1</b>	Introduction and thesis outline	8
<b>CHAPTER 2</b>	Microbial production of short and medium chain esters: Enzymes, pathways, and applications	28
<b>CHAPTER 3</b>	Eat1-like alcohol acyl transferases from yeasts have high alcoholysis and thiolysis activity	62
<b>CHAPTER 4</b>	Enzymes in a box: Repurposing clostridial microcompartments for ester production	90
<b>CHAPTER 5</b>	Multiplex genome engineering in <i>Clostridium beijerinckii</i> using CRISPR-Cas12a	134
<b>CHAPTER 6</b>	Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas	172
<b>CHAPTER 7</b>	Summary and general discussion	222
	Thesis summary	224
	General discussion	226
<b>REFERENCES</b>		278

# CHAPTER 1





# **Chapter 1**

## **Introduction and thesis outline**

### Prologue

The memories I have from my home place, Cyprus, are frequently associated with food. Enjoying a cold KEO beer by the beach at Konnos Bay in Ayia Napa, sipping slowly the almost 3000 year old Commandaria wine in the middle of the pine forests at the Troodos mountain, eating the warm freshly-made Halloumi and Anari (a by-product of Halloumi) cooked by my yiayia (grandma) Kallou or having the warm freshly-made sour Trahana cooked by my yiayia Christina. I can faintly taste the flavors and even smell the fragrances of the various products we have on the island, by simply closing my eyes...

All the aforementioned food and beverage (F&B) products are simply the result of a marriage made in heaven. It is the perfect match between ingredients (e.g. grapes or milk), microorganisms (yeasts or bacteria), and the cook. The ingredients (e.g. sugar) serve as substrate for the microorganisms to grow and replicate, the microorganisms use the substrate as carbon and energy source for growth but also to produce other products (e.g. alcohol, flavors and fragrances), and the cook eventually brings ingredients and microorganisms together, under the ideal conditions (e.g. temperature), to convert raw materials into, in my opinion, extraordinary products. Obviously, each one of the elements is necessary but the role and importance of microorganisms in this process is what shaped human cultures for millennia.

## Old-school versus New-school biotechnology

The production of F&B through the use of microorganisms as mediators dates back to 11,000 BC (1) and can be considered as old-school biotechnology. In old-school biotechnology, simple and largely unrefined ingredients such as milk and grape juice are used by natural (referred as wild-type in scientific terms) microorganisms to generate end products that add value or functionality to the final F&B product. Examples of end products made by microorganisms include ethanol (alcohol) or ethyl acetate (fragrance) produced by yeasts during beer and wine production, the production of lactic acid by bacteria which occurs during the making of yogurt and sauerkraut and gives the sour taste to these products, or simply the production of the CO<sub>2</sub> gas by yeasts which gives the characteristic dough rise during bread making. The number of products created through old-school biotechnology can simply not be listed as I would do injustice to several fine products which are produced in some forgotten corners of our world. What can be listed, however, are the main microbial workhorses which have been used widely for centuries (Table 1).

**Table 1. A short list of the main microbial workhorses used in food fermentation. The list is not exhaustive. The table has been adapted from (2-6).**

Group	Genera/species	Product/application(s)
Bacteria	<i>Acetobacter aceti</i> subsp. <i>aceti</i>	Vinegar
	<i>Bacillus subtilis</i>	Fermented soybeans
	<i>Bifidobacterium animalis</i> subsp. <i>lactis</i> , <i>B. breve</i>	Fermented milks with probiotic properties; common in European fermented milks
	<i>Brevibacterium flavum</i>	Malic acid, glutamic acid, lysine, monosodium glutamate (food additives)
	<i>Corynebacterium ammoniagenes</i>	Cheese ripening
	<i>Enterobacter aerogenes</i>	Bread fermentation
	<i>Enterococcus durans</i>	Cheese and sourdough fermentation
	<i>Lactobacillus acetotolerans</i>	Ricotta cheese, vegetables
	<i>L. brevis</i>	Bread fermentation; wine; dairy
	<i>L. delbruecki</i> subsp. <i>bulgaricus</i>	Yogurt and other fermented milks, mozzarella
	<i>L. kefir</i>	Fermented milk (kefir), reduction of bitter taste in citrus juice
	<i>L. kimchii</i>	Kimchi
	<i>Oenococcus oeni</i>	Malolactic fermentation of wine
	<i>Propionibacterium acidipropionici</i>	Meat fermentation and biopreservation of meat
	<i>Zymomonas mobilis</i> subsp. <i>mobilis</i>	Beverages
Yeasts	<i>Candida famata</i>	Fermentation of blue vein cheese and biopreservation of citrus; meat
	<i>C. krusei</i>	Kefir fermentation; sourdough fermentation
	<i>Kluyveromyces marxianus</i>	Cheese ripening; lactase (food additive)



## Chapter 1

	<i>Saccharomyces bayanus</i>	Kefir fermentation; juice and wine fermentation
	<i>S. cerevisiae</i>	Beer, bread
	<i>S. sake</i>	Sake fermentation
	<i>Schizosaccharomyces pombe</i>	Wine
	<i>Zygosaccharomyces rouxii</i>	Soy sauce
Filamentous moulds	<i>Aspergillus niger</i>	Beverages; industrial production of citric acid; amyloglucosidases, pectinase, cellulase, glucose oxidase, protease (food additives)
	<i>A. oryzae</i> , <i>A. sojae</i>	Soy sauce, beverages; $\alpha$ -amylases, amyloglucosidase, lipase (food additives)
	<i>Penicillium camemberti</i>	White mold cheeses (camembert type)
	<i>P. roqueforti</i>	Blue mold cheeses
	<i>Rhizopus oligosporus</i>	Tempe fermentation
	<i>R. oryzae</i>	Soy sauce, koji

Although old-school biotechnology has been exploiting microorganisms for millennia, the existence, functionality and importance of microorganisms in F&B was not realized until the 1660s-1670s. Who was the first to conceptualize the existence of microorganisms or the first to observe microorganisms is not a clear-cut story (7) but, frequently, literature refers to Antonie van Leeuwenhoek as “the Father of Microbiology” who discovered the “microcosmos” through his microscopic lenses around 1670. Later, around 1837, Charles Cagniard de la Tour, Theodor Schwann and Friedrich Traugott Kützing, independently described that the “metabolic activity of yeast cells is the cause of fermentation, and demonstrated that sugar and wort is used for growth and multiplication of the yeast” (8). Following, in the mid-1850s, Louis Pasteur further described the presence and importance of bacteria in milk products and developed the well-known process of pasteurization (a process where packaged and non-packaged food is heated to around 50-60°C for a short time to kill most microorganisms and extend the shelf life of a product). In 1897 a real turning point occurred in the world of microbiology, when Eduard Buchner demonstrated that yeast extract (with no living yeast) can form alcohol from a sugar solution, which led to the conclusion that fermentation is the result of enzymes (proteins capable in catalyzing a chemical conversion) produced by the microorganisms present in the F&B production process (9). Eduard Buchner’s observations marked the birth of biochemistry as a discipline and in 1907 Eduard Buchner was awarded the Nobel prize in chemistry.

Advances in microbiology, biochemistry and fermentation technologies in the 19<sup>th</sup> century led to a plethora of novel applications and the generation of new products. This marked the birth of new-school biotechnology. New-school biotechnology is essentially the controlled and

conscious exploitation of microorganisms for biotechnological applications. New-school biotechnology goes beyond the scope of using wild-type microorganisms for simple fermentation purposes with F&B as the final products. It simply aims at using (micro)organisms to convert anything (available substrates) to everything (desired products), with the main goal to use sustainable production processes (i.e., avoiding the use of petrochemical alternatives). New-school biotechnology essentially dissects old and novel (micro)organisms to their molecular level with the aim to use or re-purpose the (micro)organism towards the efficient generation of the product of interest. This includes (i) genomic analysis of the genome (DNA) of the microorganism to identify and express the genes that encode proteins/enzymes that are responsible for converting substrates to certain compounds, (ii) metabolomic analysis to characterize end products and, if necessary, (iii) the genetic manipulation of the microorganism to improve the fermentation process. Following this approach, new-school biotechnology (often referred to as industrial biotechnology) focuses on the large-scale production of pure chemicals used as biofuels, commodity chemicals, fine chemicals, pharmaceutical chemicals and even proteins and enzymes for food, commodity, cosmetic and pharmaceutical applications (Table 2). With the rapid advancements in the field of biotechnology (including microbiology, genetic engineering, process engineering), it is indubitably a very exciting time to study, develop, apply and even invest in biotechnology.

**Table 2. Short list of commercially available products through fermentation biotechnology. The companies and products are not exhaustive. Most of the information in this table is acquired from online sources or from a recent review (10).**

Chemical/protein name	Characteristic; Use	Production company
<b>Flavors and fragrances</b>		
Valencene	Citrus flavor; beverages and fragrances	Evolva, Isobionics
Nootkatone	Grapefruit flavor; beverages and fragrances	Evolva, Isobionics
Santalol	Woody odor; fragrance	Isobionics, Amyris
$\beta$ -elemene	Herbal; flavor and fragrances	Isobionics
$\beta$ -bisabolene	Green, flowery, citrus like odor; beverage industry, food, fragrances and in personal and skin care products	Isobionics
$\delta$ -cadinene	Floral, light orange tones, pine cone; flavor and fragrances	Isobionics
$\delta$ -germacrene	Floral, soapy and green odor; beverages, food and personal and skin care products	Isobionics
Farnesene	Herbal, fresh green, sweet odor; beverages and food	Isobionics
Trans- $\alpha$ -bergamotene	Citruy, fresh, earthy green odor; beverages, food and personal and skin care products	Isobionics

## Chapter 1

Vanillin	Vanilla flavor; flavors and fragrances	Evolve, IFF, Solvay
Patchouli	Fragrance	Amyris, Firmenich
$\gamma$ -decalone	Peach flavor; beverages, personal care products and pharmaceutical applications	Conagen
Sclareol	Fragrance material	Amyris, Firmenich
Manool	Woody, amber notes; fragrance	Amyris
<b>Proteins/Enzymes</b>		
Amylase AD11MDP	Starch digestion	Biocatalysts
Glucose oxidase 789L	Oxidase	Biocatalysts
CalB Immo Plus	Lipase B	c-Lecta
ATA-01 <i>Aspergillus fumigatus</i>	$\omega$ -transaminase	Enzymicals
IREDO4 <i>Paenibacillus elgii</i> B69	Imine reductase	Enzymicals
Esterase 01 <i>Bacillus subtilis</i>	Esterase	Enzymicals
ADH 030, lyophilized powder, NAD-dependent enzyme	(S)-alcohol dehydrogenase	Evovx Technologies
Addzyme TL 165G, immobilized <i>Thermomyces lanuginosus</i>	Lipase	Evovx Technologies
ADH A <i>Pyrococcus furiosus</i>	Alcohol dehydrogenase	GECCO
<i>TfuDyP Thermobifida fusca</i>	Dye-decolorizing peroxidase	GECCO
MetZyme® Brila™ Enzyme product family	Recycle fiber applications	Metgen Oy
MetZyme® Suno™ Enzyme product family	Lignocellulosic chemicals production	Metgen Oy
Alcalase® 2.4LFG from <i>Bacillus licheniformis</i>	Endo-peptidase	Novozymes
rTrypsin® 8.0L from <i>Fusarium venenatum</i>	Serine protease	Novozymes
Cephalosporins, Penicillins, Fluoroquinolones, Macrolides, Carbapenems	Antibiotics	Pfizer, Merck & Co, Johnson & Johnson, Novartis, GlaxoSmithKline, Bayer
<b>Bioplastics</b>		
Polyethylenefuranoate (PEF)	Bioplastics	Corbion, Synvina
Polylactic acid (PLA)	Bioplastics	Total-Corbion
Polypropylene (PP)	Bioplastics	LanzaTech
Polyhydroxyalkanoates (PHAs)	Bioplastics	Danimer Scientific, Yield 10 Nioscience, Bio-on, Tepha, Biocycle, Biomer
<b>Other</b>		
Cannabinoids	Therapeutic	Hyasynth Bio, Librede, Demetrix, Biomedican, Amyris, Cronos in collaboration with Ginkgo Bioworks, InMed Pharmaceuticals, Organigram
Opioids	Therapeutic	CB therapeutics, Octarine Bio and PsyBio
Reb M	Sweetener	Amyris
Squalene	Active component in an important class of vaccine adjuvants	Amyris
Squalane	Emollient; skincare, sun care, color cosmetics, makeup removal, and deodorants	Amyris
Artemisinin	Molecule used as an antimalarial	Amyris

trans- $\beta$ -farnesene	Scaffold for specialty chemical applications	Amyris
Bisabolol	Skin healing anti-microbial, anti-inflammatory, and anti-irritant properties; personal care	Amyris
Hemisqualane	Conditioners, treatments, and styling products	Amyris
L-ergothioneine	Antioxidant properties; support healthy aging and mitigate the chronic inflammation caused by oxidative stress	Conagen
Butadiene	Tires, polymers, latex	Genomatica
hexamethylenediamine, caprolactam and adipic acid	Carpet, clothing, fibers, engineering plastics directly used for nylon 6, nylon 6,6 and polyurethanes	Genomatica
1,3-butylene glycol	Cosmetics, personal care	Genomatica
1,4-butanediol	Athletic apparel, running shoes, electronics and automotive uses	Genomatica
Butylene glycol	Humectant, preservative booster, solubilizer or stabilizer; personal care applications	Genomatica
Resveratrol	Dietary and health supplement	Evolva
L-Arabinose	Sugar blocker; food and beverage	Evolva
Silk	Materials and clothing	AMSilk, Spiber Inc, Bolt Threads
Citric acid	Preservative; food and beverage	Citrique Belge
Lactic acid	Preservative; food and beverage	Corbion
Succinic acid	Solvents and lubricants, food, pharmaceuticals, deicer solutions, cosmetics	BioAmber, Succinity, Reverdia
Vitamin B2	Dietary and health supplement	DSM
Vitamin B12	Dietary and health supplement	DSM
Arachidonic acid	Dietary and health supplement	DSM
Beta-carotene	Colorant	DSM
Natamycin	Antifungal preservative	DSM
Aspartame	Sweetener	DSM
Betalain	Color; Food and beverage, cosmetic	Phytolon
Ethanol	Fuel	Lanzatech

## Industrial biotechnology: Some striking examples

Although not all biotechnology-based ideas are commercialized, several companies have done this successfully for a wide range of products and have now established the strength and potential of biotechnology as a sustainable and profitable alternative to petrochemical processes (Table 2). Several market or product “trends” are identified and are discussed below.

Current “trends” are obviously surrounding the flavor and fragrance compounds. Such ingredients are in high demand mainly because of the enormous size of the F&B industry (\$36.3

## Chapter 1

bn by 2028 with a compound annual growth rate of 4.9%) (11). Also, due to their volatile nature, flavors and fragrances can be captured and purified efficiently, hence decreasing their purification costs and increasing their purity levels. Interestingly, microbially-produced flavors and fragrances can “by-pass” any strict genetically modified organism (GMO) related regulation which is attributed again to the volatile nature of fragrance compounds and the absence of GMO material during their purification. Important players in this market sector are mentioned in Table 2.

Plastic, although a wondrous material, is highly pollutant with annual estimates of 9 to 23 million metric tons of plastic waste into lakes, rivers and oceans, and 13 to 25 million metric tons into terrestrial environment (12,13). The estimated emissions are expected to double by 2025, if we do not change our production and disposal approaches (14). An alternative is already developed by biotech companies such as Corbion, Synvina, Total, LanzaTech, Danimer Scientific, Yield 10 Nioscience, Bio-on, Tephra, Biocycle and Biomer (amongst others; Table 2). Such companies produce bio-polymers such as polyethylenefuranoate (PEF), polylactic acid (PLA), polypropylene (PP) and polyhydroxyalkanoates (PHAs) that are used for the production of materials with plastic-like qualities. Obviously, the advantage of bioplastics is their biodegradation and their sustainable production through biotech-based methods. Nevertheless, since bioplastics differ in composition, waste management should be a priority to avoid potential leakage into ecosystems (15).

The production of “non-regulated” cannabinoids for use as therapeutics is also trending at the moment, with companies such as Hyasynth Bio, Librede, Demetrix, Biomedican, Amyris, Cronos in collaboration with Ginkgo Bioworks, InMed Pharmaceuticals or Organigram (the list is not exhaustive) focusing on the production of tetrahydrocannabinol (THC), cannabidiol (CBD), cannabigerol (CBG), cannabinol (CBN) and/or tetrahydrocannabivarin (THCV) (Table 2). Also, another big trend is the production of the more potent psychedelic compounds such as the opioids psilocybin and psilocin by CB therapeutics, Octarine Bio and PsyBio (the list is not exhaustive; Table 2). The success of these technologies will depend on the undergoing clinical trials for the potency of these compounds as therapeutics.

Color production is also a hot area as several start-up companies including Pili Bio, Phytolon, Chromologics, MiChroma and Colorifix are using natural or synthetic microorganisms to produce colors destined for the F&B, the cosmetic and textile industries (Table 2). So far, none of these companies has released a product on the market but this may be only a matter of time.

Legislation issues may arise for this sector of products, especially if the color compounds are novel (in chemical structure and in quantity used in a certain product) to the industry. Also, since the color compounds are not volatile, purification of these products from GM organisms requires approval as safe and non-GMO.

Microbial production of human milk oligosaccharides (HMOs) is an upcoming trend in the industry as HMOs were proven to have a positive effect on allergies, infection and immunity (16). Glycom, Friesland Campina, Jennewein Biotechnologie GmbH (acquired by Chr. Hansen) and BASF are some of the companies which are commercially producing HMOs through the fermentation of engineered *Escherichia coli* cells (Table 2) (17).

More onto the material side of industrial biotechnology, silks are gaining popularity due to their impressive mechanical properties and potent applications in sportswear and military equipment (e.g. protective helmets and vests). Companies involved in the production of silk using microorganisms as their production platform are AMSilk (*E. coli*), Spiber Inc (unknown host) and Bolt Threads (yeast) (Table 2) (18). Since silk molecules are very long chains of amino acids, optimal protein production is required for the success of this product.

## Which microorganism and why?

Selecting the “perfect” microorganism to produce the product-of-interest is a critical aspect in biotechnological processes. The choice should be based on the chemical and physical properties of the product of interest and through a “start-at-the-end” approach (Fig. 1). The end of a biotechnological process (referred to as downstream processing) is the purification of the final product. If the product is, for example, a volatile compound, downstream processing is relatively simple since volatile compounds can be recovered from the headspace of the fermentor. However, the volatility of different compounds differs and is based on their chemical and physical properties (19). Therefore, to increase volatility and consequently increase the recovery of the final product, a higher temperature during the fermentation process would be an advantage. Obviously, this would require a microorganism that can survive and grow at high temperatures (thermophile). Another crucial criterion is that the production organism is amenable to genetic modification, especially if the expression of exogenous genes is required for the production of the compound of interest. For this, defined growth conditions, transformation protocols, replicative plasmids and genome engineering tools should be available/known or they should be developed. Next, the microorganism of choice should,

## Chapter 1

preferably, produce large amounts of precursor molecules that can be channeled to the product of interest without affecting the growth of the microorganism. Also, the final product should not be toxic to the microorganism as this will severely affect the growth and hence the final production levels of the product of interest. Lastly, the chosen microorganism should be able to utilize and grow on the carbon source of interest (e.g. sugars, gasses, methanol or formate).

Unfortunately, there is not a single “universal” prokaryotic or eukaryotic (micro)organism that is suited to produce all kinds of products. This is mainly due to genetic and metabolic differences between microorganisms which have evolved throughout billions of years to thrive in a specific ecological niche. For example, some microorganisms are able to grow and replicate by metabolizing sugars (e.g. glucose, xylose or rhamnose) whereas some others are able to grow by metabolizing C1 molecules (methane, CH<sub>4</sub>; methanol, CH<sub>3</sub>OH; formate HCOO<sup>-</sup>; carbon monoxide, CO; or carbon dioxide, CO<sub>2</sub>). Some microorganisms are able to grow optimally at extremely low temperatures (ranging from -20°C to 10°C; called psychrophiles), some at moderate temperatures (20°C to 45°C; called mesophiles) and some at very high temperatures (60 to 80°C; called thermophiles or 80 to 122°C; called hyperthermophiles). Also, some are able to grow only in aerobic (presence of O<sub>2</sub>; called obligate aerobes) conditions, some grow only anaerobically (absence of O<sub>2</sub>; called obligate anaerobes) and some can grow in both aerobic and anaerobic conditions. Moreover, the quantity and type of precursor compounds and final products produced by different microorganisms also differs, making the choice of the “perfect” microorganism a complex quest.

Despite the lack of a “universal” production organism, a few microorganisms have been used extensively in biotechnology as the main workhorses. This type of microorganisms is referred to as model microorganisms since extensive knowledge is acquired regarding their genome, their metabolism and growth conditions, and tools are available for their genetic manipulation. The most studied prokaryotic example is the model gut bacterium *Escherichia coli* whereas the most applied eukaryotic example is the model yeast *Saccharomyces cerevisiae* (also referred to as Baker’s yeast). Both microorganisms have been used extensively for fundamental and applied research, and humanity owes a lot to the scientists who contributed to the characterization and exploitation of these microorganisms.

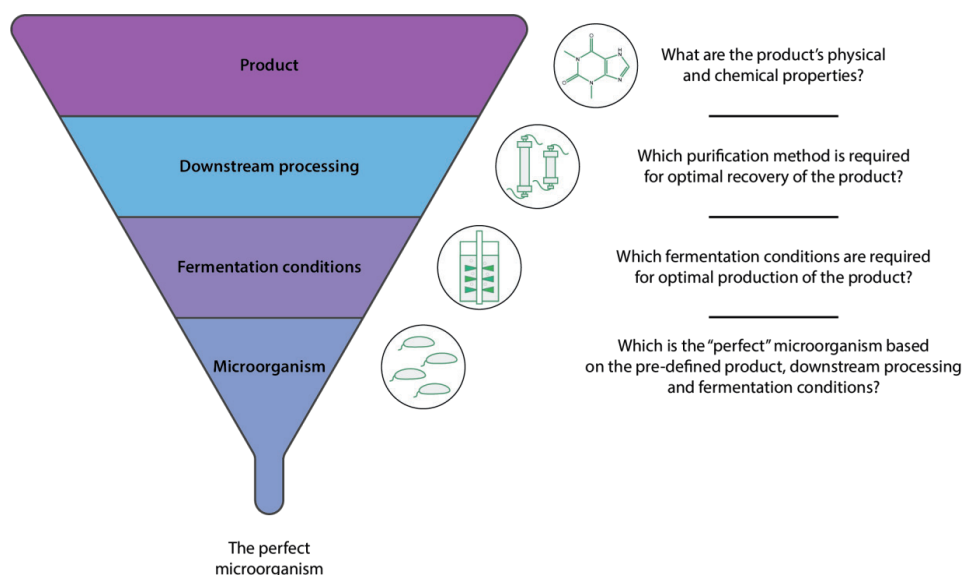


Figure 1. Start-at-the-end approach for selecting the “perfect” microorganism for the product of interest.

## Clostridia and esters

The bacterial species within the genus *Clostridium* include several important human pathogens (e.g. *Clostridium botulinum*, *C. perfringens*, *C. tetani*) but also several workhorses of biotechnology (e.g. *C. acetobutylicum*, *C. beijerinckii*, *C. thermocellum*, *C. saccharoperbutylacetonicum*). *Clostridium* species were used extensively in the twentieth century because of their ability to produce acetone, butanol and ethanol (ABE). Their importance is highlighted during World War I since *Clostridium* species were used to produce cordite/gunpowder, lacquer solvents and jet fuel (20-23). As a result of the advancements in petrochemical processes in the mid-twentieth century, the production of ABE-related products using *Clostridium* species was outcompeted by petrochemically-based products. However, due to economic and environmental reasons and due to advances in biotechnology, sustainable ABE fermentation (i.e. use of renewable sources) using *Clostridium* species has re-emerged (24).

The basic ABE metabolism of *Clostridium* species is illustrated in Figure 2. Acetone, butanol and ethanol are terminal products with acetate and butyrate being important intermediates. Other important intermediates include acyl-CoAs (e.g. acetyl-CoA and butyryl-CoA) which are the main substrates for the formation of acids and ABE. Due to the high production of ABE



(>100 mM) and acids (~20 mM) researchers figured out that the combination of the two in the form of an ester (acid joint to an alcohol) may increase the value of the final product (25-32).

Esters are versatile compounds that are frequently used as an additive or as a building block in the food and chemical industries. Their production is currently based on the traditional and well-established Fischer-Speier esterification (33), although alternative and more sustainable methods have been developed using microbial cell factories (34). The production of esters in *Clostridium* species has been demonstrated either through the expression of exogenous alcohol acyl transferases (AATs; to combine endogenously produced acyl-CoAs and alcohols) or through the supplementation of lipases (to combine endogenously produced acids and alcohols) (25-31). A recent landmark study paved the path towards the efficient production of butyl esters (butyl acetate and butyl butyrate) in *Clostridium saccharoperbutylacetonicum* N1-4 by testing several AATs and by using multiple metabolic engineering approaches (32).

In the course of this PhD thesis, we focused on a newly discovered AAT, called Eat1, and the use of *Clostridium beijerinckii* NCIMB 8052 as a host organism. Eat1 was discovered in our lab by Alex Kruis *et al.* (2017) and was demonstrated to be the main AAT enzyme for the production of ethyl acetate in yeasts (35). Eat1 belongs to the  $\alpha/\beta$ -hydrolase family and was shown to have AAT, esterase and thioesterase activity. Due to its major role for the production of ethyl acetate and other esters in yeasts (35,36), we hypothesized that Eat1 would be the ideal enzyme to express in *Clostridium beijerinckii* NCIMB 8052 to produce a variety of esters including ethyl acetate, ethyl butyrate, butyl acetate and butyl butyrate by combining the endogenously produced alcohols (ethanol, butanol) with the endogenously produced acyl-CoAs (acetyl-CoA, butyryl-CoA) (Fig. 3). The produced esters can then be used as commodity chemicals, flavors and fragrances or even as biofuels (19,37).

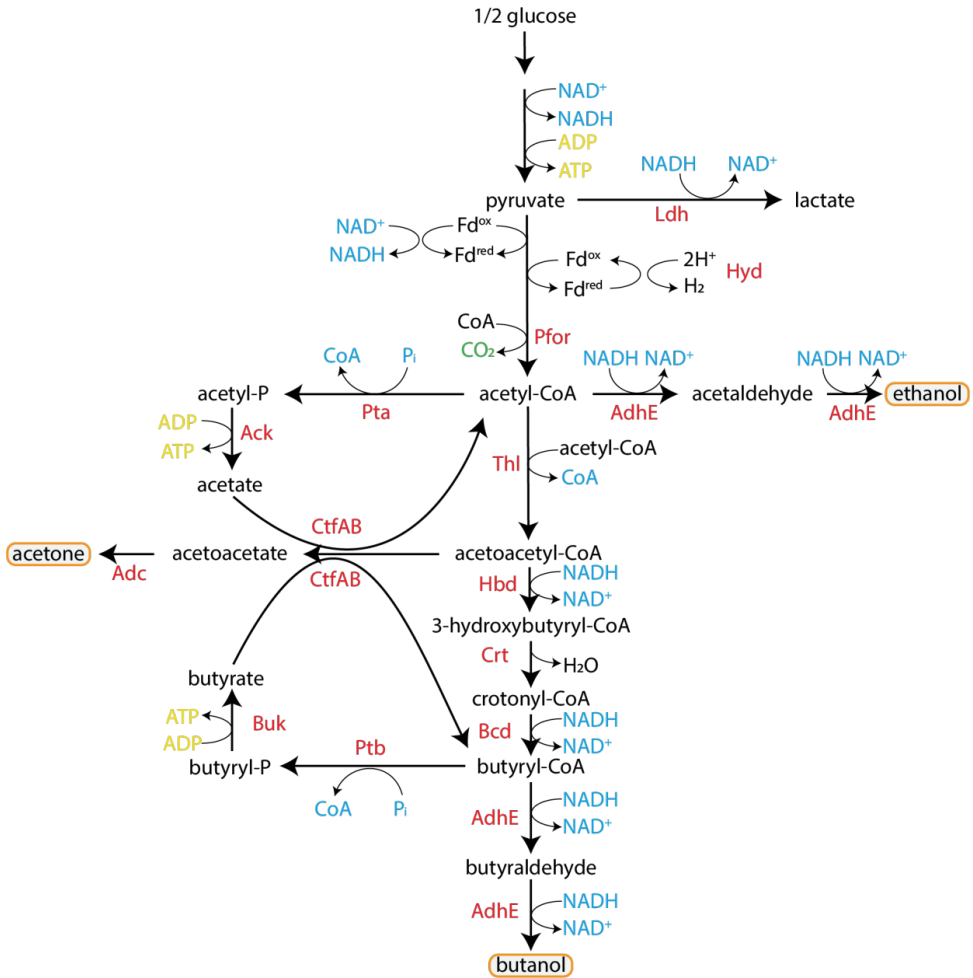


Figure 2. ABE metabolism in *C. beijerinckii* NCIMB 8052 when glucose is the main carbon source.

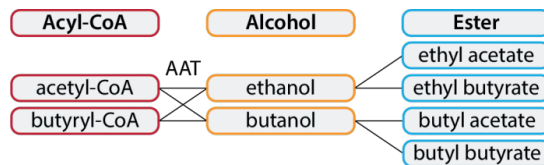


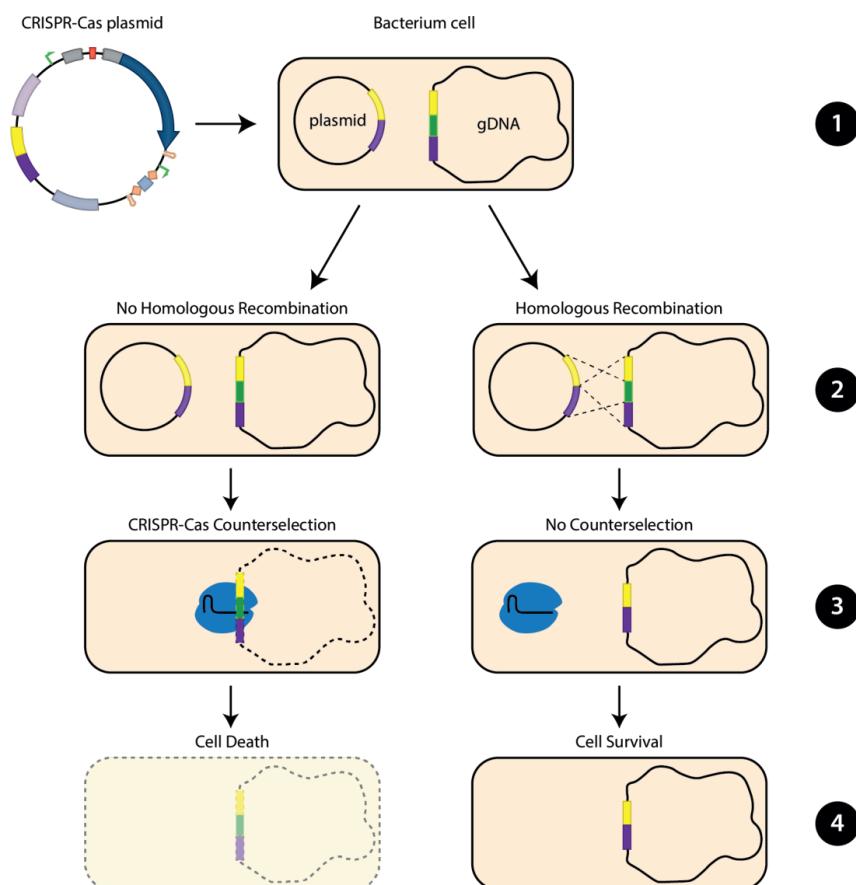
Figure 3. Potential ester products by using an AAT (alcohol acyl transferase) in *Clostridium beijerinckii* NCIMB 8052.

## CRISPR-Cas as a powerful tool for genome engineering in bacteria

Without a doubt, it is an exciting era to be a genome engineer. Genome engineering has never been simpler, faster and more accurate and all of this is attributed to the discovery and development of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated proteins (CRISPR-Cas) (38-47). CRISPR-Cas has revolutionized the field of biology and has accelerated scientific and industrial advances in an unprecedented speed. This is reflected by the quantity of CRISPR-Cas-associated peer-reviewed scientific papers which ranges at ~100 papers per week (Note: the number of weekly papers was derived by using the pubcrawler service and by using “CRISPR-Cas” as a search term for titles and abstracts) and also by the award of the Nobel prize in chemistry (2020) to Emmanuelle Charpentier (Max Planck Unit for the Science of Pathogens) and Jennifer Doudna (University of California, Berkeley) for their discoveries and advances in CRISPR-Cas technologies. Important to mention is the immense number of patents (~23.000-30.000) that contain at least the word “CRISPR” in their title or description text (Note: the number of patents was retrieved by searching for “CRISPR” at the google patents and Espacenet patent search platforms).

With such a rapidly developed and applied technology, it is becoming hard to keep up with the CRISPR-Cas advancements. Novel proteins and novel applications are being discovered and developed constantly. Applications of CRISPR-Cas range from genome editing (gene knock-out, gene knock-in, and precision editing including base editing and prime editing), gene silencing and gene activation to uses such as molecular diagnostics. Thankfully, great reviews are summarizing the CRISPR-Cas advances and applications and are listed here (48-60).

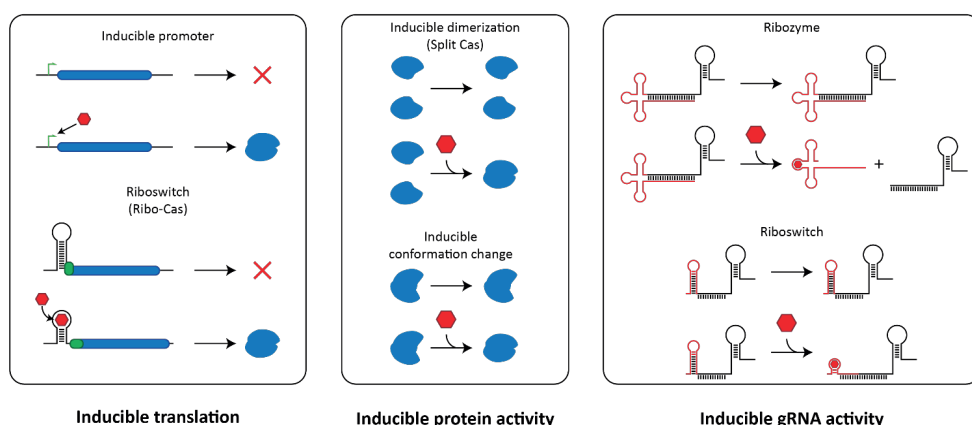
Important to microbiologists is the repurposing of CRISPR-Cas as a genome engineering tool for bacteria. The first report to use CRISPR-Cas to engineer the genome of bacteria was in 2013 when Jiang and colleagues used CRISPR-Cas9 to edit the genome of *Streptococcus pneumoniae* and *Escherichia coli* (61). The success of this tool is based on homologous recombination (HR) combined with the elegant counter-selection mechanism of CRISPR-Cas that kills unedited cells whilst allowing for the survival of edited cells (Fig. 4). HR combined with CRISPR-Cas counter-selection has been applied successfully to many different bacterial species (48,51,53,55). However, it often remains cumbersome to use CRISPR-Cas genome editing in newly characterized, non-model, bacterial species.



**Figure 4. Principle of CRISPR-Cas counterselection for efficient selection of edited cells.** 1) A plasmid DNA bearing the CRISPR-Cas system (CRISPR array and Cas protein) and homologous arms to knock-out the gene of interest, is transformed to the bacterium of interest. 2) Bacterial cells will either recombine to edit the target site (right path; Homologous Recombination) or they will remain unedited and retain their wild-type genome (left path; No Homologous Recombination). 3) Expression of CRISPR-Cas will target and cut the unedited wild-type genome (left path) whereas it will not target the edited genome (right path). 4) Unedited cells that are targeted and cut with CRISPR-Cas will die (left path) whereas the edited cells will survive. gDNA: genomic DNA.

Based on Figure 4, for high editing efficiency, HR should precede CRISPR-Cas counterselection. Unfortunately, this is not the case in many prokaryotes (including non-model organisms) where HR efficiency is typically low. Low HR efficiency combined with the strong counter-selective pressure of CRISPR-Cas leads to cell death and therefore unsuccessful genome editing. Two approaches can be taken to overcome this drawback: (1) Increase the HR

efficiency and/or (2) delay the expression and activity of CRISPR-Cas, granting more time for HR to occur before counterselection by CRISPR-Cas. Whilst the increase of HR has been applied successfully through the heterologous expression of phage recombinases in model organisms such as *E. coli* and *P. putida* (62-64), it is often laborious (maintenance of multiple plasmids) and success is not guaranteed as the recombinases may be incompatible with the target organism. The second alternative (delay the expression and activity of CRISPR-Cas), is a widely used approach and has been used successfully for genome engineering. Examples include inducible translation of the Cas protein, inducible Cas protein activity and inducible gRNA activity (Fig. 5) (65-81).



**Figure 5. Mechanisms to delay the expression and activity of CRISPR-Cas.** Inducible translation controlled by inducible promoters or riboswitches, inducible protein activity through protein dimerization or conformational change and inducible gRNA activity through ribozymes and riboswitches are shown as examples. The mechanisms illustrated in this figure are not exhaustive.

Typically, the existing approaches to control the expression and activity of CRISPR-Cas are organism-, promoter-, Cas protein- or gRNA-specific. To overcome this limitation and be able to expand the application of CRISPR-Cas genome engineering to virtually any bacterium of interest, a control method that can be applied to a wide range of bacterial species is required. As described in this thesis (see below), we developed such an approach (named SIBR-Cas) which can be applied virtually to any mesophilic bacterium of interest. The wide applicability of SIBR-Cas and the simplicity of our system is expected to be the “first-to-try” (and maybe only) approach for genome engineering in mesophilic prokaryotes.

## Thesis outline

**Chapter 1** provides a general overview of microbiology and biotechnology and the combination of the two for the production of human destined products. It is clear that microbiology has been a major player, consciously or unconsciously, that contributed to shaping human cultures. After spectacular developments over the past decades, the gained insights are currently used to support human cultures through biotechnological approaches. The state-of-the-art in biotechnologically produced products is reviewed and a short guideline is provided on how to pick the “perfect” microorganism for the product of interest. The revolutionary CRISPR-Cas technology is also introduced and the importance of CRISPR-Cas for microbiologists is highlighted.

### **Chapter 2 | Microbial production of short and medium chain esters: Enzymes, pathways, and applications**

The second chapter reviews the latest developments for the microbial production of esters. The underlying ester-forming enzymatic mechanisms are discussed and compared, with particular focus on alcohol acyltransferases (AATs). The review also focusses on the chemical and physical properties of short and medium chain esters, as these may be used to simplify downstream processing, while limiting the effects of product toxicity. Finally, the perspectives and major challenges of microorganism-derived ester synthesis are presented.

### **Chapter 3 | Eat1-like alcohol acyl transferases from yeasts have high alcoholysis and thiolysis activity**

The third chapter covers the activity portfolio of the newly discovered Eat1 enzyme. The alcoholysis and thiolysis activities are added to the activity portfolio of Eat1 to complement the already known alcohol acyl transferase (AAT), esterase and thioesterase activities. We discovered that the *in vitro* alcoholysis activity is 445 times higher than the previously described alcohol acyl transferase activity which most likely affects the net ester generation, both quantitatively and qualitatively, in food and beverage production processes.

### **Chapter 4 | Enzymes in a box: Repurposing clostridial microcompartments for ester production**

The fourth chapter uses an alternative approach to bring enzymes and substrates in close proximity in a confined space. To do this, we used the bacterial microcompartments (BMCs)

## Chapter 1

of *C. beijerinckii* NCIMB 8052 and attempted to re-purpose them to produce propyl propionate as a proof of principle. Different AATs were targeted for encapsulation in the lumen of BMCs by using N-terminal encapsulation peptides (EPs). Unfortunately, during the time course of this thesis, we did not achieve our target. However, we show for the first time the encapsulation of GFP in the BMCs of *C. beijerinckii* NCIMB 8052.

## Chapter 5 | Multiplex genome engineering in *Clostridium beijerinckii* using CRISPR-Cas12a

In the fifth chapter we expanded the CRISPR-Cas toolbox in *Clostridium beijerinckii* NCIMB 8052 and developed a CRISPR-FnCas12a genome engineering tool. We achieved single (25-100% knock-out efficiency) and multiplex (18% knock-out efficiency) gene knock-outs demonstrating the functionality of our tool.

## Chapter 6 | Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas

The sixth chapter describes the development of a widely applicable CRISPR-Cas tool for bacteria. The tool is named SIBR-Cas (Self-splicing Intron-Based Riboswitch-Cas) and is generated from a mutant library of the theophylline-dependent self-splicing T4 *td* intron that allows for tight and inducible control over CRISPR-Cas counter-selection. SIBR-Cas was successfully applied to knock-out several genes in three wild-type bacteria (*Escherichia coli* MG1655, *Pseudomonas putida* KT2440 and *Flavobacterium* IR1). Furthermore, we propose that SIBR can have a wider application as a simple gene expression and gene regulation control mechanism for any gene or RNA of interest in bacteria. We named that tool SIBR-X.

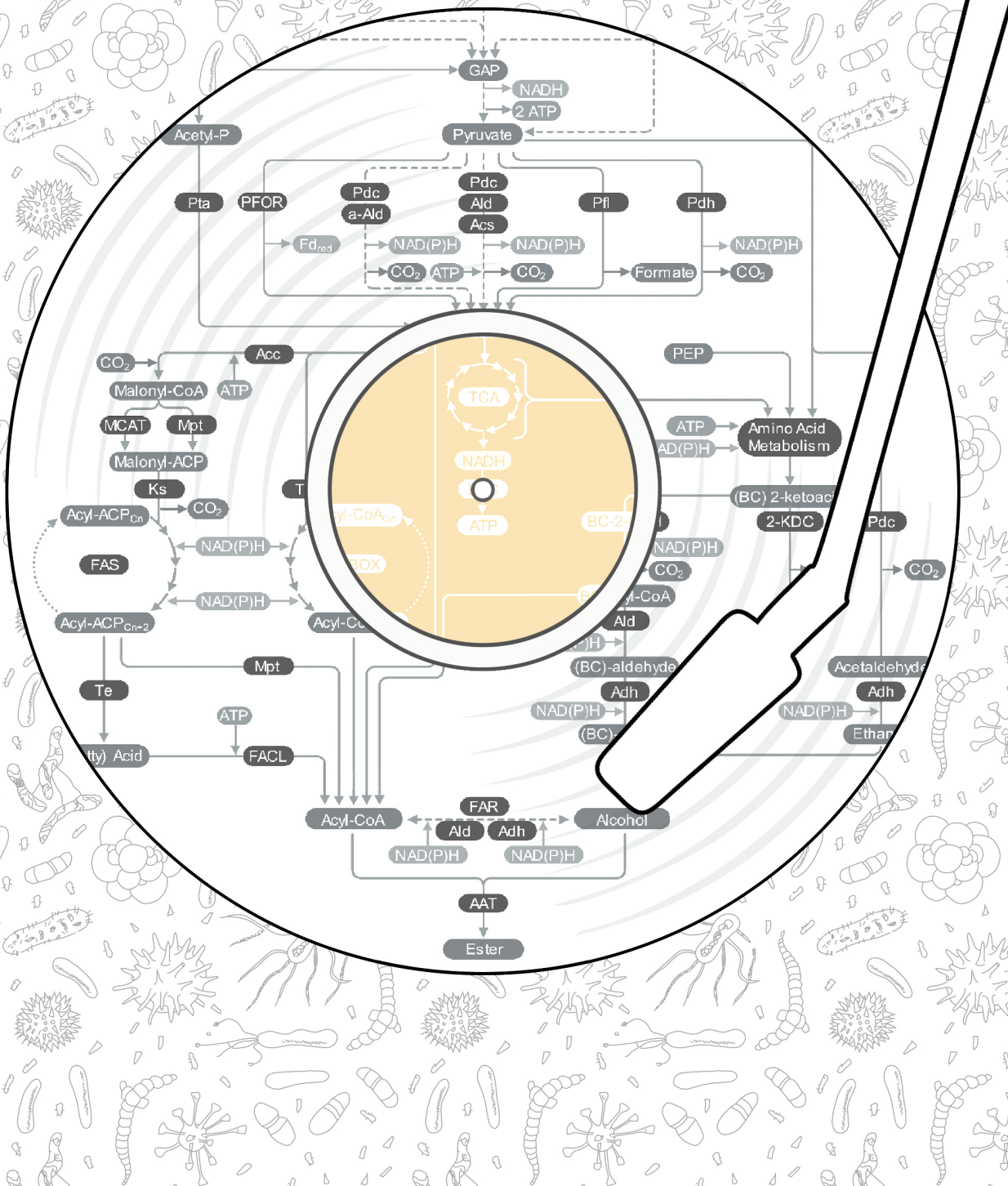
## Chapter 7 | Summary and general discussion

The last chapter gives a summary of the work performed in this thesis. It also presents experimental outcomes obtained in the course of this PhD project, that are not reported in the preceding chapters, and discusses the outcome of those experiments. In addition, outstanding practical and philosophical questions are addressed leaving an open end for the curious mind.





# CHAPTER 2



## Chapter 2

# Microbial production of short and medium chain esters: Enzymes, pathways, and applications

Aleksander J. Kruis<sup>a,b,1</sup>, Anna C. Bohnenkamp<sup>a,b,1</sup>, Constantinos Patinios<sup>b,a,1</sup>, Youri M. van Nuland<sup>a</sup>, Mark Levisson<sup>c</sup>, Astrid E. Mars<sup>d</sup>, Corjan van den Berg<sup>a</sup>, Servé W.M. Kengen<sup>b</sup>, Ruud A. Weusthuis<sup>a\*</sup>

<sup>a</sup>Bioprocess Engineering, Wageningen University & Research, the Netherlands

<sup>b</sup>Laboratory of Microbiology, Wageningen University & Research, the Netherlands

<sup>c</sup>Laboratory of Plant Physiology, Wageningen University & Research, the Netherlands

<sup>d</sup>Wageningen Food & Biobased Research, Wageningen University & Research, the Netherlands

<sup>1</sup>These authors contributed equally to the work

\*Correspondence: Prof. Ruud A. Weusthuis (ruud.weusthuis@wur.nl)

Chapter adapted from publication:

Biotechnology Advances; doi.org/10.1016/j.biotechadv.2019.06.006

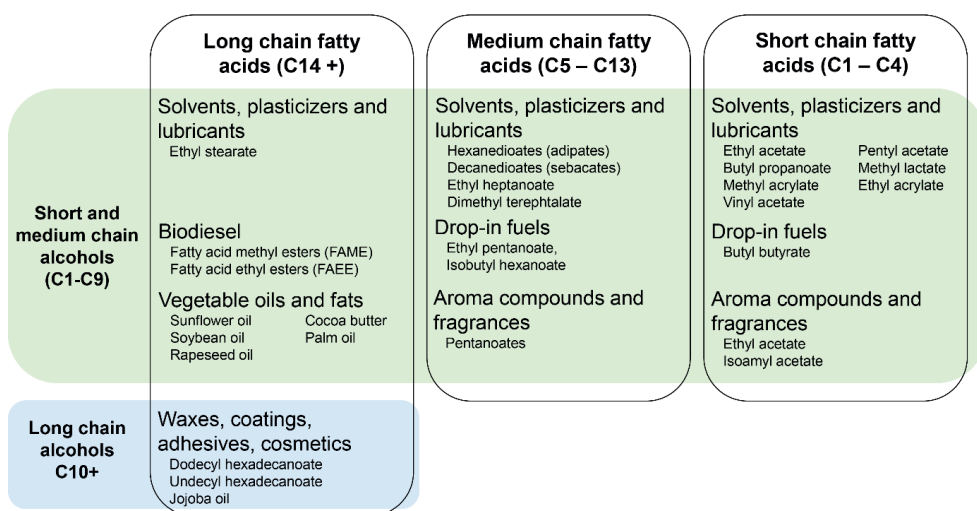
### Abstract

Sustainable production of bulk chemicals is one of the major challenges in the chemical industry, particularly due to their low market prices. This includes short and medium chain esters, which are used in a wide range of applications, for example fragrance compounds, solvents, lubricants or biofuels. However, these esters are produced mainly through unsustainable, energy intensive processes. Microbial conversion of biomass-derived sugars into esters may provide a sustainable alternative. This review provides a broad overview of natural ester production by microorganisms. The underlying ester-forming enzymatic mechanisms are discussed and compared, with particular focus on alcohol acyltransferases (AATs). This large and versatile group of enzymes condense an alcohol and an acyl-CoA to form esters. Natural production of esters typically cannot compete with existing petrochemical processes. Much effort has therefore been invested in improving *in vivo* ester production through metabolic engineering. Identification of suitable AATs and efficient alcohol and acyl-CoA supply are critical to the success of such strategies and are reviewed in detail. The review also focusses on the physical properties of short and medium chain esters, which may simplify downstream processing, while limiting the effects of product toxicity. Furthermore, the esters could serve as intermediates for the synthesis of other compounds, such as alcohols, acids or diols. Finally, the perspectives and major challenges of microorganism-derived ester synthesis are presented.

**Keywords:** ester; metabolic engineering; alcohol acyltransferase; alcohol; carboxylic acid; bulk chemical; microbial synthesis.

# 1 Introduction

Carboxylate esters are versatile compounds that find various applications in the food and chemical industry. They are naturally produced by yeasts such as *Saccharomyces cerevisiae* and define the taste and odour of fermented beverages (82). As natural food additives they are used to enhance the flavour and odour profile of various food products (83). Ethyl acetate, isoamyl acetate or propyl acetate are common fragrance and aroma compounds, that also find application as bulk chemicals (Fig. 1) (84,85). Various esters are used as industrial solvents due to their biodegradability and low toxicity, or as plasticizers and polymer additives (86-88). Further, they find applications as lubricants, coatings and are explored for their potential as drop-in fuels or biodiesels (Fig. 1) (89-91).



**Figure 1. Applications of various esters as bulk chemicals or high value products.** Due to their versatility, esters may be found in various groups of applications and are not limited to those mentioned here.

Traditional ester production processes make use of the Fischer-Speier esterification (33). Alcohols and carboxylic acids, produced from fossil resources, are condensed in the presence of an acid catalyst at elevated temperatures (92). Water is released in the process, which leads to the formation of the desired ester. The process however, is dictated by an equilibrium that reduces the reaction rate with time and prevents a complete conversion of all acid and alcohol substrates. In addition, water has been found to inhibit the catalytic activity of the acid catalyst (93). To allow the reaction to proceed, the water is continuously removed from the system in energy intense distillation or adsorption steps. While transesterification of vegetable oils is

## Chapter 2

evaluated for the production of biodiesel, this process has only limited applicability (94,95). For other esters esterification of biobased alcohols and acids might be considered, but as equilibrium reactions remain part of the process other solutions should be evaluated.

Microbial conversion systems will be key in developing efficient ester-producing bioprocesses and have already received much attention for sustainable bulk chemical synthesis (96-99). Their vast array of enzymes is able to perform hundreds of chemical conversions at ambient conditions. To catalyse multiple metabolic pathways simultaneously, enzymes have evolved high specificity for their substrates, and the formation of by-products in industrial microbial bioprocesses is thereby limited (100).

Bulk chemicals, such as esters used as solvents, polymers, drop-in fuels or lubricants, have low market prices and low profit margins compared to biopharmaceutical products or industrial enzymes (101,102). To become economically competitive with their petrochemical counterparts, the biobased conversion processes must achieve high product titres, yields and volumetric productivities, and should be followed by efficient downstream processes. This is a major challenge in current biotechnology that has prevented many products from moving to large-scale production (103).

Due to their broad application range, especially short chain esters such as acetate esters, lactate esters, butyrate esters or acrylate esters would benefit from sustainable production processes. Companies such as Corbion or Vertec BioSolvents are among the first to launch green lactate ester products for coatings, paints and cleaning (104,105). Many commercially interesting esters are important building blocks for various polymer structures, such as terephthalates. However, many are currently produced by chemical conversion processes and to date, no biobased approaches have been described. Such esters are therefore not further considered in this review (106,107).

This review article focuses on how and why microorganisms synthesize esters, particularly those that can be applied as commodity chemicals. We provide a detailed overview of the enzymatic reactions that produce esters in microorganisms and how these enzymes have been applied to improve ester formation. An overview of the metabolic engineering strategies aimed at increasing ester production is provided and notable examples are highlighted. We also consider the physical and chemical properties of esters and how they might benefit the

production of other valuable bulk chemicals, such as alcohols, carboxylic acids, and  $\alpha,\omega$ -diols. Lastly, the major challenges that lie ahead of biobased ester production are summarised.

## 2 Microbial ester production

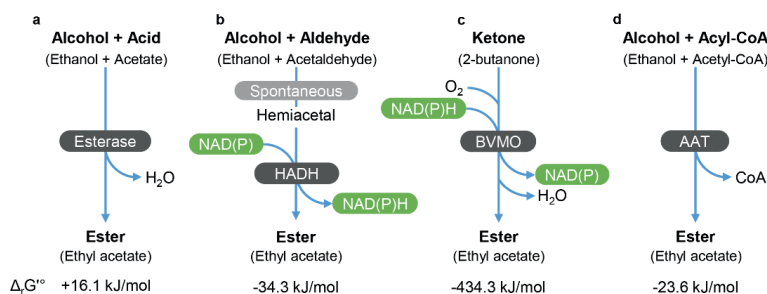
Natural ester production by microorganisms, such as yeasts and lactic acid bacteria is well established and has historically been applied in food production. Volatile esters are among the most important aroma compounds in fermented foods, such as beer, wine and dairy products. In low concentrations, esters impart a sweet, fruity aroma, but are also considered as off-flavours when present in high amounts (108). Ethyl acetate is the most abundant volatile ester in food. Its concentration ranges from ~50 to 100 mg/L in dairy products and from ~8 to 64 mg/L in beer and wine (82,108). Other volatile esters, such as isoamyl acetate, phenylethyl acetate, ethyl hexanoate and many others do not exceed concentrations of 1 mg/L. These concentrations lie around or just above the human detection threshold (82,109) and therefore greatly affect the aroma of food products. The amounts of esters naturally produced by microorganisms are typically low, although some exceptions exist, such as bulk ethyl acetate-producing yeast or wax-ester accumulation by *Euglena gracilis*. This microalga can accumulate wax esters to as much as 65% of the cell dry weight under anaerobic conditions (110).

The ability of certain yeasts to produce high amounts of ethyl acetate was observed more than 120 years ago (111). Yeasts such as *Kluyveromyces marxianus*, *Wickerhamomyces anomalus* and *Cyberlindnera jadinii* are able to synthesise ethyl acetate from sugars or ethanol (88,112-116). Growth under iron-limited conditions is the main trigger for bulk ethyl acetate production in yeasts (112,117,118). In some yeasts, ethyl acetate formation has also been induced by oxygen limitation (35,119). Ethyl acetate production in *K. marxianus* has been investigated in most detail. Several strains have been identified that form ethyl acetate from whey sugars (120,121), glucose (122), and cassava bagasse supplemented with glucose (123). *K. marxianus* is able to catabolise lactose and can utilise whey, a side stream of the cheese industry, to produce ethyl acetate. Moreover, this yeast is able to grow at elevated temperatures, which facilitates ethyl acetate removal from the fermentation broth. These traits make *K. marxianus* an attractive cell factory for the production of biobased ethyl acetate (88,121,124,125). In one study, lactose was converted to ethyl acetate by *K. marxianus* at 42 °C. The yield reached 56.2 % of the pathway maximum, which is the highest reported yield for a natural ethyl acetate-producer reported to date (126).

Because of the structural and chemical diversity of esters, no singular physiological role can be defined for ester synthesis and some are even still debated. Ethyl acetate is a major fermentation product of certain yeast species and contributes to balancing their central carbon metabolism under sub-optimal growth conditions (88,117,127). High concentrations of ethyl acetate also repress growth of competitive organisms (128,129). Volatile esters serve as metabolic intermediates during growth on alkanes or cyclic alcohols in some bacterial species (130,131). Some esters, such as isoamyl acetate, may help yeast to disperse in the environment by attracting insects (132). Some specialised esters can even act as bacterial virulence factors (133). Wax esters are produced as intracellular storage compounds in *Acinetobacter baylyi* and *Marinobacter hydrocarbonoclasticus* (134,135), or as an anaerobic fermentation product in *Euglena gracilis* (136).

## 2.1 Ester-forming reactions

Four main enzymatic ways of ester biosynthesis have been described, using i) esterases (Fig. 2a), ii) hemiacetal dehydrogenation (HADH, Fig. 2b), iii) Baeyer-Villiger monooxygenases (BVMOs, Fig. 1c) and iv) alcohol acyltransferases (AATs, Fig. 2d). The reactions catalysed by AATs and esterases are redox neutral, whereas, BVMOs and HADH require respectively NAD(P)H or NAD(P) (Fig. 2).



**Figure 2. Enzymatic reactions that result in ester production.** The  $\Delta_r G^\circ$  of the reactions were estimated using eQuilibrator (137) under standard conditions and pH 7. Abbreviations: AAT – alcohol acyltransferase, HADH – hemiacetal dehydrogenation, BVMO – Baeyer-Villiger monooxygenase.

Other ester-forming enzyme classes exist, but have not been applied extensively for ester synthesis. These include S-adenosyl methionine (SAM) dependent O-methyltransferases and polyketide synthase associated proteins. SAM dependent O-methyltransferases transfer the methyl group from SAM to free fatty acids to form FAME (138). Polyketide associated proteins are better known for their role in the synthesis of secondary metabolites. One such protein,

PapA5 is able to transfer alcohols to growing polyketide chains, forming complex esters with interesting biological activities in the process (139,140).

### 2.1.1 Esterases

Esterases, including lipases, are ubiquitous enzymes that have been identified in all domains of life. In aqueous environments, they catalyse the hydrolysis of ester bonds, resulting in the formation of an alcohol and a carboxylic acid. Water activity, pH and substrate concentrations play a critical role in esterase reactions and special consideration should be given to the thermodynamics of the reaction. Under aqueous conditions, the reverse esterase reaction has a positive  $\Delta_r G^\circ$ , making the reaction thermodynamically unfavourable (Fig. 2a). Hence, industrial production of esters *via* esterases is typically performed in non-aqueous systems using organic solvents or high substrate concentrations (141,142). Nonetheless, several studies have reported *in vivo* reverse esterase activity as the cause of ester formation by microorganisms (143-148). In *Acetobacter pasteurianus*, disruption of the esterase-encoding gene *estI* eliminated ethyl acetate and isoamyl acetate production (147). The Est1 enzyme was characterised and showed significant reverse esterase activity *in vitro* (149).

### 2.1.2 Hemiacetal dehydrogenation

Hemiacetals are formed *in vivo* by spontaneous reaction of an aldehyde with an alcohol. The subsequent NAD(P)-dependent hemiacetal dehydrogenation results in ester formation (Fig. 2b). The enzymes that catalyse the reaction are in some cases referred to as hemiacetal dehydrogenases (150,151). In a strict sense, this type of enzyme does not exist, since the HADH reaction is a side activity of certain alcohol dehydrogenases. It was proposed that the side activity is due to the resemblance of hemiacetals to secondary alcohols (152). Hemiacetal dehydrogenation was first observed in methylotrophic yeast (153). When the yeasts were grown on methanol or ethanol as carbon source, high concentrations of formaldehyde and acetaldehyde accumulated, respectively, which are toxic for most organisms (154). Hemiacetal dehydrogenation may act as mechanism to detoxify aldehydes by converting them to esters *via* hemiacetals (155). This activity has been observed for methyl formate synthesis in *Pichia methanolica*, *Candida boidinii* and *S. cerevisiae* (150,156,157). Hemiacetal dehydrogenation may also contribute to ethyl acetate formation in *Neurospora crassa*, *S. cerevisiae*, *Cyberlindnera jadinii*, and *Kluyveromyces marxianus*, but this has not been confirmed *in vivo* (150,151,158,159).



### 2.1.3 Baeyer-Villiger monooxygenases

BVMOs are flavin-containing enzymes that require NAD(P)H as cofactor. They catalyse the insertion of oxygen between a C-C bond in aldehydes and ketones (Fig. 2c). BVMOs are characterized by a FXGXXXHXXXW(P/D) sequence motif and have been found in all domains of life (160,161). In nature, BVMOs participate in the synthesis of secondary metabolites (162,163) or enable utilisation of unconventional carbon sources, such as alkanes, ketones or cyclic alcohols (130,164,165). They catalyse the conversion of ketones to esters, which are further hydrolysed to readily metabolisable alcohols and acids. Such pathways enable *Gordonia sp.* or *Pseudomonas veronii* to grow on propane or methyl ketones, respectively (131,166,167). The structures, functions and applications of BVMOs have been reviewed extensively elsewhere (160,168-171).

### 2.1.4 Alcohol acyltransferases

AATs are a large and diverse group of enzymes. They are the main ester-producing enzymes in plants, yeast, filamentous fungi and some bacteria (172-179). AATs form esters by transferring the acyl moiety from an acyl-CoA molecule to an alcohol (Fig. 2d). They vary significantly in their specificities for their alcohol and acyl-CoA substrates (Table 1). As a result, a plethora of esters are produced in nature, ranging from short chain esters such as ethyl acetate, to long chain wax esters (35,178,180).

The most-studied microbial ester-producing AATs are derived from yeasts (Table 1). One of the few other well studied microbial AATs is the TAG and wax ester-producing AtfA (sometimes referred to as wax synthase – WS) from *Acinetobacter baylyi* (181,182). Five yeast AATs have been described thus far, mostly in *S. cerevisiae*. The first yeast AATs with a determined protein sequence are the *S. cerevisiae* Atf1 and its paralog Atf2 (175,183). Atf1 in particular is responsible for 50 % of acetate ester production in *S. cerevisiae* (184), which was associated with attracting insects as a means of dispersal in the environment (132). For *S. cerevisiae* Atf2, a role in sterol metabolism has also been proposed (185). The second paralog-pair of *S. cerevisiae* AATs are Eht1 and Eeb1 (176). They produce medium chain fatty acid (MCFA) ethyl esters as a way to detoxify MCFAs that accumulate during fatty acid synthesis (186).

The most recently identified yeast AATs compose the Eat1 family. Homologs of this enzyme are responsible for 80 % of ethyl acetate synthesis in *K. lactis* and are present in all other known

bulk ethyl acetate producing yeasts. A homolog of Eat1 also contributed to 50 % of ethyl acetate synthesis in *S. cerevisiae* (35). The role of Eat1 in ethyl acetate synthesis by *S. cerevisiae* was confirmed when the enzyme was identified through polygenic trait analysis (187). The *S. cerevisiae* Imo32 is distantly related to Eat1 and also contributed to ethyl acetate production when overexpressed in *S. cerevisiae* lacking *atf1* (187). Eat1 mainly contributes to the high amounts of ethyl acetate produced by yeasts such as *K. marxianus* and *W. anomalus* (35,88). Bulk ethyl acetate production was suggested to be associated with suboptimal growth conditions under which excess acetyl-CoA accumulates in yeast mitochondria (188,189). Ethyl acetate synthesis by Eat1 may relieve this accumulation by converting acetyl-CoA to ethyl acetate. The localisation of Eat1 in yeast mitochondria supports this hypothesis (117,190,191). The cellular location of the remaining AATs has not been clearly linked to their proposed biological function. The *S. cerevisiae* Atf1 and Atf2 localise to the endoplasmic reticulum (ER), but are also associated to lipid droplets in the cytosol (190,192,193). Curiously, Atf1 and Atf2 homologs from *W. anomalus* and *K. lactis* did not localise to lipid droplets when expressed in *S. cerevisiae* (192). Eht1 was traced to the ER, the outer mitochondrial membrane, and lipid bodies of *S. cerevisiae* (190,194,195), while its paralog is located in the mitochondria (195) (Table 1).

#### **2.1.4.1 Catalytic features of microbial AATs**

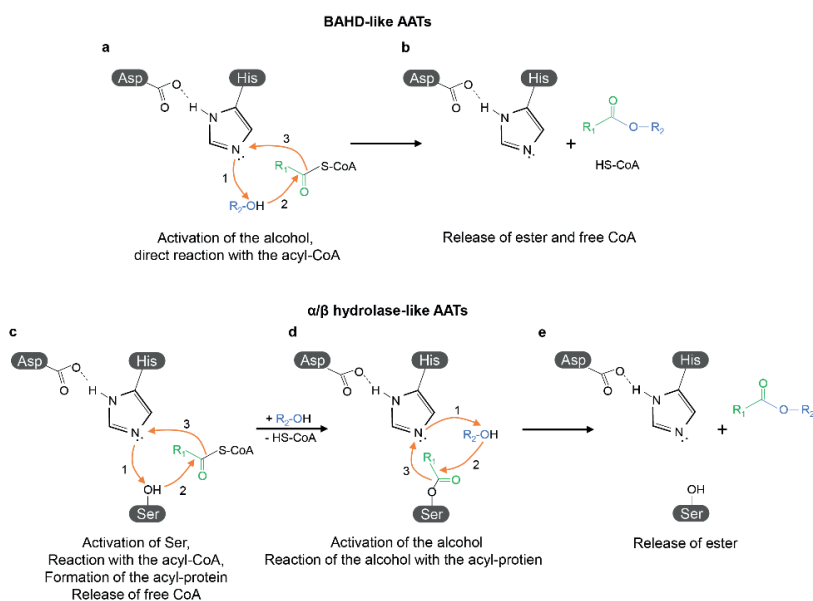
The mechanisms that catalyse the AAT reaction are not entirely understood. There are no crystal structures of AATs available yet, although several computational models have been used to study their catalytic mechanisms (196-198). Microbial AATs belong to two structurally unrelated protein families (Table 1). Atf1, Atf2 and AtfA share some characteristics of the BAHD O-acyltransferases, which are primarily found in plants and bacteria. The abbreviation BAHD is derived from the names of the first four members of this enzyme superfamily that were originally characterised in plants (199). Its members share the conserved HXXXD and DFGWG motifs (196,199). The HXXXD motif participates in transferring the acyl group from the acyl-CoA substrate to the alcohol without forming a covalent acyl-protein intermediate (Fig. 3ab) (181,198). Atf1 and Atf2 lack the DFGWG motif (200), which stabilises the solvent channel of the enzyme and is not directly involved in the catalytic mechanism (201,202). It should be noted that Atf1 and Atf2 are not closely related to AtfA (181) and they belong to different AAT subfamilies (Table 1).

Table 1. Characteristics of the most studied and applied microbial AATs.

	AitA (WS)	AitI	Ait2	EhtI	EebI	EatI
Source, species examples	Bacteria, <i>A. boylii</i>	Yeast, <i>S. cerevisiae</i>	Yeast, <i>S. cerevisiae</i>	Yeast, <i>S. cerevisiae</i>	Yeast, <i>S. cerevisiae</i>	Yeast, <i>W. anomalous</i> , <i>K. marxianus</i> , <i>S. cerevisiae</i>
Typical ester product	Wax esters and TAG	Acetate esters of various alcohols	Acetate esters of various alcohols	MCFA ethyl esters	MCFA ethyl esters	Ethyl acetate, other acetate esters
Proposed biological functions	Synthesis of storage compounds (181)	Attracting insects for dispersion (132)	Sterol metabolism (185)	MCFA detoxification (176)	MCFA detoxification (176)	Relieving acetyl-CoA accumulation (88,117)
Cellular location	Lipophilic inclusion in the bacterial cytosol (182)	ER, Cytosolic lipid droplets (192,193)	ER, Cytosolic lipid droplets (192)	ER, Mitochondrial outer membrane, lipid droplets (190,194,195)	Mitochondria (195)	Mitochondria (117,190,191)
<sup>1</sup> Protein family	WES acyl-transferase-like domain (PF03007)	AA Tase domain (PF07247)	AA Tase domain (PF07247)	$\alpha/\beta$ -hydrolase family 4	$\alpha/\beta$ -hydrolase family 4	$\alpha/\beta$ -hydrolase family 6
Catalytic or conserved regions	HXXXXD, DFGWG (181)	HXXXXD (200)	HXXXXD (200)	Ser-Asp-His triad (176)	Ser-Asp-His triad (176)	Ser-Asp-His triad (35)
Side activities	Unknown	Thioesterase ( <i>in vitro</i> ) (200)	Unknown	Esterase, thioesterase ( <i>in vitro</i> ) (176,203)	Esterase ( <i>in vitro</i> ) (176)	Esterase ( <i>in vitro</i> ), Thioesterase ( <i>in vitro</i> ) (35)
<sup>2</sup> Alcohol specificity	Broad (C4-C20) unbranched primary alcohols ( <i>in vitro</i> ) (182)	Broad towards primary alcohols ( <i>in vitro</i> ) (200)	Primary alcohols ( <i>in vitro</i> ) (178) Sterols ( <i>in vivo</i> ) (185)	Ethanol, Phenylethyl alcohol ( <i>in vivo</i> ) (176,203,204)	Ethanol, ( <i>in vivo</i> ) (176)	Ethanol ( <i>in vitro</i> ) Primary alcohols ( <i>in vivo</i> ) (36)
<sup>2</sup> Acyl-CoA specificity	Preferred long acyl-CoAs (C14-C18) ( <i>in vitro</i> ), lower activity towards shorter and longer acyl-CoAs (182)	Acetyl-CoA only ( <i>in vitro</i> ) (200)	Acetyl-CoA ( <i>in vivo</i> ) (184,185)	MCFA-CoA (C4-C8) ( <i>in vitro</i> ) (176) MCFA-CoA (C6-C12) (205) Caffeoyl-CoA ( <i>in vivo</i> ) (204)	MCFA-CoA (C6-C10) ( <i>in vitro</i> ) (176,205)	Acetyl-CoA ( <i>in vitro</i> ) (35) Propionyl-CoA ( <i>in vivo</i> ) (36)

1 – The PF number represent the classification according to PFAM (206), the  $\alpha/\beta$ -hydrolase family numbers are taken according to the classification in the ESTHER database (207). 2 – Acyl-CoA and alcohol specificities were taken from *in vivo* studies when *in vitro* information was not available. Abbreviations: AATase – alcohol acyltransferase, ER – endoplasmic reticulum, MCFA – medium chain fatty acid, TAG – triacylglycerides, WES – wax ester synthase.

The second group of AATs consists of the paralogs Eht1 and Eeb1, and the recently discovered Eat1 AAT family. They are defined by an  $\alpha/\beta$  hydrolase fold and a Ser-Asp-His catalytic triad (Fig. 3cde) (35,176,203). This fold is typical for a large group of hydrolytic enzymes that includes proteases, esterases, lipases and peroxidases. Eht1 and Eeb1 are only distantly related to Eat1 (35). The hypothetical catalytic mechanism of the  $\alpha/\beta$ -hydrolase fold-containing AATs likely resembles the mechanism observed in other esterase-like acyltransferases. Unlike the catalytic mechanism of BAHD-like acyltransferases, a covalent protein-acyl intermediate is formed and transferred to an alcohol (Fig. 3d). The preference for the acyltransferase reaction of AATs over hydrolysis is mediated by the specific environment within the three-dimensional protein structure (208). They can either exclude water from the active site, favour binding of alcohols over water, or decrease the reactivity of water compared to alcohols (208-210). It is tempting to speculate which mechanism promotes the AAT activity, but such information on the catalytic mechanism can only be obtained by studying the crystal structure of the enzymes and their acyl-protein intermediate (211).



**Figure 3. Hypothetical catalytic mechanisms of BAHD-like and  $\alpha/\beta$  hydrolase-like AATs.** (a, b) – BAHD-like AATs form the ester by activating the alcohol, which reacts directly with the acyl-CoA. No covalent bonds with the enzyme are formed. (c, d, e) – AATs containing the  $\alpha/\beta$ -hydrolase fold most likely perform the reaction *via* a covalent acyl-protein intermediate. The transition states in both hypothetical reaction mechanisms are not depicted to improve the figure clarity. Orange arrows indicate the sequence of the reaction transfer of electrons in the reactions. Numbers indicate the sequence of the reactions.

Despite their structural differences, AATs show remarkably similar catalytic (mis)behaviour. Thioesterase and/or esterase activities have been observed in Eht1, Eeb1 and Eat1 *in vitro* (Table 1). The hydrolytic activity of these AATs is likely related to their  $\alpha/\beta$ -hydrolase fold, and has been described in unrelated acyltransferases containing the same fold as well (208). Thioesterase activity has also been demonstrated in the *S. cerevisiae* Atf1 (200), even though Atf1 does not resemble  $\alpha/\beta$  hydrolases. It was able to act as a thioesterase on longer acyl-CoA substrates, but it could only utilise acetyl-CoA to produce esters in the AAT reaction. While hydrolytic side activities seem to be a characteristic of AATs in general, the interplay of AAT, thioesterase and esterase activities is poorly understood. In Eat1, both thioesterase and esterase activity are repressed by the presence of ethanol *in vitro* (35). The alcohol seems to be the preferred substrate that is able to displace water from the active site of the enzyme, but it is not clear how this occurs. Factors that control the reaction type in other AATs, such as Eht1, Eeb1 and Atf1 have not been determined yet. The catalytic mechanism of AATs should be investigated further to prevent unwanted product or substrate degradation.

### **2.1.4.2 Substrate specificities and kinetic parameters of microbial AATs**

*In vitro* data on substrate specificities is lacking for some AATs. For Eat1, only the specificity towards acetyl-CoA and ethanol has been determined *in vitro* (35). There are also no studies reporting *in vitro* alcohol specificities of the *S. cerevisiae* Eht1 and Eeb1. *In vivo* studies where AAT genes were expressed or deleted may offer some indications on their substrate specificities. However, factors like substrate availability or cellular localisation of the AAT may mask the true substrate specificity of the enzyme, which can only be determined *in vitro*. Two types of enzymatic assays have been developed to measure AAT activity. The first couples the release of CoA (Fig. 2d) to NADH formation or a coloured reaction (35,203,205). As AATs can often display thioesterase activity, care should be taken in interpreting the outcome of such coupled assays. It has been observed that the AAT and the thioesterase reaction can have different substrate specificities. For example, Atf1 could only accept acetyl-CoA as an AAT *in vitro*, but could hydrolyse longer acyl-CoA substrates as a thioesterase (200). The second type of *in vitro* AAT assay measures the ester product directly (*e.g* by gas chromatography). Such assays may provide a more reliable alternative to measuring CoA release (35,182,200,212). Both types of assays have been used to determine some enzyme kinetics of AATs.

AATs seem to be relatively unspecific towards the alcohol substrate (Table 1). For instance, the *S. cerevisiae* Atf1 and Atf2 have a broad specificity towards primary alcohols *in vitro*

(178,200). This was reflected *in vivo* when the genes were overexpressed in *E. coli* and *S. cerevisiae*, leading to a broad increase in acetate ester production (85,184). The overexpression of several *eat1* genes from different yeasts in *S. cerevisiae* also resulted in increased levels of various acetate esters (36). AtfA showed a particularly broad substrate specificity towards alcohols and accepted anything from short alcohols to long wax alcohols (182). Less information is available for the alcohol specificities of the *S. cerevisiae* Eht1 and Eeb1. They are able to synthesise ethyl esters *in vivo* (176,203,213). Recently, Eeb1 was shown to utilise phenylethyl alcohol *in vivo*, indicating some degree of promiscuity of this enzyme (204). Promiscuity towards the alcohol substrate was also observed in plant AATs (172,205). Even the antibiotic resistance marker chloramphenicol acetyltransferase (Cat), which normally acetylates the antibiotic chloramphenicol, was able to form low amounts of aliphatic esters in *E. coli*. (85).

**Table 2. Enzyme kinetics of the most studies and applied microbial AATs.**

AAT	Substrate	$K_M$ (mM)	Kcat (s <sup>-1</sup> )	Kcat/ $K_M$ (s <sup>-1</sup> M <sup>-1</sup> )	Co-substrate	Measurement method	Reference
<b>AtfA</b>	Palmitoyl-CoA;	0.029	n.a.	n.a.	Hexadecanol	Detection of ester product	(182)
<b>Atf1</b>	Isoamyl alcohol	32	n.a.	n.a.	Acetyl-CoA	Detection of ester product	(178)
	Isoamyl alcohol	26	2.9	113	Acetyl-CoA	Coupled to CoA release	(214)
	Acetyl-CoA	0.061	0.4	6656	Ethanol	Coupled to CoA release	(200)
<b>Atf2</b>	Isoamyl alcohol	22	1.6	74	Acetyl-CoA	Coupled to CoA release	(214)
<b>Eht1</b>	Octanoyl-CoA	0.002	0.28	150000	Ethanol	Detection of ester product	(203)
<b>Eat1</b>	Ethanol	3.1	n.a.	n.a.	Acetyl-CoA	Detection of ester product	(35)
	Acetyl-CoA	2.4	n.a.	n.a.	Acetyl-CoA	Detection of ester product	(35)

The specificity of AATs towards their acyl-CoA substrate is higher compared to alcohols in some cases. *S. cerevisiae* Atf1 exclusively utilised acetyl-CoA in *in vitro* AAT assays (200). Other AATs, such as the *A. baylyi* AtfA are more flexible and can accept multiple long acyl-CoAs. However, the number of acyl-CoAs accepted by AtfA is lower compared to the variety of accepted alcohols (Table 1) (182). The *S. cerevisiae* Eht1 and Eeb1 accept a broader range of MCFA-CoA substrates *in vitro* and *in vivo* (Table 1). It should be noted that the preferred acyl-CoA substrates of *S. cerevisiae* Eht1 and Eeb1 differ slightly among studies. For example, the *S. cerevisiae* Eht1 was initially named Ethanol hexanoyl-transferase because it showed the highest activity against hexanoyl-CoA (176), but appeared to be more active towards octanoyl-CoA in other studies (203,205). Nevertheless, all studies agree that the *S. cerevisiae* Eht1 and

Eeb1 have activity against a range of MCFA-CoA substrates. In one study, the *S. cerevisiae* Eht1 was even able to utilise caffeoyl-CoA, which contains an aromatic ring in its structure (204).

Various enzymatic properties of AATs, such as  $K_M$ , or the  $K_{cat}/K_M$  ratio have been determined (Table 2). However, there is still a significant lack of detail in this area. For some AATs, such as Eeb1, no kinetic parameters have been determined to date. For other AATs, the knowledge is still incomplete. Based on information available for Atf1 and Eat1, it seems that these yeast AATs have lower affinity towards alcohols, compared to acetyl-CoA. In case of Atf1, the catalytic efficiency ( $K_{cat}/K_M$ ) for isoamyl alcohol is also significantly lower relative to acetyl-CoA (Table 2). Whether such observations can be made in other families of AATs is still unknown. It is also unclear what the relevance of such low affinities is regarding *in vivo* ester synthesis.

### 3 Metabolic engineering for *de novo* ester synthesis

Competing with the petrochemical industry is challenging due to the low market prices of commodity chemicals. Bioconversion of substrates into products must therefore be as efficient as possible (103). The amounts of esters naturally produced by microorganisms are generally too low to support cost-competitive biobased processes and considerable metabolic engineering efforts have been invested to enhance this production. The crucial factors in any metabolic engineering strategy are the selection of a suitable catalyst and a sufficient supply of metabolic precursors. Most studies on ester production with microorganisms use sugars, particularly glucose as substrate and are the focus of this review.

#### 3.1 Selection of the catalyst

Four main enzymatic reactions are available for engineering ester production *in vivo* (Fig. 2). Esterases and lipases have been applied extensively for ester synthesis *via* transesterification or reverse esterase activity in nearly non-aqueous environments (215-218). However, the reverse esterase reaction is thermodynamically unfavourable in aqueous conditions under which microbial fermentations occur. Metabolic engineering of ester production using HADH has not been reported yet. To produce esters *via* this route, accumulation of aldehydes would be necessary. This may be challenging due to their toxicity (154). Furthermore, hemiacetal formation is spontaneous and requires an acid catalyst, which is not present under physiological pH-neutral conditions.

Biotransformation of cyclic ketones into lactones (cyclic esters) by BVMOs have been studied extensively (169). Several BVMOs exist that are also active towards linear ketones, but they are rarely applied *in vivo* (219-221). Such *in vitro* processes rely on external supply of expensive cofactors, such as NAD(P)H. Direct synthesis of esters from cheaper substrates like sugars may therefore be more economical. BVMO-catalysed ester production also depends on the supply of ketones, which are not common microbial metabolites.

AATs have dominated the field of metabolic engineering of *in vivo* ester production. This reaction is thermodynamically favourable and does not require the input of reducing equivalents (Fig. 2d). AATs convert an alcohol and an acyl-CoA to an ester, releasing free CoA in the process. Both the alcohol and acyl-CoA are readily produced in biological systems and can serve as efficient precursors for ester synthesis. We here outline the general metabolic engineering strategies that have been applied to increase alcohol, acyl-CoA and ester synthesis in microorganisms *via* AATs, and highlight several notable achievements (Table 3).

### 3.2 The building blocks of ester synthesis

The AAT reaction determines the efficiency of the final catalytic step in ester formation. However, the reaction also depends on the supply of the alcohol and acyl-CoA substrates. In several studies precursors for ester synthesis have been added to the cultivation medium, such as acids or alcohols (35,85,222-224). These studies have been useful for studying ester formation capacities of e.g. engineered strains. However, *de novo* ester synthesis from renewable substrates is often preferred. Ideally, carbon sources which do not compete with food and feed would be used for ester synthesis, for example lignocellulosic biomass. Layton and Trinh have also proposed anaerobic waste digestions as platform for ester synthesis. The final products for such processes are a mix of MCFA. They, or their intermediates can be used to produce a mix of esters (37,225). However, the majority of lab-scale studies on ester production have used crystalline glucose as carbon source, which will be the focus of this review (Fig. 4).

#### 3.2.1 Basic metabolites of ester synthesis

Metabolites, such as phosphoenolpyruvate (PEP), pyruvate and acetyl-CoA lie at the core of acyl-CoA, alcohol, and consequently ester synthesis. These carbon compounds are formed during the oxidation of glucose in the glycolysis. During this conversion, reducing equivalents (NADH, NADPH or  $Fd_{red}$ ) and metabolic energy (ATP) are released (Fig. 4a). In the presence of oxygen, ATP is also produced by oxidative phosphorylation. The carbon compounds,



reducing equivalents, and ATP are subsequently used to produce acyl-CoAs (Fig. 4b) and alcohols (Fig. 4c). Finally, the ester is formed by an AAT (Fig. 4d).

Several variants of glycolysis exist (Fig. 4a). The main differences between them are the number and types of reducing equivalents produced, and their ATP yield (226,227). The Embden-Meyerhof-Parnas (EMP) pathway is the conventional glycolytic pathway, which produces NADH and yields 2 ATP per glucose (Fig. 4a). Parallel glycolytic routes include the pentose phosphate pathway (PPP) and the Entner-Doudoroff (ED) pathway, which produce both NADPH as well as NADH (Fig. 4a) (228,229). In the PPP pathway carbon is lost in the form of CO<sub>2</sub>. The final product of these glycolytic pathways is pyruvate, which can be used for e.g. alcohol synthesis (Fig. 4c), or converted further to acetyl-CoA (Fig. 4a).

Acetyl-CoA can be used directly to produce a variety of acetate esters, or it can be condensed into longer acyl-CoA moieties (Fig. 4b) (152,230). The oxidative decarboxylation of pyruvate to acetyl-CoA has been engineered and reviewed extensively (231,232). Pyruvate formate lyase (Pfl), pyruvate dehydrogenase (Pdh) and pyruvate-ferredoxin oxidoreductase (PFOR) catalyse the conversion in a single enzymatic step. Other pathways convert pyruvate to acetyl-CoA in a series of reactions (Fig. 4a). These include the coupling of pyruvate decarboxylase, acetaldehyde dehydrogenase and acetyl-CoA synthetase (Pdc-Ald-Acs, respectively). This pathway is sometimes referred to as the Pdh bypass. An alternative pathway couples Pdc and the acetylating acetaldehyde dehydrogenase (Pdc-aAld) to produce acetyl-CoA directly (Fig. 4). In most cases, CO<sub>2</sub> and a reduced cofactor are produced. The exception is Pfl, which functions in bacteria under anaerobic conditions and releases formate instead (233) (Fig. 4a). Most naturally occurring acetyl-CoA-forming pathways release NADH as the cofactor, although variants of NADP-dependent Pdh and Ald have been described and engineered (Fig. 4a) (234-236).

Most acetyl-CoA forming reactions do not consume ATP, with the exception of acetyl-CoA formation *via* Pdc-Ald-Acs (Fig. 4a). This pathway is the main acetyl-CoA forming pathway in the *S. cerevisiae* cytosol (237). Many industrially interesting compounds, including esters rely on the supply of cytosolic acetyl-CoA in this yeast. Engineering of energy-efficient bypass reactions has therefore been the focus of several studies. These bypass mechanisms include cytosolic expression of the acetylating acetaldehyde dehydrogenase (aAld) or Pdh (Fig. 4a) (238,239). The availability of cytosolic acetyl-CoA in yeast has also been increased by introducing the ATP-citrate lyase, which converts mitochondrial citrate to acetyl-CoA and

oxaloacetate at the cost of ATP. Expression of a phosphoketolase (Xpk) also resulted in an increased flux towards acetyl-CoA (240). This enzyme cleaves acetyl-P units from xylulose-5P and fructose-6P (241). The acetyl-P can then be converted to acetyl-CoA *via* phosphotransacetylase (Pta) (Fig. 4a). The availability of acetyl-CoA in *E. coli* could also be manipulated by upregulating the production of the cofactor CoA. Overexpression of pantothenate kinase, which limits CoA synthesis in *E. coli*, resulted in improved production of CoA as well as acetyl-CoA (242).

The amounts of acetyl-CoA, reducing equivalents, and ATP that are required for ester biosynthesis are determined by the subsequent biosynthetic pathways for the AAT substrates acyl-CoA and alcohol (Fig. 4cb). These compounds are biotechnologically relevant in their own right for a number of applications, such as production of sustainable chemical building blocks, fragrances or biodiesels. The metabolic engineering approaches that were used for engineering acyl-CoA and alcohol supply for these products can also be applied to ester synthesis.

Due to the high energetic demand of fatty acid elongation, the activity of the FAS complex is tightly regulated. Increasing the activity of this complex has been the target of many studies aimed at increasing *de novo* fatty acid and fatty acyl-CoA production. Strategies include increasing the activity of rate limiting steps, such as acetyl-CoA carboxylase (Acc) (Fig. 4b), or removing the regulatory mechanisms that repress the FAS cycle (243-247). The efficiency of FAS can also be limited by the supply of acetyl-CoA. Implementation of alternative acetyl-CoA generating routes, such as Pfl or Xpk-Pta (Fig. 4a) in the cytosol of *Yarrowia lipolytica* (*Y. lipolytica*) increased the lipid titre 1.5 and 1.6-fold, respectively (248). The operation of the FAS cycle also requires high supply of NADPH. In yeast such as *Y. lipolytica*, NADPH is derived from the PPP pathway, which limits the maximum lipid yield due to the decarboxylation step in the pathway (Fig. 4a). To circumvent carbon loss, the metabolism of the yeast was rewired for more efficient NADPH supply, resulting in the highest fatty acid titre to date of 99 g/L (249).

The reversed  $\beta$ -oxidation (RBOX) provides an energetically efficient alternative pathway for acyl-CoA synthesis (250). As the name suggests, the RBOX is the reversal of the  $\beta$ -oxidation pathway that normally oxidises fatty acyl-CoAs (251). Elements of the RBOX pathway are naturally present in *E. gracilis* and are the main source of acyl-CoA moieties used for the synthesis of wax esters under anaerobic conditions (136). The key difference between the FAS and RBOX is the reaction used for the entry of acetyl-CoA into the cycle (Fig. 4b) (252). The

RBOX uses the thiolase (Thl)-catalysed entry of acetyl-CoA into the cycle which avoids the ATP-consuming malonyl-CoA formation (Fig. 4b). In most engineered RBOX systems NADH instead of NADPH is used to reduce the acyl-CoA intermediate. However, by combining the Thl reaction with the FAS II system in *E. coli*, an NADPH-dependent RBOX metabolism was engineered (253). The RBOX operates *via* CoA, and not ACP intermediates (Fig. 4b). An acyl-CoA is thus the direct final product of the RBOX pathway, which can be used for further synthesis. The net cost of an RBOX cycle per acetyl-CoA incorporated is 2 NAD(P)H while the reaction does not consume ATP. As a result, glucose can be converted to acyl-CoAs at the theoretical maximum yield while still generating ATP (250). Implementation of the RBOX in *E. coli* or *S. cerevisiae* required extensive modifications of the regulatory mechanisms that control the native  $\beta$ -oxidation (250,254). The pathway has been used to produce a variety of acyl-CoAs from butyryl-CoA (C4) to decanoyl-CoA (C10), which are precursors for a variety of MCFA, dicarboxylic acids, and other biotechnologically relevant compounds, including esters (250,252,255,256).

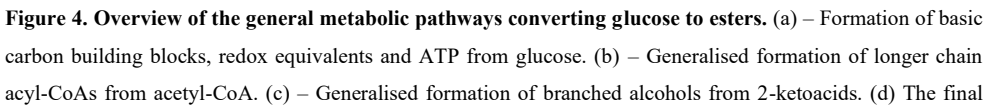
Another class of acyl-CoAs are the branched chain (BC) acyl-CoAs. The formation of these precursors is tightly linked to the generation of higher alcohols (Fig. 4c). They are derived from BC-2-ketoacids, which are in turn derived from the amino acid metabolism (257). The reaction is catalysed by the BC-2-ketoacid dehydrogenase (BC-2-KDH) and is analogous to acetyl-CoA formation by Pdh (Fig. 4ac) (258). Introduction of this pathway in *E. coli* enabled the production of isovaleryl-CoA, 3-methylvaleryl-CoA and isobutyryl-CoA from leucine, isoleucine and valine, respectively (85). BC- acyl-CoAs can also be generated by converting BC-alcohols *via* an appropriate aldehyde dehydrogenase (Ald) and alcohol dehydrogenase (Adh) (85). Alcohols and acyl-CoA can generally be interconverted (Fig. 4d), although the reaction from acyl-CoA to alcohol is thermodynamically more feasible (137).

Table 3. Overview of metabolic engineering approaches for increasing microbial ester production.

Target ester	Organism	Goal/Strategy	AAT catalyst	Acyl-CoA supply	Alcohol supply	Effect/Titre	Notes	References
2-methyl-1-butyl acetate, 2-phenylethyl acetate	<i>E. coli</i>	Acetate ester prod.	See AtfI	Native Acetyl-CoA	Overexpression of Lla 2-KDC gene	≈ App. 20 mg/L		(85)
Isobutyl acetate	<i>E. coli</i>	Isobutyl ac. prod.	See AtfI	Native Acetyl-CoA	Overexpression of 2-keto pathway genes for isobutanol	17 g/L	80% theor. max. yield, Hexadecane layer as organic extraction phase	(85)
Isobutyl acetate	<i>E. coli</i>	Isobutyl ac. prod.	See AtfI	Native Acetyl-CoA	Overexpression of 2-keto pathway genes for isobutanol	36 g/L	42 % theor. max. yield	(214)
Isoamyl acetate	<i>E. coli</i>	Isoamyl ac. prod.	See AtfI	Native Acetyl-CoA	Overexpression of 2-keto pathway genes for isoamyl alcohol	386 mg/L	11% theor. max.	(214)
Isoamyl acetate	<i>E. coli</i>	Increased CoA and acetyl-CoA supply	See Atf2	Increased CoA production (1), Combined with Pdh overexpression (2)	Isoamyl alcohol supplementation	2.3-fold increase (1) 6.2-fold increase (2)	Suppl. of pantothenic acid needed (CoA precursor)	(259)
Ethyl acetate	<i>S. cerevisiae</i>	Pathway colocalization	See AtfI	Ald-Acs targeted to AtfI (lipid droplet)	Native ethanol	2-fold increase		(260)
Butyl butyrate	<i>Clostridium acetobutylicum</i>	Butyl butyrate production	<i>Mallus sp.</i> (apple) AAT	Native butyryl-CoA	Native 1-butanol	45 mg/L	Butyl acetate and ethyl butyrate also detected	(27)
Isobutyl Isobutyrate	<i>E. coli</i>	Branched esters production	See EhtI Cat	Isobutyryl-CoA synthesis via heterologous BC-2-KDH, Introduced RBOX	Conversion of isobutyryl-CoA to isobutanol by native Adh	27 mg/L		(85)
MCFA ethyl ester	<i>S. cerevisiae</i>	Implement RBOX	Native AATs		Native ethanol	Increased MCFA ethyl esters	Quantity not reported	(254)
Butyrate ester platform	<i>E. coli</i>	Synthesis of various butyrate esters	Fan AAT (SAAT)	Butyryl-CoA module ( <i>Clostridium</i> pathway)	Ethanol, isopropanol, isobutanol, butanol modules (2-keto pathway)	Production of ethyl-, isopropyl-isobutyl, and butyl butyrate esters	First engineered <i>de novo</i> synthesis of esters from glucose under anaerobic conditions: Dodecane as extraction phase	(261)
Ethyl and isobutyl ester platforms	<i>E. coli</i>	Conversion of waste carboxylates to esters	Fan AAT (SAAT) or Fve AAT (VAAT)	Activation of carboxylates (e.g. acetate, propionate, pentanoate, hexanoate) to acyl-CoAs	Ethanol and isobutanol modules (2-keto pathway)	Production of a library of expected ethyl or isobutyl esters	Esters such as pentyl pentanoate were also produced.	(225)
Lactate esters platform	<i>E. coli</i>	Formation of lactate esters, focus on ethyl- and isobutyl lactate	Fve AAT (VAAT)	Activation of lactate to lactyl-CoA via Per	Ethanol or isobutanol modules	App. 10 mg/L ethyl lactate and isobutyl lactate	First <i>de novo</i> synthesis of lactate esters from glucose	(262)

<b>1,6-diacytoxy hexane</b>	<i>E. coli</i>	Diterminal acetylation of $\alpha,\omega$ -alcohols	See AtfI	Native acetyl-CoA	Diterminal oxidation of hexane to hexanediol	4.3 g/L	Resting cells, supplementation of hexane, BEHP as organic extraction phase	(263)
<b>Diethyl adipate</b>	<i>E. coli</i>	Diterminal ethylation of $\alpha,\omega$ -fatty acids	See EebI	Diterminal activation of adipic activation via acyl-CoA ligase	Ethanol supplementation	20 mg/L		(213)
<b>FASBE</b>	<i>S. cerevisiae</i>	Combination of FAS and 2-keto pathway	<i>Marinobacter sp.</i> WS	Native FAS synthase; Deletion of <i>tpd3</i> and <i>opi1</i> (repressors of fatty acyl-CoA synthesis)	Overexpression of 2-keto pathway to isobutanol, isoamyl alcohol, amyl alcohol	230 mg/L FASBE		(264)
<b>FASBE</b>	<i>E. coli</i>	Combination of FAS and 2-keto pathway	Aba WS	Prevention of $\beta$ -oxidation	Overexpression of 2-keto pathway to isobutanol, isoamyl alcohol, amyl alcohol	1000 mg/L FASBE		(265)
<b>FAEE</b>	<i>E. coli</i>	Increased FAS activity	Aba AtfA	FAS – overexpressed thioesterase and MCAT (initial step in FAS)	Introduced Pde-Adh	674 mg/L FAEE (C12-C18)	9.4% of th. max yield	(266)
<b>FAEE</b>	<i>S. cerevisiae</i>	Additional acetyl-CoA and NADPH	Mhy WS2	Xpk+Pta bypass	Native ethanol	4.6 mg/g CDW	1.6-fold increase	(240)
<b>FAEE</b>	<i>Y. lipolytica</i>	Change cellular location of AAT	Per-AbaAtfA ER-AbaAtfA	Native acyl-CoA metab. in compartments	Native ethanol	111 mg/L 137 mg/L	15-fold increase 19-fold increase	(248)
<b>WE</b>	<i>A. baylyi</i>	Novel FAR	Native WS	Native fatty acyl-CoA	Overexpression of novel FAR	450 mg/L	Highest WE titre to date on glucose only	(267)
<b>WE</b>	<i>A. baylyi</i>	Product diversification	Native WS	Native fatty acyl-CoA	Changed by Pflu LuxCDE	Shift from C18 to C16 WE		(268)

Routine disruptions of common by-product pathways are not listed and can be found in the original references. Abbreviations: Aba – *Acinetobacter baylyi*, Adh – alcohol dehydrogenase, Acs – acetyl-CoA synthetase, Aid – aldehyde dehydrogenase, BC-2-KDH – branched chain 2-ketoacid dehydrogenase, BEHP – bis(2-ethylhexyl) phthalate, Cat – chloramphenicol acetyltransferase, ER – Endoplasmic reticulum, Fan – *Fragaria x ananassa*, FAS – fatty acid synthase, FAR – fatty acid reductase, FASBE – fatty acid short- and branched-chain esters, Fve – *Fragaria vesca*, Lia – *Lactococcus lactis*, MCFA – medium chain fatty acid, Mhy – *Marinobacter hydrocarbonoclasticus*, Per – peroxisome, Pct – propionyl-CoA transferase, Pdh – pyruvate dehydrogenase, Pdc – pyruvate dehydrogenase, Pflu – *Pseudomonas fluorescens*, Pta – phosphotransacetylase, RBOX – reverse  $\beta$ -oxidation, See – *S. cerevisiae*, WS – wax synthase, WE – wax esters, Wan – *W. anomalus*, Xpk – phosphoketolase.



## Chapter 2

ester formation reaction catalysed by an AAT. Also shown is the interconversion of acyl-CoAs and alcohols. Dashes indicate multi-enzyme conversion pathways. Abbreviations: a-Ald – acetylating acetaldehyde dehydrogenase, AAT – alcohol acyl transferase, Acc – acetyl-CoA carboxylase, ACP – Acyl carrier protein, Acs – acetyl-CoA synthetase, Adh – alcohol dehydrogenase, Ald – acetaldehyde dehydrogenase, BC – branched chain, BC-2-KDH – branched-chain 2-ketoacid dehydrogenase, ED-Entner-Doudoroff pathway, EMP – Embden-Meyerhof-Parnas pathway, ETC – electron transport chain, FACL – fatty acid CoA-ligase, FAR – Fatty acid reductase, FAS – fatty acid synthase, Fd – Ferredoxin, GAP – Glycerol 3-phosphate, 2-KDC – 2-ketoacid carboxylase, Ks – ketoacyl synthase, MCAT – Malonyl-CoA acyl carrier protein transacylase, Mpt – malonyl/palmitoyl transferase, Pdc – Pyruvate decarboxylase, Pdh – pyruvate dehydrogenase, PEP – Phosphoenolpyruvate, Pfl – pyruvate formate lyase, PFOR – pyruvate-ferredoxin oxidoreductase, PPP – Pentose phosphate pathway, Pta – phosphotransacetylase, RBOX – reverse  $\beta$ -oxidation, R5P – Ribulose 5-phosphate, TCA – tricarboxylic acid cycle, Te – thioesterase, Thl – thiolase, Xpk – phosphoketolase, Xu5P – Xylulose 5-phosphate.

### 3.2.2 Synthesis of higher and fatty alcohols

The biosynthesis of alcohols is linked to the production of pyruvate, acyl-CoAs and the amino acid metabolism (Fig. 4abc). Ethanol is a simple alcohol and a common microbial fermentative product. In yeast and *Zymomonas mobilis*, ethanol is produced *via* the non-oxidative decarboxylation of pyruvate to acetaldehyde, which is then reduced with 1 NAD(P)H to ethanol. Pyruvate decarboxylase (Pdc) and alcohol dehydrogenase (Adh) catalyse the reactions, respectively (Fig. 4ac). In *E. coli*, ethanol is produced *via* oxidative decarboxylation of pyruvate to acetyl-CoA by pyruvate dehydrogenase (Pdh) (Fig. 4a). In the next step, acetyl-CoA is reduced with 2 NAD(P)H, first to acetaldehyde, and then to ethanol (Fig. 4d). The reactions are catalysed by an aldehyde dehydrogenase (Ald) and an alcohol dehydrogenase (Adh), respectively. In *E. coli*, the bifunctional alcohol/acetaldehyde dehydrogenase (AdhE) performs both reactions in one step.

The catalytic steps that form higher and fatty alcohols are analogous to the conversion of pyruvate to ethanol *via* acetyl-CoA. For example, the 1-butanol produced during the clostridial ABE fermentation is formed from butyryl-CoA *via* an Ald and Adh at the cost of 2 NADH (Fig. 4cd) (269). Fatty acyl-CoAs that are produced in the FAS and RBOX cycles (Fig. 4b) can be converted to fatty alcohols *via* the same reactions. However, this conversion is usually performed by a single enzyme, the fatty acid reductase (FAR) (270) (Fig. 4d). By introducing FAR genes into *S. cerevisiae*, fatty alcohols, such as 1-hexadecanol could be produced from fatty acyl-CoAs (271).

Some alcohols, such as 1-propanol, isobutanol, isoamyl alcohol, 1-butanol and 2-phenylethanol can be produced from 2-ketoacids that are derived from the amino acid metabolism (also

referred to as the 2-keto pathway) (272). This pathway is the source of higher alcohol synthesis in yeast such as *S. cerevisiae*. The specific 2-ketoacids are first decarboxylated to an aldehyde by a 2-ketoacid decarboxylase (2-KDC) and then reduced to an alcohol by Adh. The reactions are analogous to the conversion of pyruvate to ethanol *via* Pdc and Adh (Fig. 4c) (257). Typical approaches for increasing amino acid-derived alcohol production are the disruption of by-product formation, overexpression of 2-ketoacid biosynthetic genes, and the introduction of appropriate 2-KDC and Adh enzymes (85,214,273). This approach enabled the production of 22 g/L isobutanol from glucose at 86 % of the maximum yield (274). The synthesis of amino acid-derived alcohols relies on the supply of NADPH (Fig. 4c). Engineering additional supply of this cofactor in *S. cerevisiae* improved the production of isobutanol and isoamyl alcohol (275).

### 3.3 Engineering microbial ester production

Ester production requires the supply of alcohols, acyl-CoAs and the selection of a suitable AAT. Most metabolic engineering efforts aimed at improving ester production have focussed on improving the supply of alcohols and acyl-CoAs. A common initial strategy is increasing the availability of basic cellular building blocks, such as acetyl-CoA. This is typically achieved by channelling the carbon flux away from competing pathways e.g. by disrupting lactate and acetate production in *E. coli* (276,277). The carbon flux has also been channelled towards the relevant metabolic precursors by introducing alternative or more efficient metabolic pathways. For example, FAEE synthesis in *S. cerevisiae* is limited by the availability of cytosolic acetyl-CoA, as well as NADPH, which fuel the FAS cycle. To overcome the limitation, the Xpk-Pta pathway (Fig. 4a) was expressed in the *S. cerevisiae* cytosol, improving FAEE production by 1.6-fold (Table 3) (278). In *E. coli*, acetyl-CoA supply was improved by increasing the availability of the cofactor CoA. This approach increased isoamyl acetate production in *E. coli* by 6.3-fold when combined with overexpression of the Pdh complex (259) (Table 3).

The synthesis of more complex esters, such as isoamyl- or isobutyl-acetate relies on the activity of pathways that produce the appropriate alcohols. The 2-keto pathway (Fig. 4c) is commonly used to produce such higher alcohols and is naturally present in organisms such as yeast and some lactic acid bacteria. The key enzyme of the pathway is 2-KDC (Fig. 4c) that converts 2-ketoacids to aldehydes. When 2-KDC from *Lactococcus lactis* was introduced in *E. coli* along with the *S. cerevisiae* Atf1, up to 20 mg/L acetate esters were produced (85). Efficient ester synthesis only commenced after the entire 2-keto pathway towards isobutanol production was



introduced (85,274). By using this pathway, isobutyl acetate was produced from glucose at 80 % of the maximum theoretical yield. Other authors have reported similar improvements for the production of isobutyl and isoamyl acetate *via* the 2-keto pathway, using the *S. cerevisiae* Atf1 as the catalyst (214). Some clostridia utilise an alternative pathway *via* reactions that resemble the RBOX to generate higher alcohols, particularly 1-butanol (Fig. 4c). By introducing an AAT from apple (*Malus sp.*), the natural 1-butanol production was used to produce butyl butyrate from glucose (Table 3) (27). Introducing novel pathways for the synthesis of uncommon acyl-CoA, such as lactyl-CoA has enabled the production of lactate esters (262). This pathway relies on the activation of lactate to lactyl-CoA *via* propionate-CoA transferase (Pct), in a reaction that consumes acetyl-CoA. By coupling the synthesis of lactyl-CoA with ethanol and isobutanol production modules, ethyl lactate and isobutyl lactate were produced *de novo* from glucose (Table 3).

The production of FAEE in *E. coli* and yeast can be limited by the supply of fatty acids. The activity of FAS is often the limiting factor, as this complex is tightly regulated. Its activity in *E. coli* was improved by overexpressing one of the first steps of the pathway, malonyl CoA-acyl carrier protein transacylase (MCAT) as well as the thioesterase (Te) that terminated the FAS cycle (Fig. 4b). This allowed FAEE production at almost 10 % of the theoretical maximum on glucose (Table 3) (266). The production of FAEE could also be improved by disrupting the transcriptional repressors that downregulate the synthesis of acyl-CoAs, or by preventing their degradation in the  $\beta$ -oxidation pathway (264,265). Alternatively, the RBOX pathway was used to supply acyl-CoAs instead, resulting in MCFA ethyl ester production in *S. cerevisiae* (254). Increased fatty acid supply could also be achieved by translocating the *A. baylyi* AtfA to the endoplasmic reticulum or peroxisome of *Y. lipolytica*. The fatty acid precursors are more abundant in this organelle, which resulted in a 15- to 19-fold improved FAEE synthesis in this yeast (Table 3).

Beyond the supply of metabolic precursors, the selection of a suitable AAT determines the final ester product. Microbial AATs (Table 1) have been used most extensively for improving ester production in microbial hosts. Plant AAT genes have also been used to evoke ester production. These enzymes belong to the BAHD family of AATs and share structural similarities to AtfA, Atf1 and Atf2 (Table 1). Two members, SAAT and VAAT, have often been used for engineering ester production in microbial hosts (27,37,225). SAAT originates from the garden strawberry (*Fragaria x ananassa*) and VAAT was derived from the wild strawberry (*Fragaria vesca*). SAAT and VAAT can accept a variety of alcohols ranging from linear alcohols, such

as ethanol, butanol, and isoamyl alcohol to aromatic alcohols like geraniol (172). They are also flexible in the acyl-CoA moiety and have been used to synthesise a variety of acetate, propanoate, butanoate, pentanoate, hexanoate, as well as lactate esters (27,37,262). Often, homologs of the same enzyme originating from different organisms exhibit altered kinetic parameters or substrate specificities (178,279). The first step in many studies has therefore been bioprospecting of several AAT candidates before selecting the optimal one (85,214,279). Other engineering approaches targeting the AAT enzyme or reaction are not commonplace, but they can be effective. For example, co-localising the *S. cerevisiae* Atf1 with the acetyl-CoA generating Ald-Acs pathway (Fig. 4a) led to a metabolic channelling effect and resulted in a 2-fold increase in ethyl acetate formation (Table 3) (260).

The flexibility of the AAT reaction hypothetically allows for any alcohol to react with any acyl-CoA. This has enabled the development of modular *E. coli* cell factories, where each metabolic pathway towards the synthesis of an alcohol or acyl-CoA constitutes a submodule (261). By combining a metabolic module for butyryl-CoA synthesis with various alcohol modules, the synthesis of ethyl butyrate, propyl butyrate, isobutyl butyrate, and butyl butyrate from glucose was observed (Table 3). A similar principle was applied to convert various carboxylates (waste products of anaerobic digestion of lignocellulose) to ethanol or isobutanol esters (37). The acyl-CoA module was based on activating the carboxylates *via* FACL (Fig. 4b) while ethanol and isobutanol were produced through their respective pathway submodules (Table 3).

In *S. cerevisiae* and *E. coli*, the 2-keto production pathway towards isobutanol or isoamyl alcohol was combined with the synthesis of fatty acid acyl-CoA to give rise to a number of fatty acid short and branched-chain esters (FASBE) (264,265,280), which can be applied as biodiesels. In *A. baylyi*, the spectrum of wax esters could be adapted by expressing alternative FAR enzymes. These enzymes converted the fatty acyl-CoAs produced by the organism to fatty alcohols. By using an alternative FAR system from *Pseudomonas fluorescens*, the product spectrum shifted from C18 wax esters to C16 wax esters (Table 3) (268). The versatility and broad substrate specificities of some AATs also allow for the design of novel esters that may not exist in nature (Table 3). For example, the *S. cerevisiae* Atf1 is able to acetylate both termini of various  $\alpha,\omega$ -diols, such as hexanediol and pentanediol (213). Similarly, Eht1 could react ethanol with activated  $\alpha,\omega$ -fatty acids, resulting in the production of di-ethyl esters (263).

Recent understanding and engineering of (heterologous) ester production has enabled the development of *in silico* models, such as MODCELL and ModCell2 (281,282). These tools are

able to predict and design modular cells and metabolic pathways, that can couple the production of several ester platforms, based on their common metabolic reactions. By this approach, genetic engineering for the design of novel ester producing strains can be combined and optimised, reducing the overall workload.

## 4 Beyond esters as final products

The various metabolic engineering strategies and increased understanding of microbial ester synthesis should boost developments towards production of biobased esters as high value compounds and bulk chemicals. As esters can be easily hydrolysed to organic acids and alcohols, which are valuable bulk chemicals in their own right (226), the biobased production of these compounds may profit from enhanced ester production as well, e.g. by decreasing product toxicity or facilitating product removal. In addition, esterification of intermediates improves microbial production of  $\alpha,\omega$ -diols as they act as a protective group during conversions.

### 4.1 Esters for the production of alcohols and carboxylic acids

By adding a hydrolysis step, the organic acid and alcohol portion of the produced esters can be readily recovered in a potential bioprocess. The following section highlights a few cases in which alcohol or acid production *via* an ester intermediate might be beneficial based on their physical properties.

A minimum product titre of 50 g/L is considered acceptable when implementing a biobased process or the downstream processing (DSP) steps would become too cost intense (103). Product toxicity often prevents reaching sufficiently high titres during microbial production of chemicals, especially regarding alcohols and acids. The presence of 4 g/L (50 mM) 2-butanol already negatively affected the growth rate of *S. cerevisiae*, *E. coli* and *B. subtilis*, while concentrations of 16 g/L butanol even inhibit growth and continuation of the fermentation in solventogenic clostridia (283,284). A similar effect is observed for organic acids such as acetic acid, propionic acid, or butyric acid where growth is completely inhibited at concentrations exceeding 5 g/L, 11 g/L or 6 g/L, respectively (285). Only for ethanol, fairly tolerant hosts were found, like *S. cerevisiae*, where ethanol tolerances may exceed 100 g/L ethanol (286,287).

Based on their physical characteristics, a relation has been found between compound hydrophobicity, expressed in the  $\log P_{o/w}$  value, and microbial toxicity. The  $\log P_{o/w}$  value

describes how well a compound distributes over an octanol phase in comparison to an aqueous phase (288) but can also be linked to the molecular toxicity of the compound for microorganisms during fermentation (289-292). Straathof and colleagues correlated both the  $\log P_{o/w}$  and the aqueous solubility of a compound to the experimentally determined critical concentration  $C_{crit}^{aq}$  retrieved from various literature, at which cell growth is no longer possible. While this value might slightly vary with the microorganism of choice, the general correlation between toxicity and polarity stays valid (292). The critical concentrations derived from the  $\log P_{o/w}$  correlation of some industrially important esters, acids and alcohols highlight the above-mentioned bottleneck of reaching insufficient titres and dealing with rather severe toxic effects of the products (Table 4). As a rule of thumb, compounds with a  $\log P_{o/w}$  between 0.7 and 4 are considered toxic to an organism. At higher  $\log P_{o/w}$  values the molecular toxicity effect is avoided as the compound is so apolar that it forms a second phase, thus, is no longer interfering with the aqueous system of the microbe. Then, however, one has to account for the toxicity that the second phase imposes on the microorganism, the phase toxicity (293). At lower  $\log P_{o/w}$  values the toxic effects are circumvented as the compound is so hydrophilic that it is no longer interfering with the cell membrane.

Most of the industrially relevant alcohols, acids, as well as esters (Fig. 1) are in the toxic range based on their corresponding  $\log P_{o/w}$  values (Table 4). Nevertheless, microbial production of these esters may be more promising than microbial production of their alcohol or acid precursors. In line with the  $\log P_{o/w}$  values, C2 to C8 esters were generally less toxic to *E. coli* than alcohol or acid equivalents (294). Moreover, the C-mol-based Heat of Vaporization at standard conditions ( $H_{vap}^{\circ}$ ), the energy needed to evaporate a compound of interest, shows that short and medium chain lengths esters consistently require less energy input than their acid or alcohol counterpart (Table 4). *In situ* product removal (ISPR) by, for instance gas stripping or phase extraction has proven an efficient way to increase final yields and titres and is better applicable to esters than to alcohols and acids. Introducing a biphasic system using hexadecane, enabled a yield of 80%, reaching a final titre of 17 g/L isobutyl acetate (85). With gas stripping, isobutyl acetate yields could be increased from 28.8% without stripping to 42% of the theoretical maximum, and reached 50% of the theoretical maximum yield in an ethyl acetate stripping experiment (214,295). For ethyl acetate this would correspond to liquid titres twice as high as the predicted critical concentration, for isobutyl acetate the increase is even more than 10-fold (Table 4). Therefore, this approach offers a way to keep up high productivities by avoiding accumulation of inhibitory product concentrations.

## Chapter 2

**Table 4. Characteristic parameters of several esters and their alcohol and acid derivatives.**  $\log P_{o/w}$  values and the heat of vaporization ( $H_{vap}^{\circ}$ ) were retrieved from PubChem and Chemeo database, the critical concentration ( $C_{crit}$ ) was calculated based on the Straathof correlation (292). The partition coefficient on decane/water was estimated using the LSER approach (296). Solute parameters were calculated by ACDLabs (297) while solvent parameters were exported from (298).

Compound	#C	$\log P_{o/w}$	$C_{crit}$ (M)	$H_{vap}^{\circ}$ (kJ C-mol <sup>-1</sup> )	Partition coefficient (decane/water)
Ethanol	2	-0.31	2.0700	21.20	9.60E-03
Acetic Acid	2	-0.17	1.6787	25.80	1.10E-03
Ethyl acetate	4	0.73	0.1686	8.83	1.94E+00
Butyric acid	4	0.79	0.1343	14.50	n.a.
Butanol	4	0.88	0.1604	13.10	1.69E-01
Butyl acetate	6	1.78	0.0228	7.17	3.63E+01
Ethyl butyrate	6	1.85	0.0149	7.00	3.91E+01
Hexanoic acid	6	1.92	0.0861	11.88	1.02E+00
Hexanol	6	2.03	0.0577	10.17	2.53E+01
Ethyl hexanoate	8	2.40	0.0044	6.08	1.45E+03
Butyl butyrate	8	2.83	0.0021	5.32	7.43E+02
Octanol	8	3.00	0.0024	8.66	n.a.
Dodecanoic acid	12	4.20	0.0001	5.48	n.a.
Hexyl Hexanoate	12	4.40	<0.0001	4.29	4.06E+05
Dodecanol	12	5.10	<0.0001	7.50	n.a.
Oleic acid	18	6.50	<0.0001	4.39	5.27E+06
Ethyl oleate	20	8.00	<0.0001	3.46	7.44E+09

n.a. – not available

Acetic acid, ethanol and ethyl acetate are abundant industrial compounds, with moderate toxicity ( $\log P_{o/w} < 0.7$ ). While a lot of research has been performed on improving ethanol production itself, the latest advances focus on ISPR *via* gas stripping (299,300). A similar approach has been pursued for ethyl acetate production (295). In direct comparison to ethanol, higher volatilities for ethyl acetate relate to a more favourable Henry coefficient. As a consequence, the liquid/gas distribution of ethyl acetate, even at 25°C is approximately 25-fold more beneficial compared to ethanol (301). For acetic acid in turn, gas stripping cannot be recommended as low volatility paired with a high boiling point and good solubility push the equilibrium to the liquid phase. Considering applications of acetic acid as a bulk chemical however, only few studies obtained sufficiently high titres to encourage further research on a biobased process (302). Latest technologies and challenges for the bulk production of acetic acid have been summarized recently and predict yields, DSP efficiency and costs as major challenges (303). Therefore, the production of acetic acid *via* hydrolysis of the ester intermediate could be an attractive alternative, as it leads to a second valuable by-product, ethanol.

When comparing other compounds of the alcohol, acid or ester family, most of them have a poor volatility paired with a boiling point well above the one of water. This makes product removal *via* gas stripping less feasible. Nevertheless, when comparing butanol and butyric acid with esters such as butyl butyrate or ethyl butyrate, biobased production *via* an ester intermediate might again be an interesting solution. Looking at the respective extractabilities of these compounds by an organic phase such as decane, the outcome favours the ester derivative once again, not in the least due to the absence of functional polar groups (Table 4). Despite the fact that all compounds are fairly toxic to any production host ( $\log P_{o/w}$  below 4), ISPR by applying an organic phase might be applied to keep the dissolved concentration low. The potential for liquid/liquid extraction of butanol *via* an organic phase was already mentioned before when different ISPR strategies were compared (283,304,305). Moreover, Oudshoorn *et al.* (2009) concluded, the energy requirements for steam stripping and distillation were approximately 66% of the combustion energy of recovered 1-butanol while extraction and adsorption showed the lowest energy costs with still significant losses of 25 and 22%, respectively (306). For butyl butyrate the extractability by decane is more than 1000-fold better than for butanol, indicating even more potential for energy savings when taking a route *via* the ester intermediate.

Due to the low solubility or good volatility, esters can be extracted from the system with less effort. If these benefits in ester production and extraction are sufficient to compensate for the energy spend on subsequent hydrolysis and recovery of acid and alcohol derivatives, remains to be answered. Especially for the above-mentioned examples, all classes (esters, acids and alcohols) are settled in the most toxic range according to their  $\log P_{o/w}$  values. Microbial production will therefore always be limited by strain robustness and efficiency of ISPR. For these reasons, it might be worthwhile to focus on esters outside the critical  $\log P_{o/w}$  range. These include, but are not limited to hexyl hexanoate ( $\log P_{o/w}=4.4$ ), ethyl decanoate ( $\log P_{o/w}=4.6$ ) or butyl decanoate ( $\log P_{o/w}=5.4$ ). Extraction of these compounds in microbial production systems might be straightforward and energy efficient, as an organic phase is formed spontaneously. The effect of phase toxicity of the different compounds needs to be evaluated.

One benefit of the proposed approaches however, is undebatable in all cases: it enables tuneable processes with esters as fermentation output or the co-production of carboxylic acids and alcohols upon hydrolysis. This provides great flexibility on industrial scale and enables fast responses to changing market demands.

## 4.2 Esters for the production of $\alpha,\omega$ -bifunctional monomers

Esterification can also be exploited for the production of  $\alpha,\omega$ -bifunctional monomers (BMs). These compounds contain a functional group (e.g., hydroxy, carboxy, amine) on both terminal carbon atoms, which allow them to be used as building blocks for a vast array of polymers. BMs are industrially interesting due to their high market demands which are higher compared to molecules with a single functionality. Examples of BMs are  $\alpha,\omega$ -diols and  $\alpha,\omega$ -dicarboxylic acids, which are currently produced by multistep, energy-intensive processes. Hence, much research has been devoted to the development of more environmentally friendly production routes. For medium-chain molecules, the major challenge is the introduction of a second functional group (e.g., hydroxy, carboxy) on the  $\omega$ -position. Alkane monooxygenases can  $\omega$ -oxidize primary alcohols and fatty acids, but efficiencies are low with medium-chain molecules (255,263,307-310). This is likely caused by the presence of the polar group that hampers entry into the hydrophobic binding pockets of the monooxygenase. Esterification of fatty acids hides the polar group and allows alkane monooxygenase AlkB to efficiently  $\omega$ -oxidize esterified fatty acids (263,311-313). Instead of adding esterified fatty acids directly, the fatty acids can also be esterified *in vivo* and then be combined with a monooxygenase to achieve  $\omega$ -oxidation. To achieve this, *E. coli* was equipped with acyl-CoA ligase AlkK and either *A. baylyi* AtfA or *S. cerevisiae* Eeb1 to produce esters from medium-chain fatty acids. The resulting esters were efficiently converted to mono-ethyl dicarboxylic acids. Also, di-ethyl esters accumulated (263). A challenge for such process is the inherent toxicity of medium-chain fatty acids to microorganisms. This could be circumvented by fed-batch addition of the fatty acid, or to produce them *de novo* with subsequent esterification. The same concept can be applied to produce the mono-ethyl dicarboxylic acids from n-alkanes (263). Ideally, this approach should be coupled or combined to microbial alkane production from glucose, which is reviewed elsewhere (314).

For the production of  $\alpha,\omega$ -diols another challenge has to be tackled. Alkane monooxygenases tend to overoxidise primary alcohols to aldehydes and carboxylic acids. Furthermore, under aerobic conditions, alcohol and aldehyde dehydrogenases that are present in the production host can do the same. Esterification of the alcohols is a promising strategy to prevent overoxidation. This protective group chemistry is common in organic chemistry, and has been applied to protect overoxidation of unactivated  $sp^3$  C-H bonds (315-317).

## 5 Challenges and perspectives

This review provides an overview of microbial ester production on a fundamental level, with special attention towards the potential applications in the bulk chemical arena. There are some major challenges ahead before biobased ester production can move to the industrial level. The key ester producing enzyme AAT is not understood well, and studies that focus on fundamental aspects of its function are scarce. For example, it is still not clear why many AATs display thioesterase and esterase activities, and how to control these activities. In case of Eat1, ethanol is the determining factor that shifts the enzyme from a hydrolase to an AAT (35). The controlling factors in other AATs have not been established yet. It is furthermore unclear how efficiently AATs are expressed and translated in heterologous hosts, which may hinder the development of efficient bioprocesses.

Another major challenge is how to fully employ the vast array of AATs that are available in nature. Metabolic engineering studies have hitherto utilised a relatively limited subset of AATs. However, it has been shown that even homologs of the same AAT can show remarkable differences in their substrate specificities. Mining the unknown AATs may provide us with enzymes that are able to perform conversions that are currently either not possible or inefficient. AATs with narrower substrate specificities than the ones employed currently should also be identified. Catalytic promiscuity can be useful as the same enzyme can form different products. This can simplify the development of ester producing-strains, particularly in laboratory research where many metabolic engineering strategies are typically tested at the same time. However, promiscuity can also be a disadvantage, especially on large-scale, where a single fermentation product is desired. Special consideration should also be given to discovering catalytically efficient AATs. In many studies aimed at ester production, the titres achieved were relatively low. The reasons for this are not well understood, but it is likely that some AATs may be catalytically inefficient to allow metabolic flux towards ester synthesis. To enable development of ester-producing processes, efficient AATs should be discovered, or alternatively evolved from known AATs.

The advances in metabolic engineering of ester production are closely linked to the developments in engineering towards production of acyl-CoA (acids) and alcohols. The metabolic diversity of the alcohol and acyl-CoA pathways indicate that a practically limitless number of esters can be designed, assuming an AAT with the appropriate specificity is available. The real challenge of engineering ester production will be to balance the supply of



## Chapter 2

the alcohol and acyl-CoA substrates. Ideally, both substrates will be produced in a 1:1 ratio to ensure efficient conversion of sugars into esters. This is challenging as both substrates are produced in complex and intertwined metabolic pathways.

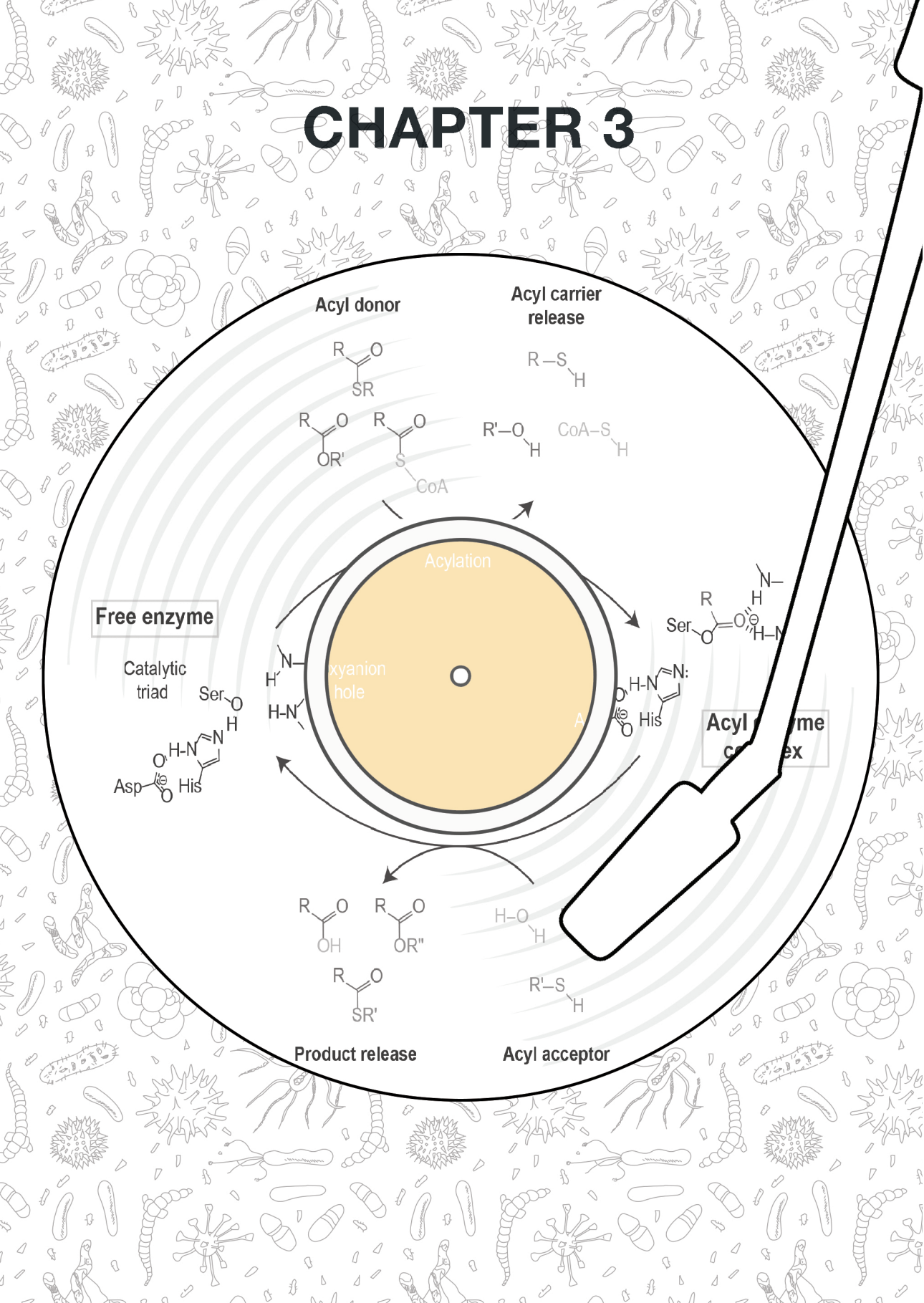
The commercialisation of ester production will require efficient DSP development. Based on the physical properties of esters, they can be removed by gas stripping, extracted into an organic phase, or may even form their own phase. In comparison to their alcohol or acid precursor ISPR is benefitting from the absence of polar groups, making esters generally easier to remove from aqueous systems. Small, volatile esters can freely diffuse from the cell. However, this is not the case for large, insoluble esters, such as wax esters of FAEE. These esters mostly accumulate in the cell and need to be extracted, adding costs to the DSP.

Finally, the true potential of esters as a platform chemical for alcohol and carboxylic acids should be investigated. From a bioprocess engineering perspective, ester formation could be favourable to alcohol and acid formation due to lower toxicity and easier extractability. Esterification could also facilitate the production of high value compounds such as  $\alpha,\omega$ -diols and  $\alpha,\omega$ -dicarboxylic acids and unleash the full potential of biobased ester production.

## 6 Acknowledgements

We would like to thank Nouryon (previously AkzoNobel Speciality Chemicals) and the BE-Basic foundation for financial support.



[illegible]

## Chapter 3

# Eat1-like alcohol acyl transferases from yeasts have high alcoholysis and thiolysis activity

Constantinos Patinios<sup>1,2</sup>, Lucrezia Lanza<sup>1</sup>, Inge Corino<sup>1</sup>, Maurice C. R. Franssen<sup>3</sup>, John Van der Oost<sup>1</sup>, Ruud A. Weusthuis<sup>2</sup>, Servé W. M. Kengen<sup>1\*</sup>

<sup>1</sup>Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Wageningen, the Netherlands

<sup>2</sup>Laboratory of Bioprocess Engineering, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Wageningen, the Netherlands

<sup>3</sup>Laboratory of Organic Chemistry, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Wageningen, the Netherlands

\*Correspondence: Dr. Servé W. M. Kengen ([serve.kengen@wur.nl](mailto:serve.kengen@wur.nl))

Chapter adapted from publication:

Frontiers in Microbiology; doi: 10.3389/fmicb.2020.579844

### Abstract

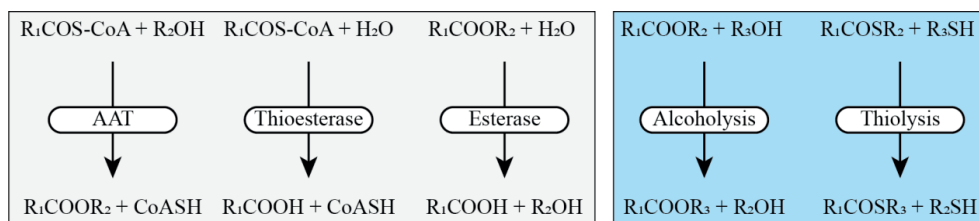
Esters are important flavor and fragrance compounds that are present in many food and beverage products. Many of these esters are produced by yeasts and bacteria during fermentation. Whilst ester production in yeasts through the alcohol acyl transferase reaction has been thoroughly investigated, ester production through alcoholysis has been completely neglected. Here, we further analyze the catalytic capacity of the yeast Eat1 enzyme and demonstrate that it also has alcoholysis and thiolysis activities. Eat1 can perform alcoholysis in an aqueous environment *in vitro*, accepting a wide range of alcohols (C2-C10) but only a small range of acyl donors (C2-C4). We show that alcoholysis occurs *in vivo* in several Crabtree negative yeast species but also in engineered *Saccharomyces cerevisiae* strains that overexpress Eat1 homologs. The alcoholysis activity of Eat1 was also used to upgrade ethyl esters to butyl esters *in vivo* by overexpressing Eat1 in *Clostridium beijerinckii*. Approximately 17 mM of butyl acetate and 0.3 mM of butyl butyrate could be produced following our approach. Remarkably, the *in vitro* alcoholysis activity is 445 times higher than the previously described alcohol acyl transferase activity. Thus, alcoholysis is likely to affect the ester generation, both quantitatively and qualitatively, in food and beverage production processes. Moreover, mastering the alcoholysis activity of Eat1 may give rise to the production of novel food and beverage products.

**Keywords:** Alcoholysis, Thiolysis,  $\alpha/\beta$ -hydrolase, Alcohol Acyl Transferase (AAT), Yeast, Ester, *Clostridium beijerinckii*

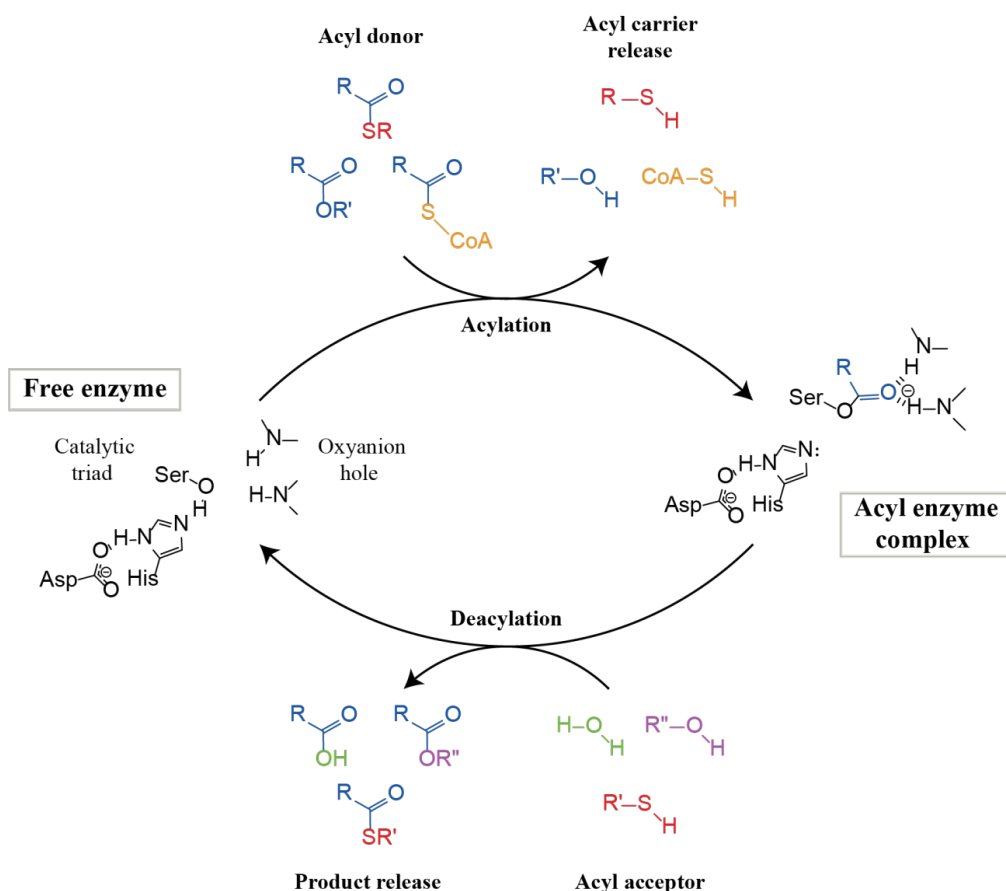
## Introduction

The  $\alpha/\beta$ -hydrolase fold superfamily of enzymes belongs to one of the largest groups of structurally related enzymes sharing a typical  $\alpha/\beta$ -sheet with a conserved active site. Members of this family have diverse catalytic functions, including peptidase (EC 3.4), alcohol acyl transferase (AAT) (EC 2.3.1), esterase (EC 3.1.1), thioesterase (EC 3.1.2), lipase (EC 3.1.1.3) and others (318,319). In fact, members of this superfamily can catalyze 17 different reactions through the same Ser-His-Asp catalytic triad (208). Because of their wide portfolio of catalytic reactions, their high catalytic activity, their high stability and their high regio- and stereoselectivity, some  $\alpha/\beta$ -hydrolase fold enzymes are widely used in biotechnology (320). Applications range from the production of flavors and fragrances such as butyl butyrate and cinnamyl propionate in the food industry, the production of biologically-potent pharmaceutical chemicals, such as (R)-indanol and lutein dipalmitate, and the synthesis or degradation of polymers such as poly ( $\beta$ -caprolactone) or polyhydroxyalkanoates, respectively (321).

A new member of the  $\alpha/\beta$ -hydrolase fold superfamily is the Eat1 protein family. Eat1 was recently discovered in ester-producing yeast species, such as *Kluyveromyces marxianus*, *Kluyveromyces lactis* and *Wickerhamomyces anomalus*, and was identified as the main enzyme responsible for ethyl acetate formation in these yeasts (35,191). Ethyl acetate synthesis by Eat1 is catalyzed through the transfer of the acetate moiety of acetyl-CoA to a free molecule of ethanol in an AAT reaction (Fig. 1). It is hypothesized that the role of Eat1 is to regenerate the free CoA pool in the cell during iron or oxygen limitation through ethyl acetate production (88,117,118,127,188). This hypothesis was strengthened by the localization of Eat1 in the mitochondria of *K. lactis* and *K. marxianus* and its upregulated expression in *W. anomalus* during iron-limited conditions (35,117,191).



**Figure 1. Schematic representation of the Eat1 activities.** Alcohol acyl transferase (AAT), thioesterase and esterase activities have previously been described by Kruis *et al.* (2017) (light grey box). Alcoholysis and thiolysis are described by this study (light blue box). An acyl-CoA is represented by  $\text{RCOS-CoA}$ , a free CoA by  $\text{CoASH}$ , an ester by  $\text{RCOOR}$ , a thioester by  $\text{RCOSR}$ , an alcohol by  $\text{ROH}$  and a thiol by  $\text{RSH}$ .



**Figure 2. Canonical esterase mechanism demonstrating AAT, hydrolysis, alcoholysis or thiolysis reactions.**

The catalytic triad (Asp-His-Ser) along with the oxyanion hole (two amide N-Hs) are indicated in the free enzyme state (left). The first step of the cycle is the binding of the acyl moiety of an acyl-CoA, an ester or a thioester (colored yellow, blue or red, respectively), followed by acylation and the release of the acyl carrier and the formation of the acyl enzyme complex (right). Next, depending on whether an alcohol or a thiol is present in the aqueous medium, hydrolysis, ester formation or thioester formation may occur. In the case of esterase or thioesterase, water (green) binds to the acyl enzyme complex, causing the release of the acid and the restoration of the free enzyme state. In the case of AAT or alcoholysis, an alcohol (magenta) binds to the acyl enzyme complex, which causes the release of an ester. The difference between AAT and alcoholysis is defined by the acyl carrier molecule as a CoA or an alcohol, respectively. Likewise, if a thiol (red) nucleophile is present, the product of the cycle will be a thioester.

In addition to its AAT activity, Eat1 can also catalyze esterase and thioesterase activities (Fig. 1), both through hydrolysis, in an aqueous environment (35). This is not surprising as all three activities follow the canonical esterase mechanism for hydrolysis (Fig. 2). Remarkably,

however, Eat1 was shown to favor ethanol over water as a nucleophile, resulting in a preference for the AAT activity over the competing hydrolysis reactions at elevated alcohol concentrations (35). This feature is unique amongst the  $\alpha/\beta$ -hydrolase fold enzymes as only a small fraction of these enzymes favors alcohols over water in an aqueous environment (208). Therefore, based on the observation that Eat1 can accept both esters and thioesters as acyl donors, along with its preference for ethanol over water as acyl acceptor, we hypothesized that Eat1 might also be able to catalyze alcoholysis and thiolysis (Fig. 1).

Alcoholysis is a transesterase reaction in which the alcohol moiety of an ester is replaced by another alcohol following the general equation:  $R_1COOR_2 + R_3OH \rightarrow R_1COOR_3 + R_2OH$  (Fig. 1). Alcoholysis occurs both in aqueous and in non-aqueous environments using lipases or esterases and has been widely used by the industry for biodiesel production (322). Likewise, thiolysis (or transthioesterification) describes the production of a thioester by donating an acyl group from an acyl donor to a thiol acceptor. The acyl donor can be either an ester or a thioester, but the former is generally used:  $R_1COOR_2 + R_3SH \rightarrow R_1COSR_3 + R_2OH$  (323,324).

Whilst ester production in yeasts has been thoroughly investigated through the AAT reaction, ester production through alcoholysis has been completely neglected. Alcoholysis, can play an important role in the production and distribution of esters in yeasts and consequently affect the final product quality of yeast-derived food and beverage products like beer and wine. Here we investigated the catalytic capacity of the Eat1  $\alpha/\beta$ -hydrolase with respect to alcoholysis and (to a minor extent) thiolysis. We could show that the Eat1 enzyme exhibits a high alcoholysis activity using a broad range of alcohols. In addition, we show that alcoholysis can also play a role *in vivo* in several yeast species and that expression of Eat1 leads to alcoholysis in engineered yeast species and in an engineered bacterium (*Clostridium beijerinckii*). The discovery of the high alcoholysis side activity of Eat1 sheds new light on the production and variety of short chain esters in food and beverage products and might open new research lines for the industrial production of sustainable short chain esters.



## Materials and methods

### Microbial strains and cultivation conditions

The strains used in this study are given in Table 1. Transformed *Escherichia coli* was grown at 37°C on LB agar plates containing 50 mg L<sup>-1</sup> kanamycin or spectinomycin or cultured in liquid LB media containing 100 mg L<sup>-1</sup> kanamycin or spectinomycin. For plasmid construction and propagation, chemically competent *E. coli* NEB® 5-alpha was used and for protein expression, BL21 (DE3) competent *E. coli* was used following the manual provided by the manufacturer (NEB).

The wild type (WT) yeast strains were grown in 50-ml Falcon tubes containing 10 ml of either YPD medium (10 g L<sup>-1</sup> yeast extract, 20 g L<sup>-1</sup> peptone and 20 g L<sup>-1</sup> glucose), yeast minimal medium (YMM) or YMM without iron supplementation (YMM-no-iron) as described previously (117). *In vivo* alcoholysis was enabled by adding 5 mM butyl butyrate to YMM-no-iron. As a control for esterification, WT strains were also grown in YMM-no-iron containing 5 mM butyrate. Cultures were grown in triplicates for 48 h at 30°C on a shaking platform at 200 rpm. Samples were taken after 48 h and stored at -20°C until further use.

*Saccharomyces cerevisiae* strains harboring the pCUP1 variants were grown in 50-ml Falcon tubes containing 10 ml YMM medium with the required growth factors to supplement the auxotrophic requirements of the strains (75 mg L<sup>-1</sup> tryptophan, 500 mg L<sup>-1</sup> leucine and 125 mg L<sup>-1</sup> histidine). Expression of the Eat1 homologs and *in vivo* alcoholysis was assessed by adding 100 µM CuSO<sub>4</sub> and 5 mM butyl butyrate to the growth medium. Control cultures were grown in medium containing 5 mM butyrate. Cultures were incubated at 30°C for 24 h on a shaking platform at 200 rpm and sampled once at the end of the fermentation. Samples were stored at -20°C until further use.

Transformed *C. beijerinckii* strains were grown anaerobically on mCGM agar medium (5 g L<sup>-1</sup> yeast extract, 5.51 mM KH<sub>2</sub>PO<sub>4</sub>, 4.31 mM K<sub>2</sub>HPO<sub>4</sub>, 1.62 mM MgSO<sub>4</sub>, 0.036 mM FeSO<sub>4</sub>, 17.11 mM NaCl, 15.14 mM L-asparagine, 15.14 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.03 mM L-cysteine, 69.4 mM D(+)-glucose) or fermented in GAPES liquid medium (2.5 g L<sup>-1</sup> yeast extract, 7.35 mM KH<sub>2</sub>PO<sub>4</sub>, 3.50 mM K<sub>2</sub>HPO<sub>4</sub>, 4.06 mM MgSO<sub>4</sub>, 0.02 mM FeSO<sub>4</sub>, 1.30 mM pABA, 37.65 mM CH<sub>3</sub>CO<sub>2</sub>NH<sub>4</sub>, 331.02 mM D(+)-glucose, 0.99 mM L-cysteine; (325)) containing 650 mg L<sup>-1</sup> spectinomycin. For *in vivo* alcoholysis, 0, 10, 20, 50 or 100 mM of ethyl acetate or ethyl

butyrate was added to the growth medium and the cultures were grown for 96 h. Samples were taken at 48, 72 and 96 h and stored at -20°C until further use.

**Table 1. Strains used in this study.**

Strain	Genotype	Source
<i>Escherichia coli</i> NEB® 5-alpha	<i>fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	New England Biolabs (NEB)
<i>Escherichia coli</i> BL21 (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHI ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5</i>	NEB
<i>Saccharomyces cerevisiae</i> CEN.PK2-1D	<i>MATalpha, his3D1, leu2-3_112, ura3-52, trp1-289, MAL2-8c, SUC2</i>	(326)
<i>Wickerhamomyces anomalus</i> DSM 6766	Wild type	Leibniz-Institut DSMZ
<i>Wickerhamomyces ciferrii</i> CBS 111	Wild type	Centraalbureau voor Schimmelcultures (CBS)
<i>Kluyveromyces marxianus</i> DSM 5422	Wild type	Leibniz-Institut DSMZ
<i>Kluyveromyces lactis</i> CBS 2359	Wild type	CBS
<i>Cyberlindnera jadinii</i> DSM 2361	Wild type	Leibniz-Institut DSMZ
<i>Cyberlindnera fabianii</i> CBS 5640	Wild type	CBS
<i>Hanseniaspora uvarum</i> CECT 11105	Wild type	Colección Española de Cultivos Tipo (CECT)
<i>Clostridium beijerinckii</i> NCIMB 8052	Wild type	(327)

## Plasmid construction

Table 2 lists all the plasmids used in this study. The pET26b:harmWanomala\_5543-His and pCUP1 plasmids were derived from Kruis *et al.* (2017). To construct pCOSCB3:WanEat1 the WT *W. anomalus* Eat1 sequence was amplified through PCR from the genomic DNA of *W.*

## Chapter 3

*anomalus* DSM 6766 and introduced downstream of the strong constitutive thiolase promoter derived from *C. beijerinckii* NCIMB 8052. Individual fragments containing the spectinomycin resistance gene, the pAMB1 ori and the ColE1 ori were also amplified through PCR and assembled through NEBuilder® HiFi DNA Assembly. For pCOSCB:EV, the WT *W. anomalus* Eat1 gene was omitted from the plasmid.

**Table 2. Plasmids used in this study.**

Plasmid	Host, relevant gene	Source
pCUP1:EV	<i>S. cerevisiae</i> , Empty vector	(35)
pCUP1:WanEat1	<i>S. cerevisiae</i> , <i>W. anomalus</i> Eat1	(35)
pCUP1:WciEat1	<i>S. cerevisiae</i> , <i>W. ciferii</i> Eat1	(35)
pCUP1:KmaEat1	<i>S. cerevisiae</i> , <i>K. marxianus</i> Eat1	(35)
pCUP1:KlaEat1	<i>S. cerevisiae</i> , <i>K. lactis</i> Eat1	(35)
pCUP1:CjaEat1	<i>S. cerevisiae</i> , <i>C. jadinii</i> Eat1	(35)
pCUP1:CfaEat1	<i>S. cerevisiae</i> , <i>C. fabianii</i> Eat1	(35)
pCUP1:HuvEat1	<i>S. cerevisiae</i> , <i>H. uvarum</i> Eat1	(35)
pCUP1:SceEat1	<i>S. cerevisiae</i> , <i>S. cerevisiae</i> Eat1	(35)
pET26b:EV	<i>E. coli</i> , Empty vector	Novagen
pET26b:harmWanomala_5543-His	<i>E. coli</i> , codon harmonized <i>W. anomalus</i> Eat1	(35)
pCOSCB:EV	<i>C. beijerinckii</i> , Empty vector	This study
pCOSCB3:WanEat1	<i>C. beijerinckii</i> , <i>W. anomalus</i> Eat1	This study

## Chemicals and reagents

Unless otherwise specified, all the chemical reagents were purchased from Sigma-Aldrich. 1-Octen-3-ol, 1-phenylethanol, 2-phenylethanol, cis-3-hexenol, citronellol, geraniol, linalool, trans-2-hexenol and their acetate esters were kindly provided by Axxence Aromatic GmbH, Germany.

## Eat1 expression and purification

To purify the WanEat1 protein, a previously established protocol was used (328). Briefly, *E. coli* BL21 (DE3) bearing pET26b:harmWanomala\_5543-His was grown in 1 L LB medium containing 100 mg L<sup>-1</sup> kanamycin at 37°C and shaken at 120 rpm. When an OD<sub>600</sub> of 0.5-0.6 was reached, the cultures were chilled on ice water for 15 min before they were induced with 0.2 mM IPTG. The cultures were then incubated overnight (~16 h) at 20°C and shaking at 120 rpm. After overnight incubation, the cells were harvested by centrifuging the culture for 15 min

at 6000  $\times$  g at 4°C. The cell pellet was washed with 50 ml of 50 mM potassium phosphate buffer (KPi, pH 7.5), centrifuged for 15 min at 6,000  $\times$  g at 4°C and the cell pellet was stored at -20°C until use.

For cell lysis, the cell pellet was resuspended in 20 ml of HA buffer (50 mM KPi, 300 mM NaCl, pH 8.0, 20 mM imidazole) containing 1 mini tablet cOmplete™ protease inhibitor for every 10 ml of HA buffer. Cells were disrupted by sonication using a VS 70 T tip (Bandelin SONOPLUS HD) using the following setup: 25% amplitude, 10 min total time and 1 sec ON/2 sec OFF. The cell lysate was centrifuged at 4°C at 30,000  $\times$  g for 45 min. The supernatant was collected and filtered through a 0.22  $\mu$ m membrane filter and then used for protein purification.

The cell-free extract was subjected to Q-sepharose Fast Performance Liquid (FPLC) purification through ÄKTA go Protein Purification System (GE Healthcare Life Sciences). The first purification step involved loading the cell-free extract on a 1-mL His Trap™ HP column (GE Healthcare Life Sciences) equilibrated with HA buffer and then eluted by washing the column with HB buffer (50 mM KPi, 300 mM NaCl, 500 mM imidazole, pH 8.0). The fractions containing the protein were collected, combined and diluted 5 times with CA buffer (50 mM KPi, pH 7.0). The diluted protein was loaded on a 1-ml HiTrap SP HP column (GE Healthcare Life Sciences) equilibrated with CA buffer and eluted by a NaCl gradient in CA buffer (50 mM KPi, pH 7.0, 1 M NaCl). The fractions containing the highest protein content were combined and used for further analysis.

SDS-PAGE was used to analyze the purity of the protein samples. The protein samples were denatured by heating at 98°C for 5 min in 4x Laemmli Sample Buffer (Bio-Rad) and centrifuged at 10,000 rpm for 30 sec. Proteins were then separated on Mini-PROTEAN® TGX™ precast gel (Bio-Rad) at 20 mA for 45 min and stained using Page Blue™ protein staining solution (Thermo Fisher Scientific).

### **Enzyme assays**

All enzyme assays were performed at 30°C in a phosphate buffer (50 mM KPi, 150 mM NaCl, pH 7.5).

The 4-nitrophenol release assay was performed by measuring the release of 4-nitrophenol from 0.1 mM 4-nitrophenyl acetate (diluted from a 200 mM stock in DMSO) at 405 nm using a temperature-controlled U-2010 spectrophotometer (Hitachi). The initial slope was determined as the absorption per min of 4-nitrophenol at 405 nm and the amount of 4-nitrophenol was

## Chapter 3

calculated from a calibration curve of 4-nitrophenol ( $r^2 \geq 0.99$ ) at 30°C in a phosphate buffer (50 mM KPi, 150 mM NaCl, pH 7.5).

To assess alcoholysis and thiolysis, the 4-nitrophenol release assay was performed in the presence of different concentrations of ethanol, 1-butanol, 1-hexanol, ethanethiol or butanethiol (0-80 mM). The reaction was initiated by adding purified WanEat1 protein (0.69  $\mu\text{g ml}^{-1}$  final concentration).

To analyze the reaction mixture for ester (alcoholysis) or thioester (thiolysis) formation during the 4-nitrophenol release assay, we repeated the assay as described above but we used 2.5 mM 4-nitrophenyl acetate and 2.5 mM of the alcohols or thiols. Every 10 sec and for a total of 3 min, 100  $\mu\text{l}$  of the reaction mixture was recovered and mixed with 100  $\mu\text{l}$  stop solution (0.1 N  $\text{H}_2\text{SO}_4$  with 10 mM acetone as internal standard) and 100  $\mu\text{l}$  n-hexane, vortexed vigorously and let the extraction to proceed for 15 min. The hexane layer was then analyzed by GC.

To assess whether the inhibition of Eat1 by alcohols or thiols was reversible, 11.04  $\mu\text{g ml}^{-1}$  purified WanEat1 protein was first briefly incubated with 80 mM ethanol, 1-butanol, 1-hexanol, ethanethiol or butanethiol before diluting to 5 mM final alcohol or thiol concentration. The final enzyme concentration was the same (0.69  $\mu\text{g ml}^{-1}$ ) as in the standard assay. The reaction was initiated by adding 0.1 mM 4-nitrophenyl acetate. The final reaction volume for the 4-nitrophenol release and inhibition assays was 1 ml.

*In vitro* alcoholysis and thiolysis assays were performed in triplicate in gas tight glass vials at 30°C in the presence of 13.98  $\mu\text{g ml}^{-1}$  purified protein. The total volume of the reaction was 250  $\mu\text{l}$  and was performed in a phosphate buffer (50 mM KPi, 150 mM NaCl, pH 7.5) supplemented with 2.5 mM of the ester and 2.5 mM of the alcohol or thiol. The reaction mixture was incubated at 30°C for 10 min. Then, the reaction was initiated by adding the enzyme to the reaction mixture and the reaction was terminated after 5 min of incubation by adding 250  $\mu\text{l}$  stop solution (0.1 N  $\text{H}_2\text{SO}_4$ ). 10 mM acetone was included in the stop solution as internal standard. Extraction of the mixture of substrates and products was done by adding 250  $\mu\text{l}$  n-hexane. The sample vial was vortexed vigorously and the extraction proceeded for 15 min. Following, the n-hexane layer was used for GC analysis.

## Analytical

For *in vitro* alcoholysis and thiolysis, esters, thioesters, alcohols or thiols recovered by n-hexane extraction were analyzed on a Shimadzu 2010 gas chromatograph equipped with an AOC 20i+s

### Eat1-like alcohol acyl transferases from yeasts have high alcoholysis and thiolysis activity

autosampler (Shimadzu). 1  $\mu\text{L}$  of the sample was injected on a DB-WAX UI column (30 m length, 0.53 mm inner diameter, 1  $\mu\text{m}$  film thickness, Agilent) with a split ratio 1:20. The column temperature was kept at 70°C for 1 min followed by an increase to 125°C at a rate of 50°C  $\text{min}^{-1}$  and then followed by an increase to 230°C at a rate of 50°C  $\text{min}^{-1}$  where the temperature was kept at 230°C for 1 min.

*In vivo* alcoholysis by WT yeasts or by transformed *S. cerevisiae* was analyzed by taking a 200  $\mu\text{L}$ -sample of the yeast culture followed by centrifugation at 15000 rpm for 5 min to remove any cells. 100  $\mu\text{L}$  of the supernatant was mixed with 100  $\mu\text{L}$  of MQ water containing 10 mM acetone as internal standard in a 10-ml gas tight vial. The final solution was analyzed on a Shimadzu 2010 gas chromatograph equipped with an HS-20 autosampler (Shimadzu). The sample vials were heated at 60°C for 6 min to allow evaporation of the esters and alcohols to the headspace of the vial. 1 ml of the headspace was then injected on a DB-WAX UI column (30 m length, 0.53 mm inner diameter, 1  $\mu\text{m}$  film thickness, Agilent) with a split ratio 1:20. The column temperature was kept at 50°C for 1 min followed by an increase to 90°C at a rate of 5°C  $\text{min}^{-1}$  followed by an increase to 230°C at a rate of 230°C  $\text{min}^{-1}$  where the temperature was kept at 230°C for 1 min.

Cultures of transformed *C. beijerinckii* (pCOSCB3:EV or pCOSCB3:WanEat1) were analyzed for *in vivo* alcoholysis by sampling 200  $\mu\text{L}$  of the anaerobic culture followed by centrifugation at 15000 rpm for 5 min. 100  $\mu\text{L}$  of the supernatant was mixed with 100  $\mu\text{L}$  of MQ water containing 10 mM acetonitrile as internal standard in a 10-ml gas tight vial. Samples were analyzed as described above. The column temperature was kept at 70°C for 1 min followed by an increase to 125°C at a rate of 50°C  $\text{min}^{-1}$  followed by an increase to 230°C at a rate of 70°C  $\text{min}^{-1}$  where the temperature was kept at 230°C for 1 min.

All the data presented in this study are averages of biological triplicates and the standard deviation is presented as the error bar in all figures.

### Eat1 homolog accession numbers

The accession numbers of the Eat1 homologs used in this study are: *W. anomalus* Eat1 (XP\_019041020.1), *W. ciferrii* Eat1 (XP\_011273049.1), *K. marxianus* Eat1 (KMAR\_10772), *K. lactis* Eat1 (KLLA0\_E24421g), *C. jadinii* Eat1 (CEP25158.1), *C. fabianii* Eat1 (CDR40574.1), *H. uvarum* Eat1 (D499\_0A01740), *S. cerevisiae* Eat1 (YGR015C).

## Results

### Eat1 accelerates 4-nitrophenol release through alcoholysis

The recently discovered  $\alpha/\beta$ -hydrolase Eat1 exhibits three activities: AAT, esterase and thioesterase (35). Kruis *et al.* (2017) showed that the esterase and thioesterase activities are decreased by high concentrations of ethanol. This could be due to either inactivation of the enzyme because of unfolding or preference of the enzyme for alcohols over water as acyl acceptors. Together with the ability of the enzyme to use esters and thioesters as substrates, we hypothesized that Eat1 should also be capable of catalyzing alcoholysis (and thiolysis). To demonstrate the predicted alcoholysis, we set-up a classic esterase assay using 4-nitrophenyl acetate and monitored the release of 4-nitrophenol by Eat1 in the absence or presence of various concentrations (0-80 mM) of ethanol, 1-butanol or 1-hexanol (Fig. 3A). This method enabled us also to investigate the effect of high alcohol concentrations on the Eat1 activities.

In the absence of alcohols, Eat1 catalyzed the release of 4-nitrophenol with a specific activity of  $9.52 \pm 0.49 \text{ U mg}^{-1}$ , representing its esterase activity. Interestingly, the release of 4-nitrophenol increased dramatically when ethanol, 1-butanol or 1-hexanol were added to the reaction mixture. A maximum specific activity was reached with 10 mM ethanol ( $124.30 \pm 2.54 \text{ U mg}^{-1}$ ), 5 mM 1-butanol ( $100.49 \pm 0.27 \text{ U mg}^{-1}$ ) or 2.5 mM 1-hexanol ( $60.25 \pm 0.88 \text{ U mg}^{-1}$ ).

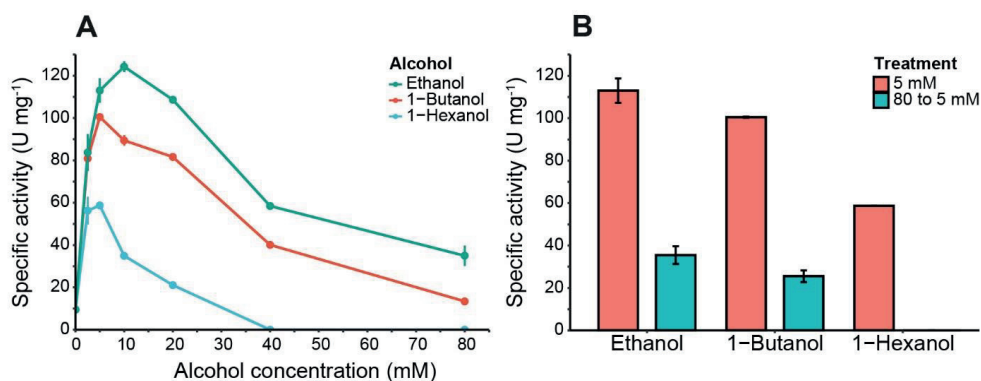
The observed acceleration of the 4-nitrophenol release confirms our assumption that Eat1 has a preference for alcohols as acyl acceptors. As such, Eat1 performs alcoholysis instead of hydrolysis in the presence of alcohols. To confirm alcoholysis over hydrolysis and to exclude non-specific acceleration of 4-nitrophenol release, we incubated 2.5 mM of 4-nitrophenyl acetate with 2.5 mM of ethanol in the presence of  $0.0138 \text{ mg ml}^{-1}$  Eat1 and analyzed the products by GC. As expected, ethyl acetate was the major product when 4-nitrophenyl acetate and ethanol were used as substrates (Fig. S1). Also, alcoholysis dominated over hydrolysis as ethyl acetate reached the maximum conversion of 2.5 mM. Similarly, butyl acetate and hexyl acetate were detected by GC (data not shown). These esters were not detected in the control samples that did not contain alcohols or where the Eat1 enzyme was omitted from the assay.

### High alcohol concentration irreversibly inhibits Eat1

Whereas a low alcohol concentration increased the release of 4-nitrophenol, higher alcohol concentrations ( $>10 \text{ mM}$ ) caused a decrease in the specific activity of Eat1 for all three tested alcohols (Fig. 3A). This reduction was more apparent at a concentration of 80 mM for ethanol

and 1-butanol where the specific activity for 4-nitrophenol release was  $23.95 \pm 4.81 \text{ U mg}^{-1}$  and  $13.41 \pm 0.84 \text{ U mg}^{-1}$ , respectively. For 1-hexanol, the release of 4-nitrophenol was abolished at 40 mM or higher.

The observed inhibition of 4-nitrophenol release with increased alcohol concentrations suggested that the alcohols irreversibly inhibit Eat1. To assess this, we repeated our previous experiment for 4-nitrophenol release, but we first incubated Eat1 with 80 mM of ethanol, 1-butanol or 1-hexanol and immediately diluted it to a final alcohol concentration of 5 mM, before starting the assay. All other constituents of the assay mixture were the same as in the standard assay. Our results demonstrate that the Eat1 enzyme was irreversibly inhibited by alcohols as it was not able to recover its activity after exposure to high alcohol concentrations (Fig. 3B).



**Figure 3. 4-nitrophenol release and Eat1 inhibition assays in the presence of different alcohols.** (A) 4-nitrophenol release by Eat1 was measured colorimetrically using 100  $\mu\text{M}$  4-nitrophenyl acetate and different concentration (0–80 mM) of ethanol (blue line), 1-butanol (red line) or 1-hexanol (green line). (B) Inhibition of Eat1 was assessed by exposing Eat1 to 80 mM of ethanol, 1-butanol or 1-hexanol and immediately diluting it to 5 mM final concentration (blue bars). The release of 4-nitrophenol was monitored and compared with an assay where pre-treatment of Eat1 with 80 mM alcohols was omitted (red bars). Error bars indicate the standard deviation. 1 U = 1  $\mu\text{mol min}^{-1}$ .

### Eat1 catalyzes alcohololysis with various alcohols in an aqueous environment

Our previous results confirmed alcohololysis by Eat1 in the presence of 4-nitrophenyl acetate and alcohols in an aqueous environment. As Eat1 is present in various ester-producing yeasts, it is important to determine the alcohol (and acyl) specificity of Eat1. Whereas 4-nitrophenyl acetate is an excellent compound for quantifying alcohololysis, it is not a physiologically relevant compound. For this reason, we replaced 4-nitrophenyl acetate with ethyl acetate (the main ester found in yeasts having Eat1) as the acyl donor in our assays and determined the alcohol



## Chapter 3

specificity of Eat1 using various primary, secondary or tertiary alcohols (C1-10) as acyl acceptors (Table 3).

**Table 3. Specific activity of Eat1 towards various primary, secondary and tertiary alcohols during alcoholysis with ethyl acetate as the acyl donor.**

Alcohol carbon length	Alcohol name	XLogP3	Specific activity (U mg <sup>-1</sup> )
<b>Primary</b>			
3	1-Propanol	0.30	7.56 ± 1.27
4	1-Butanol	0.90	4.34 ± 1.76
5	1-Pentanol	1.60	1.51 ± 0.72
6	1-Hexanol	2.00	1.20 ± 0.55
6	Cis-3-Hexenol	1.30	2.14 ± 0.32
6	Trans-2-Hexenol	1.40	1.81 ± 0.70
8	2-Phenylethanol	1.40	2.19 ± 1.04
8	1-Octanol	3.00	0.35 ± 0.08
9	1-Nonanol	4.30	0.10 ± 0.02
10	1-Decanol	4.60	0.18 ± 0.02
10	Citronellol	3.20	0.36 ± 0.02
10	Geraniol	2.90	0.43 ± 0.01
<b>Secondary</b>			
3	2-Propanol	0.30	0.75 ± 0.31
5	2-Pentanol	1.20	0.26 ± 0.07
8	1-Phenylethanol	1.40	0.45 ± 0.07
8	1-Octen-3-ol	2.60	ND
9	3-Methyloctan-4-ol	3.10	ND
10	Carveol	2.10	ND
<b>Tertiary</b>			
5	2-Methyl-3-butyn-2-ol	0.30	ND
9	2-phenyl-2-propanol	1.80	ND
10	Linalool	2.70	ND

ND, Not detected; 1 U=1 μmol min<sup>-1</sup>; ± indicates the standard deviation; The XLogP3 value was derived from <https://pubchem.ncbi.nlm.nih.gov> and indicates the predicted octanol-water partition coefficient (logP) as described by Cheng T. et al (2007) (329).

Eat1 was rather promiscuous with respect to the alcohols. Alcoholysis was observed using various primary and secondary alcohols. 1-Propanol was the best short chain primary alcohol acceptor with a specific activity of 7.56 ± 1.27 U mg<sup>-1</sup> whereas longer chain primary alcohols, such as 1-nonanol and 1-decanol, showed a specific activity of only 0.10 ± 0.02 and 0.18 ± 0.02 U mg<sup>-1</sup>, respectively. More complex primary alcohols, such as 2-phenylethanol (2.19 ± 1.04 U mg<sup>-1</sup>), citronellol (0.36 ± 0.02 U mg<sup>-1</sup>) and geraniol (0.43 ± 0.01 U mg<sup>-1</sup>), were also accepted by Eat1 even with higher rate than some simple primary alcohols.

The short chain secondary alcohols 2-propanol (0.75 ± 0.31 U mg<sup>-1</sup>), 2-pentanol (0.26 ± 0.07 U mg<sup>-1</sup>) and 1-phenylethanol (0.45 ± 0.07 U mg<sup>-1</sup>) were accepted during alcoholysis by Eat1

but longer complex secondary alcohols such as 1-octen-3-ol, 3-methyloctan-4-ol and carveol did not show any activity with Eat1. No activity was detected with any of the tertiary alcohols tested.

After determining the substrate specificity of Eat1 towards alcohols using ethyl acetate as the acyl donor, we further investigated the acyl specificity of the alcoholysis reaction (Table 4). Ethyl acetate, ethyl butyrate, ethyl valerate or ethyl hexanoate were used as the acyl donors and 1-butanol was used as the acyl acceptor. Ethyl acetate was the preferred acyl donor for Eat1 as a higher specific activity ( $4.34 \pm 1.76 \text{ U mg}^{-1}$ ) was observed compared to ethyl butyrate ( $1.70 \pm 0.43 \text{ U mg}^{-1}$ ). No alcoholysis activity was found using ethyl valerate and ethyl hexanoate.

**Table 4. Acyl specificity of Eat1 during alcoholysis.**

Acyl carbon length	Ethyl ester donor	Butyl ester product	Specific activity ( $\text{U mg}^{-1}$ )
2	Ethyl acetate	Butyl acetate	$4.34 \pm 1.76$
4	Ethyl butyrate	Butyl butyrate	$1.70 \pm 0.43$
5	Ethyl valerate	Butyl valerate	ND
6	Ethyl hexanoate	Butyl hexanoate	ND

ND, Not detected. 1 U =  $1 \mu\text{mol min}^{-1}$ ;  $\pm$  indicates the standard deviation

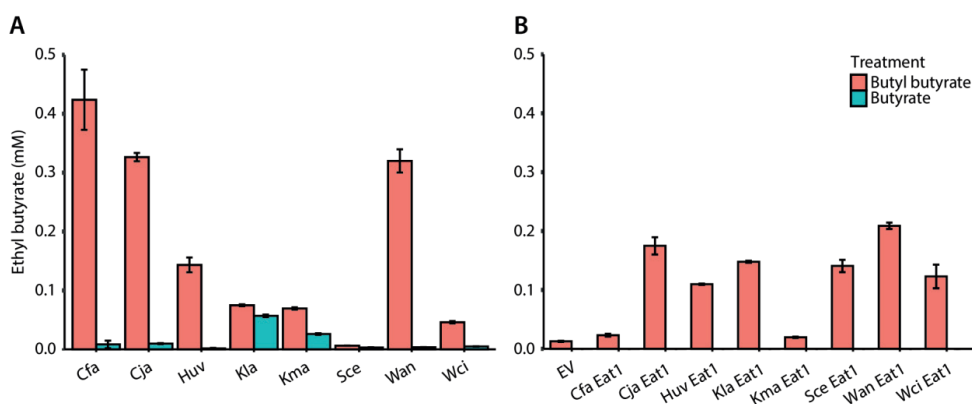
### ***In vivo* alcoholysis in WT and engineered yeasts**

Alcoholysis was previously demonstrated in cell free extract and *in vivo* in certain lactic acid bacteria (143,144,330,331). Since Eat1 homologs are present in various yeasts, we assessed the capacity of such yeasts to perform *in vivo* alcoholysis. We selected *S. cerevisiae* (control) and several Crabtree negative yeasts and grew them in YMM-no-iron in the presence of 5 mM butyl butyrate. *S. cerevisiae* was selected as a control because it is Crabtree positive and therefore does not follow the hypothesis of acetyl-CoA accumulation into the mitochondria and overexpression of Eat1 during iron starvation. Our selection of Crabtree negative yeasts and the indicated growth conditions are based on the following: i) the selected yeasts have active homologs of Eat1, ii) Crabtree negative yeasts show enhanced ethyl acetate production during iron limitation which is hypothesized to be correlated with acetyl-CoA accumulation and Eat1 overexpression in the mitochondria, iii) butyl butyrate, butanol and butyrate are not produced by the selected yeasts and thus alcoholysis between the supplemented butyl butyrate and the endogenously produced ethanol can be easily screened through the production of ethyl butyrate (35,117,118). As a control for potential esterification between free butyric acid and ethanol, we replaced butyl butyrate in the growth medium with 5 mM butyrate.

Ethyl butyrate was produced by all tested yeasts in cultures containing butyl butyrate (Fig. 4A). *C. fabianii*, *C. jadinii* and *W. anomalus* showed the highest production of ethyl butyrate ( $0.42 \pm 0.05$ ,  $0.33 \pm 0.01$  and  $0.32 \pm 0.02$  mM, respectively) whereas *K. lactis*, *K. marxianus* and *W. ciferrii* produced very little ethyl butyrate ( $< 0.1$  mM). *H. uvarum* was an intermediate ethyl butyrate producer ( $0.14 \pm 0.01$  mM) whereas the *S. cerevisiae* control produced only traces of ethyl butyrate. Yeast cultures containing free butyrate, instead of butyl butyrate, showed low ethyl butyrate production ( $< 0.06$  mM).

Our results indicate that yeast species are capable of *in vivo* alcoholysis, but that the capacity varies amongst yeast species. This might be correlated to the expression level of Eat1 under the defined conditions or the ability of the different Eat1 homologs (or other enzymes) to perform alcoholysis in the first place. To examine whether different Eat1 homologs have a different alcoholysis capacity *in vivo*, we developed a more robust assay where we overexpressed these homologs in *S. cerevisiae* under the inducible CUP1 promoter. These recombinant *S. cerevisiae* strains were grown in the presence of 5 mM butyl butyrate to assess alcoholysis or in the presence of 5 mM butyrate to assess esterification.

Ethyl butyrate was produced in all *S. cerevisiae* variants (Fig. 4B). However, Cja-Eat1, Huv-Eat1, Kla-Eat1, Sce-Eat1, Wan-Eat1 and Wci-Eat1 produced higher ethyl butyrate levels compared to the other homologs and the empty vector control. Wan-Eat1 was the best ethyl butyrate producing enzyme reaching  $0.21 \pm 0.01$  mM. Surprisingly, Cfa-Eat1 did not show high ethyl butyrate production through alcoholysis despite the observed alcoholysis in the WT yeast assays. Furthermore, Kla-Eat1, Sce-Eat1 and Wci-Eat1 could perform *in vivo* alcoholysis when overexpressed in *S. cerevisiae* even though the WT yeasts did not show this capacity. Control cultures containing butyrate did not produce ethyl butyrate through esterification.



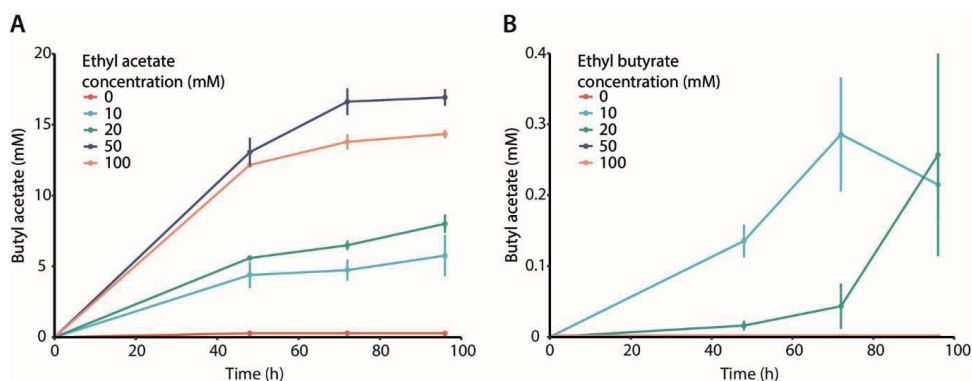
**Figure 4.** *In vivo* alcoholysis in wild type yeasts or in engineered *S. cerevisiae*. **(A)** *In vivo* alcoholysis by WT yeasts grown in the presence of 5 mM butyl butyrate (red bars) or 5 mM butyrate (blue bars) for 48 h. n = 3. **(B)** Screening of different Eat1 homologs for their capacity for alcoholysis when expressed in *S. cerevisiae* grown for 24 h in the presence of 5 mM butyl butyrate (red bars) or 5 mM butyrate (blue bars). n = 3. Abbreviations: EV – Empty vector; Cfa- *Cyberlindnera fabianii*; Cja - *Cyberlindnera jadinii*; Huv - *Hanseniaspora uvarum*; Kla – *Kluyveromyces lactis*; Kma - *Kluyveromyces marxianus*; Sce - *Saccharomyces cerevisiae*; Wan - *Wickerhamomyces anomalus*; Wci - *Wickerhamomyces ciferrii*. The error bars indicate standard deviations.

### Ester upgrade through *in vivo* alcoholysis in *Clostridium beijerinckii*

The ability of Eat1 to perform *in vivo* alcoholysis may provide an opportunity to upgrade low value esters with high value alcohol moieties *in vivo* and therefore increase their commercial value. To realize this, we chose *C. beijerinckii* NCIMB 8052 as the appropriate host for ester upgrading since it is a natural producer of butanol and thus a good candidate to produce butyl esters through alcoholysis.

Batch cultures of *C. beijerinckii* transformed either with an empty plasmid (pCOSCB:EV) or with a plasmid constitutively expressing WanEat1 (pCOSCB3:WanEat1) were grown in the presence of various concentration of ethyl acetate (0-100 mM). Without the supplementation of ethyl acetate, the Eat1 expressing cultures produced only traces of butyl acetate (Fig 5A). However, high levels of butyl acetate was produced by all cultures containing ethyl acetate and expressing Eat1. The highest production of butyl acetate ( $16.92 \pm 0.58$  mM) was reached when 50 mM ethyl acetate was supplemented in the medium after 96 h of fermentation. 100 mM ethyl acetate seemed to have a toxic effect on *C. beijerinckii* as butyl acetate production decreased. Empty vector control cultures did not show butyl acetate production when ethyl acetate was supplied in the growth medium (data not shown).

In addition to acetate esters, Eat1 can also catalyze alcoholysis with butyrate esters. To assess whether we can upgrade even more complex ethyl esters to butyl esters, we performed a similar experiment as reported above, but supplied ethyl butyrate as the acyl donor at different concentrations (0-100 mM). Butyl butyrate was produced in cultures expressing Eat1 (Fig. 5B). Maximum production of  $0.29 \pm 0.08$  mM was reached in cultures supplemented with 10 mM ethyl butyrate and grown for 72 h. *C. beijerinckii* could not grow in cultures containing ethyl butyrate concentrations equal to or higher than 50 mM. Obviously, butyl butyrate was not detected in those cultures, nor in cultures bearing the empty plasmid.



**Figure 5. *In vivo* alcoholysis in engineered *C. beijerinckii*.** (A) Production of butyl acetate by *C. beijerinckii* expressing WT WanEat1 and grown in 0 (red), 10 (light blue), 20 (green), 50 (dark blue) or 100 mM (orange) ethyl acetate. (B) Production of butyl butyrate by *C. beijerinckii* expressing WT WanEat1 and grown in 0 (red), 10 (light blue), 20 (green), 50 (dark blue) or 100 mM (orange) ethyl butyrate. Error bars indicate the standard deviation.

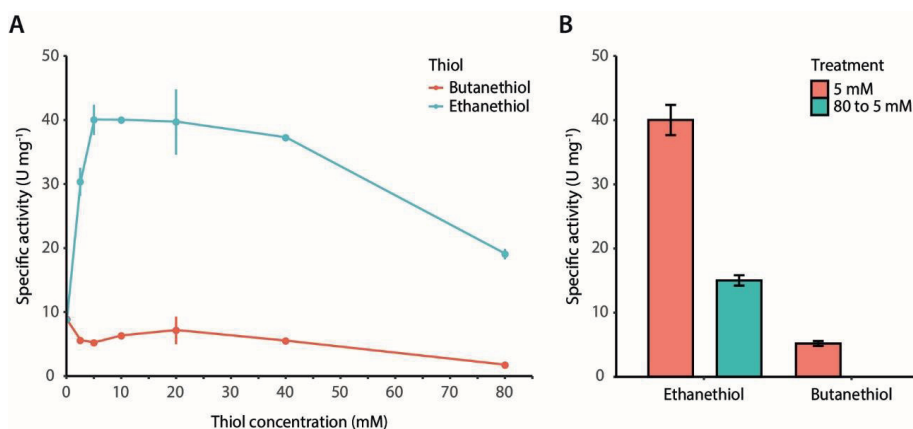
### Eat1 can also perform thiolysis

Eat1 can accept esters and thioesters as acyl donors. This triggered us to investigate the capacity of Eat1 to accept thiols as acyl acceptors and therefore perform thiolysis. To assess this, we used the 4-nitrophenol release assay developed for alcohols, but instead we replaced the alcohols with either ethanethiol or butanethiol.

The release of 4-nitrophenol from 4-nitrophenyl acetate was accelerated in the presence of ethanethiol (Fig. 6A). The production of ethyl thioacetate was confirmed by GC (data not shown). This clearly indicates that Eat1 is also capable of catalyzing thiolysis (Fig. 6A). The release of 4-nitrophenol in the presence of increasing concentrations of ethanethiol followed a similar trend as observed for alcoholysis (Fig. 3A); 4-nitrophenol release is enhanced ( $40.04 \pm$

### Eat1-like alcohol acyl transferases from yeasts have high alcoholysis and thiolysis activity

2.36 U mg<sup>-1</sup>) at low ethanethiol concentrations (5 mM) but it is decreased ( $19.06 \pm 0.80$  U mg<sup>-1</sup>) when a high (80 mM) concentration is used. Using a similar setup as used in the alcoholysis assays, we could show that the lower thiolysis activity at higher ethanethiol concentrations was a result of the irreversible inhibition of Eat1 by ethanethiol (Fig. 6B). In contrast to ethanethiol, butanethiol did not accelerate 4-nitrophenol release when it was added to the reaction medium. On the other hand, it slightly decreased the activity of Eat1 towards 4-nitrophenyl acetate compared to its esterase activity.



**Figure 6. 4-nitrophenol release assay and Eat1 inhibition in the presence of thiols.** (A) Effect of 4-nitrophenol release from 4-nitrophenyl acetate incubated with various concentrations (0–80 mM) of ethanethiol or butanethiol. (B) Inhibition assay of Eat1 by exposing Eat1 to 80 mM ethanethiol or butanethiol followed by immediate dilution to a final concentration of 5 mM. Error bars indicate the standard deviation. 1 U=1  $\mu$ mol min<sup>-1</sup>.

**Table 5. Comparison between thiolysis and alcoholysis.**

Acyl Donor	Acyl Acceptor	Product	Specific activity (U mg <sup>-1</sup> )
Ethyl acetate	Ethanethiol	Ethyl thioacetate	0.08 $\pm$ 0.01
Ethyl thioacetate	Ethanol	Ethyl acetate	6.27 $\pm$ 0.86

1 U=1  $\mu$ mol min<sup>-1</sup>;  $\pm$  indicates the standard deviation

For completeness, we performed *in vitro* thiolysis between ethyl acetate as the acyl donor and ethanethiol as the acyl acceptor (Table 5). Also, for comparison, we performed alcoholysis between ethyl thioacetate as the acyl donor and ethanol as the acyl acceptor. Ethyl thioacetate was produced through thiolysis and ethyl acetate was produced through alcoholysis. However, the specific activity to produce these compounds differed based on the acyl donor. When ethyl acetate was the acyl donor, a very low specific activity was observed with ethanethiol ( $0.08 \pm$

0.01) as the acyl acceptor. In contrast, the specific activity of Eat1 increased when ethyl thioacetate was used as the acyl donor and ethanol as the acceptor ( $6.27 \pm 0.86$ ).

### Comparison of Eat1 catalytic activities

Previously, the Eat1 enzyme was reported to exhibit 3 types of catalytic activities *viz.* AAT, esterase, and thioesterase activity (35). In Table 6, we compare the novel alcoholysis and thiolysis activities to the previously reported Eat1 activities. The specific alcoholysis activity is ~450-fold higher than the AAT activity when ethyl acetate or acetyl-CoA was used as the acyl donor, respectively. Alcoholysis is also the highest activity when 4-nitrophenyl acetate was used as the acyl donor, showing the clear preference of Eat1 to alcohols over thiols and water (alcohol>thiol>water). The thiolysis activity between ethyl acetate and ethanethiol, although ~5-fold higher than the AAT activity, is ~95-fold lower than the alcoholysis activity between ethyl acetate and 1-propanol.

**Table 6. Comparison between the different activities of Eat1.**

Activity	Acyl donor	Acyl acceptor	Specific activity (U mg <sup>-1</sup> ); (fold difference compared to AAT)	Source
AAT	Acetyl-CoA	Ethanol	0.017; (1)	Kruis <i>et al.</i> (2017)
Thioesterase	Acetyl-CoA	H <sub>2</sub> O	0.032; (1.88)	Kruis <i>et al.</i> (2017)
Esterase	Ethyl acetate	H <sub>2</sub> O	0.85; (50)	Kruis <i>et al.</i> (2017)
Esterase	4-nitrophenyl acetate	H <sub>2</sub> O	10.7; (629)	This study
Alcoholysis	4-nitrophenyl acetate	Ethanol	112; (6588)	This study
Alcoholysis	Ethyl acetate	1-Propanol	7.56; (445)	This study
Thiolysis	4-nitrophenyl acetate	Ethanethiol	40.04; (2355)	This study
Thiolysis	Ethyl acetate	Ethanethiol	0.08; (4.7)	This study

1 U=1  $\mu\text{mol min}^{-1}$

## Discussion

In this study we show that the  $\alpha/\beta$ -hydrolase Eat1 can catalyze both alcoholysis and thiolysis reactions. This discovery extends the catalytic portfolio of Eat1 to five different activities: AAT, esterase, thioesterase, alcoholysis and thiolysis (Fig. 1). Although all five activities most likely use the same canonical esterase mechanism, the acyl donor and the acyl acceptor affinity and specificity defines which of the activities prevails. For example, *in vitro*, AAT is dominant over the esterase and the thioesterase activities as Eat1 favors ethanol over water in an aqueous environment (35). The preference of Eat1 for alcohols and thiols also explains that alcoholysis and thiolysis prevail over hydrolysis in an aqueous environment. This was apparent when we tested the release of 4-nitrophenol from 4-nitrophenyl acetate by using Eat1 in an aqueous environment. In the absence of alcohols or thiols, the rate of hydrolysis is the only determinant for the half-life of the acyl-enzyme intermediate (332). However, when alcohols or thiols were present, Eat1 preferred the alcohol or thiol nucleophile over water. This resulted in a shorter half-life of the acyl-enzyme intermediate that in turn resulted in an overall faster release of 4-nitrophenol.

In the presence of high alcohol or thiol concentrations (>10 mM), Eat1 reduced the release of 4-nitrophenol from 4-nitrophenyl acetate compared to lower concentrations (<10 mM). A similar trend was also observed by other studies when different acyltransferases were tested for their capacity to catalyze alcoholysis or transesterification (332-335). The reduced activity of Eat1 by high alcohol or thiol concentrations suggested that substrate inhibition occurs. Our results, clearly demonstrate that Eat1 was irreversibly inhibited *in vitro* when exposed to high concentrations of either alcohols or thiols. The irreversible ping-pong bi-bi mechanism may explain our observations (336-338). According to this mechanism, the alcohol acts as an inhibitor of the enzyme by binding to the free enzyme thus forming ineffective, dead-end complexes. Our results obey this mechanism as Eat1 cannot recover its activity even after diluting the alcohol or thiol concentration. However, unfolding of Eat1 at high alcohol concentration should not be excluded as a possible explanation for our observations. Nevertheless, inhibition of Eat1 at 80 mM of ethanol is surprising as the ethanol production by the yeasts *K. marxianus* and *K. lactis* may exceed 100 mM (117). Still, Eat1 has been shown to be the main enzyme that contributes to high ethyl acetate production in both yeasts (117,191). Compartmentalization of Eat1 to the yeast mitochondria and low ethanol concentration inside the mitochondria may assist in preventing the inhibition of Eat1 by ethanol *in vivo*.



### Chapter 3

The novel alcoholysis activity of Eat1 described here, exceeds the former activities (AAT, esterase and thioesterase) by far. The specific alcoholysis activity is ~450-fold higher than the AAT activity and ~9-fold higher than the esterase activity (Table 6). Although these activities depend on the assay conditions, the enormous differences between the activities detected *in vitro* strongly suggest that alcoholysis does play a role *in vivo* as well. Because of the high alcoholysis activity, ethyl acetate may even outcompete acetyl-CoA as acyl donor, resulting in the recycling of ethyl acetate instead of the *de novo* synthesis from acetyl-CoA. In other words, ethyl acetate may occupy the active site and as such it would prevent access of acetyl-CoA. To avoid such a futile cycle of ethyl acetate synthesis, gas stripping can be applied to remove ethyl acetate from the medium as much as possible and allow access of acetyl-CoA to the active site of Eat1. The effect of alcoholysis may have played a role in a recent study in which Eat1 was used as an AAT to produce ethyl acetate in *E. coli* (339). In this previous study, it was reasoned that gas stripping would enhance ethyl acetate production by preventing its hydrolysis by Eat1. However, the alcoholysis activity of Eat1 was not known at that time. Hence, in retrospect, the insights gained in the present study indicate that gas stripping not only resulted in enhanced ethyl acetate production by avoiding hydrolysis, but most importantly, by avoiding alcoholysis. In total 42.8 mM of ethyl acetate was produced, which is the highest ethyl acetate production in an engineered microorganism to date (339).

Eat1 is able to accept a broad range (C3 to C10) of primary and secondary alcohols, although with different specific activities. One explanation of the observed differences may be the variable solubility of each alcohol in water since their XLogP3 values follow a comparable pattern as the specific activity by Eat1 during alcoholysis (Fig. S2). Nevertheless, this may explain only the trends observed for primary alcohols, as such a correlation did not apply to secondary alcohols. Therefore, especially for the secondary alcohols, the difference in specific activity is likely due to steric hindrance of the alcohol moieties in the Eat1 substrate-binding pocket (340,341). The interplay between the catalytic pocket and the substrates of Eat1 can only be resolved with the crystal structure of Eat1. Unfortunately, we have not yet succeeded in obtaining suitable Eat1 crystals.

With respect to acyl specificity, Eat1 can accept only short acyl chains (C2 and C4) and shows no activity towards longer acyl chains (C5 and C6). The wide alcohol and the narrow acyl specificity of Eat1, differentiates it from the yeast Eht1 and Eeb1  $\alpha/\beta$ -hydrolases. Eht1 and Eeb1 also exhibit AAT activity but their acyl specificity is broad (C4 to C12). Similar to Eat1, Eht1 and Eeb1 also have hydrolytic activities (176). Since Eht1 and Eeb1 can hydrolyze and produce

esters following the same canonical esterase mechanism as Eat1, it can also be expected that Eht1 and Eeb1 can perform alcoholysis. The extent to which activity prevails depends on the substrate specificity and availability and on the hydrophobicity of the catalytic pocket of the enzyme (342,343). Future experiments should include Eht1 and Eeb1 as potential enzymes for alcoholysis.

It was not until recently that the Eat1 enzyme was included in the list of AATs responsible for the production of esters in yeasts, including *S. cerevisiae*, *W. anomalus*, *K. marxianus*, *K. lactis* and others. Although Eat1 was demonstrated as an important AAT *in vivo*, its role for *in vivo* alcoholysis was not known. Alcoholysis, if present, can play an important role in the production and distribution of esters in yeasts. Such a role has previously been described for lactic acid bacteria where alcoholysis is the main mechanism for the production of esters (143,144,330). Similarly, we hypothesize that yeasts expressing Eat1 homologs may perform alcoholysis *in vivo*. As shown here, *C. fabianii*, *C. jadinii* and *W. anomalus* could indeed perform *in vivo* alcoholysis to recombine supplemented butyl butyrate and endogenously produced ethanol, under iron-limited conditions. Surprisingly, *K. lactis* and *K. marxianus* did not show *in vivo* alcoholysis even though their Eat1 homolog is the main contributor for the production of ethyl acetate (35,191). Although Eat1 was highly expressed in *W. anomalus* under iron-limited conditions, we do not have transcriptome data to support high expression of Eat1 for all the yeasts tested in this study (35). Therefore, it is unclear whether the Eat1 homologs are better expressed under iron-limited conditions in all yeasts even though ethyl acetate production is boosted in iron-limited conditions (117). In addition, the specificity of each Eat1 homolog towards different acids and alcohols is not known, and hence, the availability, accessibility and processing of esters and alcohols may differ between yeast species.

When we overexpressed Eat1 homologs in *S. cerevisiae*, several of the transformed strains were also able to catalyze *in vivo* alcoholysis. In the case of *C. jadinii*, *H. uvarum* and *W. anomalus* Eat1, our results agree with the observed alcoholysis in the WT yeasts, which may suggest that the Eat1 homolog is the main enzyme for alcoholysis. However, at present, no firm conclusion can be drawn since Eat1 knock-out strains could not be obtained. Whilst WT *K. lactis*, *S. cerevisiae* and *W. ciferrii* did not show alcoholysis, overexpression of their respective Eat1 homolog in *S. cerevisiae* did result in alcoholysis. This observation further strengthens the hypothesis that other factors affect the alcoholysis capacity of these yeasts such as the expression levels of Eat1 during the defined growth conditions, the availability of the substrate and the metabolic features of the different yeast species. Altogether, these *in vivo* experiments

show that various Eat1 homologs, other than the Eat1 tested in this study (*W. anomalus*), are capable of alcoholysis, but that the expression level and the specificity determines the final alcoholysis capacity of the corresponding strains. Our results further show that ester production in yeasts is not simply the result of AAT activity of Eat1, Eht1, Eeb1 and Atf1, but also of the alcoholysis activity of Eat1. This observation may revolutionize the food and beverage industry as the production of esters through alcoholysis has been completely neglected. The supplementation of short chain triglycerides (e.g. tributyrin) in the growth medium may provide the acyl substrate for the Eat1 enzyme and the subsequent production of novel esters by yeasts. We foresee that this discovery will enable food and beverage producers to further innovate with their yeast strains and their substrates in their fermentation processes and to produce novel food and beverage products.

As an alternative to current unsustainable processes, various AATs combined with extensive metabolic engineering have been used to produce esters in microorganisms (34). In our study, we applied a different approach for *in vivo* ester production by taking advantage of the alcoholysis activity of Eat1. We showed that engineered *C. beijerinckii* strains overexpressing Eat1 could upgrade ethyl acetate or ethyl butyrate to butyl acetate or butyl butyrate, respectively, through alcoholysis. The observed difference between the production level of butyl acetate and butyl butyrate is probably due to the acyl specificity of Eat1, as it has a higher specific activity with acetate esters than with butyrate esters (Table 2). In addition, we hereby show that the availability of the acyl donor (i.e. acyl-CoA) in *C. beijerinckii* is the main factor that controls the production of esters. Strains without the supplementation of ethyl esters produced only traces of butyl esters through AAT, indicating that the acyl-CoA availability is limiting for the AAT reaction. This observation was also made in a recent study where the availability of acyl-CoAs in *C. saccharoperbutylacetonicum* was the limiting factor for ester production (32). Our data show that *in vivo* ester upgrade through alcoholysis provides an alternative to traditional ester production methods by avoiding the limitation of the acyl-CoA availability. Such an approach may open new research lines for the production of esters using microorganisms.

Finally, we discovered that Eat1 also shows thiolysis activity. Thiolysis is known to be catalyzed by lipases from *R. miehei* (Lipozyme IM 20<sup>®</sup>), *C. antarctica* (Novozym 435<sup>®</sup>) or *Thermomyces lanuginosus* for the production of thiobutyl butyrate, thiobutyl valerate, thiostearyl palmitate, thiohexyl octanoate and other acetate thioesters (323,324,344-346). To our knowledge, this is the first time that an esterase was shown to catalyze thiolysis of short

chain thioesters. As thiols, like cysteine and glutathione, are present in yeasts, this activity may also be of physiological relevance or play a role in the production of thiol-based aromas. Further research is needed to assess the importance of thiolysis by Eat1 and its homologs for the *in vitro* and *in vivo* production of thioesters.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Author contributions

CP, RAW and SWMK conceived and designed the experiments. CP, LL and IC performed the experiments. MCRF and JVdO gave advice for the experimental design. CP, RAW and SWMK wrote the manuscript. All the authors have reviewed and approved the present manuscript. All the authors have contributed significantly to realize this work.

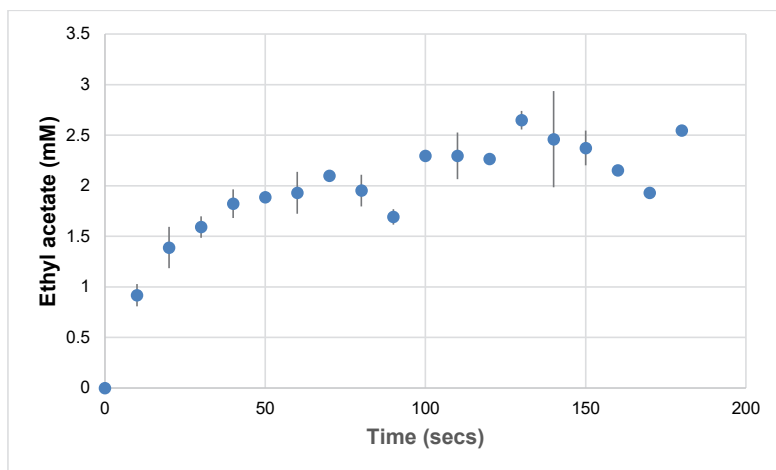
## Funding

The research presented in this article was financially supported by the Graduate School VLAG, Wageningen University and Research, the Netherlands.

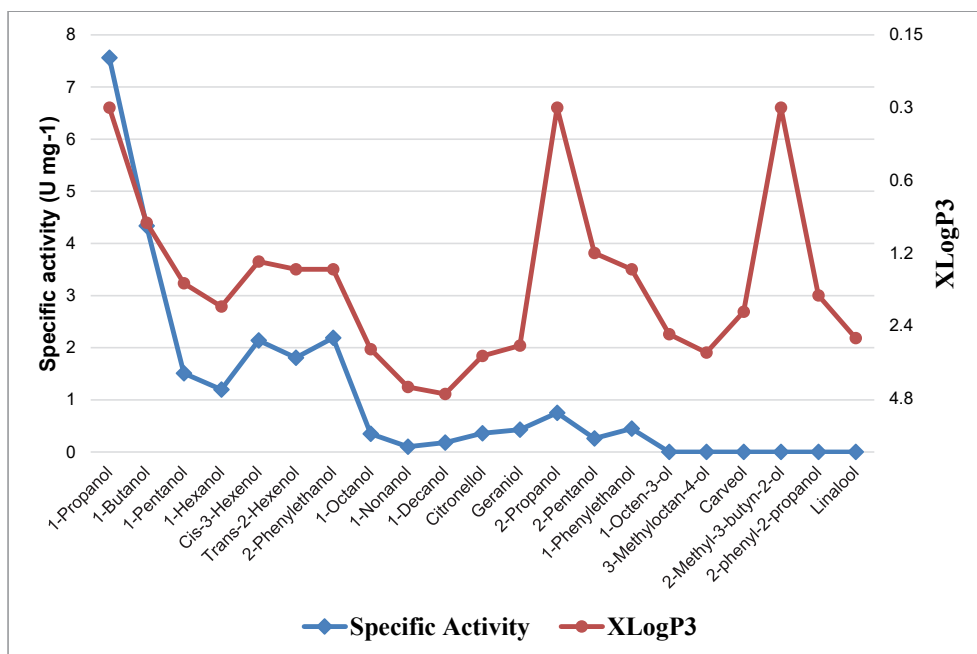
## Acknowledgements

We would like to thank Ton van Gelder, Rob Joosten and Steven Aalvink for their technical support. Also, we would like to thank dr. Alex Kruis and Dr. Mark Levisson for their advice on the Eat1 enzyme. We want to thank Axxence Aromatic GmbH, Germany, for kindly providing several chemical compounds. We want to thank Dr. Jules Beekwilder for kindly providing carveol.

## Supplementary information



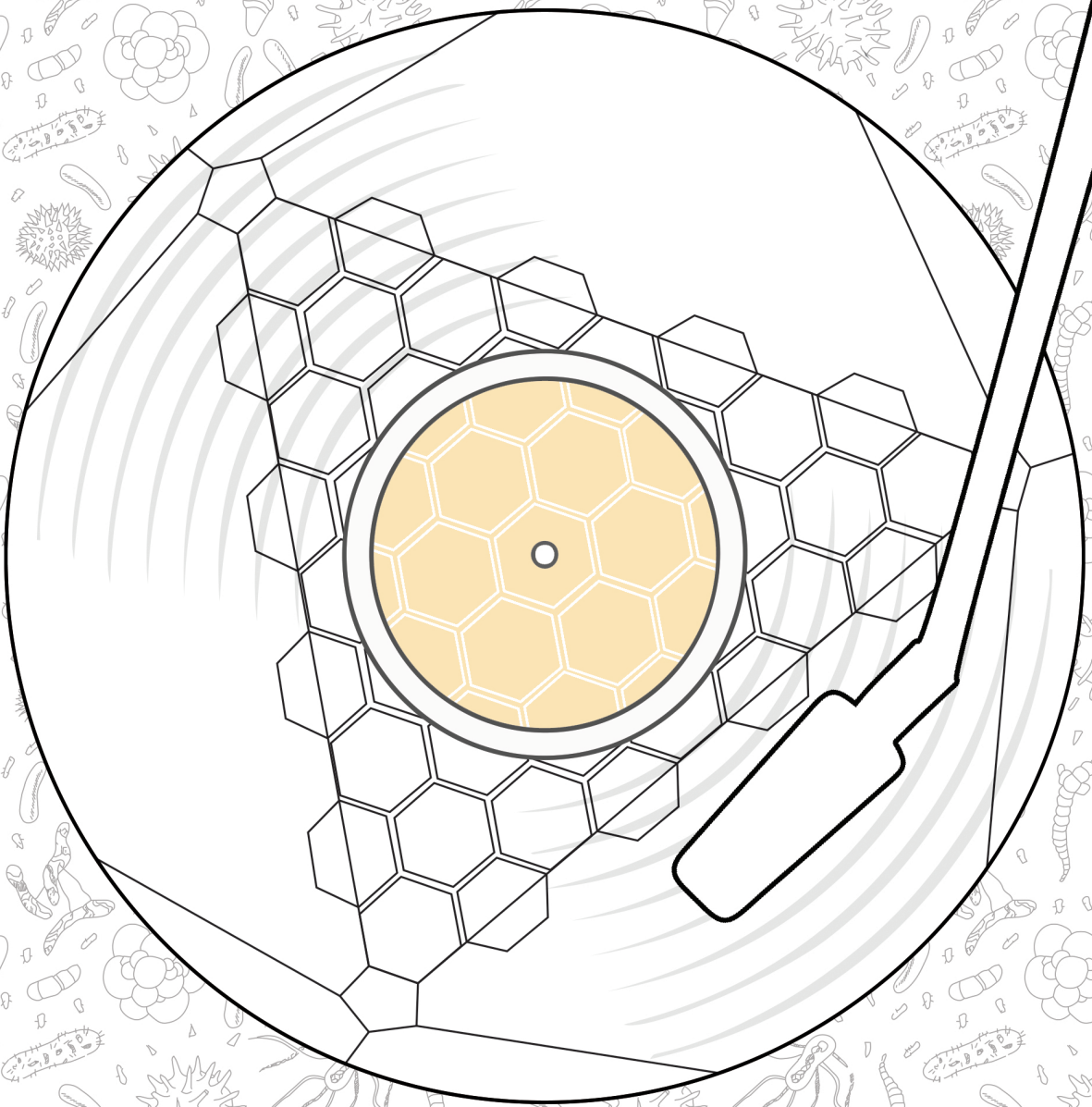
**Supplementary figure 1.** Detection of ethyl acetate in the presence of 2.5 mM 4-nitrophenyl acetate as acyl donor and 2.5 mM ethanol as acyl acceptor. 0.0318 mg ml<sup>-1</sup> WanEat1 was used.



**Supplementary figure 2.** Specific activity of Eat1 with different alcohols against their XLogP3 value. The XLogP3 value was derived from <https://pubchem.ncbi.nlm.nih.gov/>.



# CHAPTER 4



## Chapter 4

# Enzymes in a box: Repurposing clostridial microcompartments for the production of esters

Constantinos Patinios<sup>1,2,a</sup>, Caglar Yildiz<sup>1</sup>, Ana M López-Contreras<sup>3</sup>, John van der Oost<sup>1</sup>, Ruud A Weusthuis<sup>2</sup>, Servé WM Kengen<sup>1</sup>

<sup>1</sup>Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands.

<sup>2</sup>Laboratory of Bioprocess Engineering, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands.

<sup>3</sup>Bioconversion Group, Wageningen Food and Biobased Research, Bornse Weiland 9, 6708WG Wageningen, the Netherlands

Manuscript in preparation



### Abstract

Bacterial microcompartments (BMCs) are proteinaceous organelles in which dedicated metabolic processes are performed. Due to their prevalence, their crucial physiological role and their functional variability, BMCs have triggered the interest of many international research groups for both fundamental and applied research. In this study, we set the foundation for visualizing and repurposing the BMCs of the industrially relevant *Clostridium beijerinckii* NCIMB 8052 strain. We show that *C. beijerinckii* NCIMB 8052 has a glycyl radical enzyme type 3 (GRM3) BMC-operon and can metabolize L-rhamnose to 1,2-propanediol, which eventually is converted to the end products propanol and propionate. Several of the enzymes encoded by the BMC operon have N-terminal encapsulation peptides (EPs) which are responsible for the encapsulation of the enzymes into the lumen of the BMCs. We used two EPs (derived from the enzymes PduP and PduL) to encapsulate the yeast-enhanced green fluorescent protein (yEGFP) into the lumen of BMCs. The PduL EP showed distinct agglomerates, indicating successful encapsulation of yEGFP into the lumen of the BMCs. The PduL EP was then fused to the N-terminus of several alcohol acyl transferases (AATs) in an attempt to direct the AAT into the lumen of BMCs and produce propyl propionate. Unfortunately, we were yet unable to successfully show the production of propyl propionate through an encapsulated AAT.

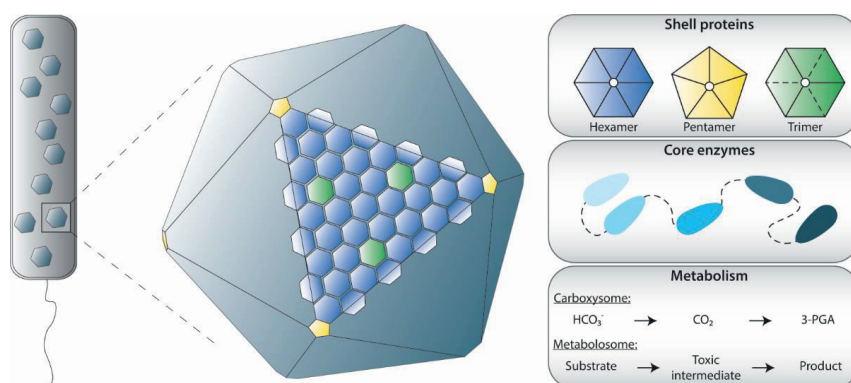
## Introduction

Specialized membrane-bound organelles, such as the nucleus and the mitochondrion, have given several advantages to the evolution of eukaryotic (micro)organisms. The nucleus stabilizes and protects the genome from mutations (347) and separates transcription from translation to allow for the slow mRNA splicing to proceed before the fast mRNA translation occurs (348). The mitochondria compartments allow for separating certain cellular metabolic processes from the cytosol, thereby increasing flux efficiencies (349). Certain bacteria also have membrane-bound organelles, such as the magnetosomes which serve as navigation organelles (350,351), and the anammoxosome compartment for anammox-related conversions that include toxic intermediates (352,353). In addition, some bacteria possess distinct protein-based organelles that perform specialized metabolic processes. Examples of these proteinaceous bacterial organelles include gas vesicles (354,355), encapsulins (356,357) and bacterial microcompartments (BMCs) (358,359).

BMCs are proteinaceous partitions that separate metabolic pathways from the rest of the cytosolic pool (Fig. 1). Several BMC types have been characterized and shown to prevent cell damage from toxic intermediates by quickly metabolizing them to non-toxic products (termed as metabolosomes) or to enhance CO<sub>2</sub> fixation by preventing its diffusion to the outside of the BMC (termed as carboxysomes) (Fig. 1) (358,359). To achieve the aforementioned activities, BMCs enclose several enzymes (termed as core enzymes) in a large quasi-icosahedral proteinaceous shell of about 100–400 nm in size (359). The shell of the BMCs is formed by hexameric, pentameric and trimeric proteins, termed as shell proteins, which have been purified and crystallized successfully, revealing pores in the center of their polymeric structures (360). These protein pores (4–14 Å) facilitate (passive) metabolite transport likely in a charge-dependent way (360,361). Thus, metabolites enter the BMCs through the pores of the shell proteins and are efficiently converted in the lumen of BMCs by encapsulated core enzymes, after which the generated products leave the BMCs via the shell pores.

Intriguingly, BMCs are widely present in the bacterial domain with more than 45 phyla having these peculiar structures (358,359,362,363). Due to their prevalence, their role in bacterial survival and their variety in function, BMCs have triggered the interest of many international research groups for fundamental and applied research (364–369). An interesting application of BMCs would be metabolic engineering in which BMCs can serve as modular “enzymes in a box”. By keeping the reaction-relevant enzymes in close proximity, BMCs can (presumably)

facilitate enhanced metabolisms by overcoming slow reaction rates, resource competition, co-factor imbalance, by-product formation or metabolite toxicity (349,370). Several studies have shown the successful repurposing of BMCs including the encapsulation of an alcohol dehydrogenase (Adh) in the lumen of BMCs for increased ethanol production, the encapsulation of a polyphosphate kinase (PPK1) for polyphosphate accumulation and the encapsulation of the lysis protein E for avoiding cell toxicity (365-369). In this study, as a proof of principle, we have chosen *Clostridium beijerinckii* NCIMB 8052 and attempted to re-purpose its BMCs for the production of ester compounds.

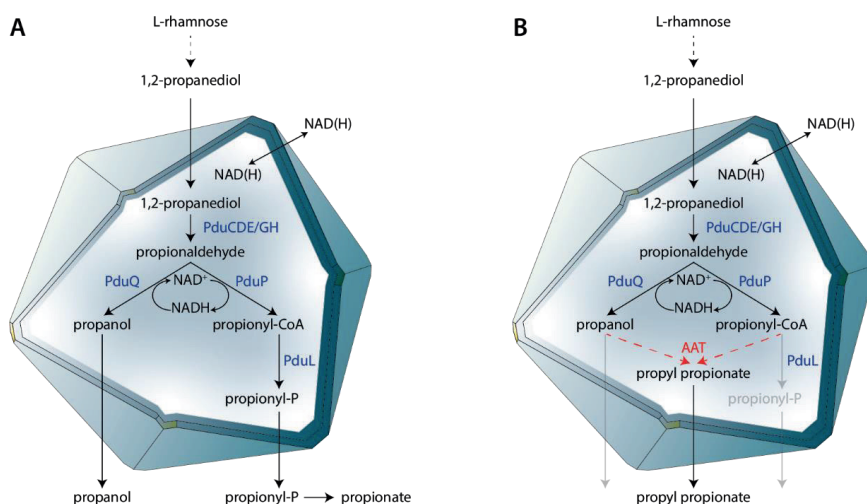


**Figure 1. Schematic of bacterial microcompartment structure and composition.** The hexameric, pentameric and trimeric shell proteins are interacting with each other to assemble the icosahedron BMC shell. The core enzymes, through interactions between them and the shell proteins, are encapsulated by the shell proteins to carry metabolic reactions: conversion of  $\text{HCO}_3^-$  to  $\text{CO}_2$  and then to 3-PGA (carboxysome), or conversion of a substrate to a toxic intermediate and then to a non-toxic terminal product (metabolosome). Unknown interactions between the core enzymes also occur and are shown with dashed lines.

Esters are versatile compounds with various applications in the chemical and food industries. Their production currently relies on the traditional Fischer-Speier esterification (33), although more sustainable methods through microbial cell factories have been developed (34). Ester production in *Clostridium* species has been demonstrated either through the supplementation of lipases (to combine endogenously produced acids and alcohols) or the expression of exogenous alcohol acyl transferases (AATs; to combine endogenously produced acyl-CoAs and alcohols) (25-31). A recent landmark study showed the production of 20.3 g/L of butyl acetate (made from endogenously produced acetyl-CoA and butanol) and 1.6 g/L butyl butyrate (made from endogenously produced butyryl-CoA and butanol) in *Clostridium saccharoperbutylacetonicum* N1-4 by testing several AATs and by using multiple metabolic

engineering approaches (32). One such approach included the enhancement of acetyl-CoA availability/flux. Since acetyl-CoA is a primary substrate for butyl acetate production, but it also serves as a precursor molecule for the production of acetate, butyrate, ethanol, acetone and butanol, its availability is critical for efficient production of acetate esters. To this end, availability of acetyl-CoA was enhanced by boosting acetate reassimilation through the conversion of acetone to isopropanol, resulting in 65% increase in butyl acetate production (32). Another approach to increase ester formation can be through the rational organization of enzymes associated with ester synthesis, like an 'assembly line' (32,349). This approach is based on placing the enzymes needed for the step-wise conversion in close proximity to avoid loss of substrates to competing pathways and to increase the local concentration of substrates and enzymes; an approach that may consequently result in increased reaction rates and productivity (371). A perfect location for such an approach would be the lumen of BMCs.

In this study, to overcome the low abundance of free acyl-CoAs in the cytosol of *C. beijerinckii* NCIMB 8052, and to increase the production of esters, we utilized the BMCs of *C. beijerinckii* NCIMB 8052 for the production of propyl propionate. Propyl propionate is a solvent with a fruity aroma (pineapple or pear) which is authorized as a food and beverage flavoring agent by EFSA and FAO (372,373). Also, because of its stronger solubility and lower odor values than acetate esters, propyl propionate is used in coatings, automotive refinishing, printing inks and as a polymerization solvent (374). Propyl propionate was chosen as the target ester compound due to the endogenous ability of the *C. beijerinckii* BMCs to produce the propionyl-CoA and propanol substrates in the lumen of the BMCs, when grown on L-rhamnose (Fig. 2A) (375). Therefore, we hypothesized that the addition of an AAT in the lumen of the BMC will result in the production of propyl propionate by combining the propionyl-CoA and the propanol substrates (Fig. 2B). Unfortunately, due to time restrictions, we did not succeed in demonstrating the encapsulation of AATs into the lumen of the *C. beijerinckii* NCIMB 8052 BMCs. However, as a proof of principle, we demonstrated the encapsulation of yEGFP by using the PduL encapsulation peptide (EP). This study is the first attempt to re-purpose *Clostridium* BMCs for the production of esters and will serve as the basis for further improvements and engineering of *Clostridium* BMCs.



**Figure 2. Native and engineered *C. beijerinckii* NCIMB 8052 BMC metabolism when grown on L-rhamnose or 1,2-propanediol. (A) Native BMC metabolism with propanol and propionate as end products. (B) Engineered BMC with propyl propionate as end product. An alcohol acyl transferase (AAT) is directed into the lumen of the BMC to facilitate propyl propionate production by combining the propanol and propionyl-CoA substrates. PduCDE: glycerol dehydratase; PduGH: glycyl-radical activating family protein; PduQ: propanol dehydrogenase; PduP: propionaldehyde dehydrogenase; PduL: phosphate propanoyl transferase.**

## Materials and methods

### Bacterial strains and handling

*E. coli* NEB® 5-alpha (NEB) was used for plasmid construction and propagation. Transformed *E. coli* cells were cultured in LB liquid medium (10 g L<sup>-1</sup> tryptone, 5 g L<sup>-1</sup> yeast extract, 10 g L<sup>-1</sup> NaCl) or on LB agar plates (LB liquid medium, 15 g L<sup>-1</sup> bacteriological agarose) containing spectinomycin (0.1 g L<sup>-1</sup>) and incubated at 37°C.

*C. beijerinckii* NCIMB 8052 was used for all the BMC related assays. Unless otherwise specified, transformed *C. beijerinckii* NCIMB 8052 was regularly grown anaerobically at 37°C in mCGM liquid medium (5 g L<sup>-1</sup> yeast extract, 0.75 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.75 g L<sup>-1</sup> K<sub>2</sub>HPO<sub>4</sub>, 0.4 g L<sup>-1</sup> MgSO<sub>4</sub> · 7H<sub>2</sub>O, 0.01 g L<sup>-1</sup> MnSO<sub>4</sub> · H<sub>2</sub>O, 0.01 g L<sup>-1</sup> FeSO<sub>4</sub> · 7H<sub>2</sub>O, 1 g L<sup>-1</sup> NaCl, 2 g L<sup>-1</sup> L-asparagine, 2 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.125 g L<sup>-1</sup> L-cysteine, 13.753 g L<sup>-1</sup> D-(+)-glucose · H<sub>2</sub>O) or on mCGM agar medium (1 g L<sup>-1</sup> yeast extract, 2 g L<sup>-1</sup> tryptone, 0.5 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 1 g L<sup>-1</sup> K<sub>2</sub>HPO<sub>4</sub>, 0.1 g L<sup>-1</sup> MgSO<sub>4</sub> · 7 H<sub>2</sub>O, 0.01 g L<sup>-1</sup> MnSO<sub>4</sub> · H<sub>2</sub>O, 0.015 g L<sup>-1</sup> FeSO<sub>4</sub> · 7 H<sub>2</sub>O, 0.01 g L<sup>-1</sup> CaCl<sub>2</sub>, 0.002 g L<sup>-1</sup> CoCl<sub>2</sub>, 0.002 g L<sup>-1</sup> ZnSO<sub>4</sub>, 2 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 55 g L<sup>-1</sup> D-(+)-glucose · H<sub>2</sub>O, 12 g L<sup>-1</sup> agar), containing spectinomycin (0.65 g L<sup>-1</sup>). For fermentation purposes, *C. beijerinckii* NCIMB 8052 was grown in GAPES liquid medium (2.5 g L<sup>-1</sup> yeast extract, 1 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.61 g L<sup>-1</sup> K<sub>2</sub>HPO<sub>4</sub>, 1 g L<sup>-1</sup> MgSO<sub>4</sub> · 7H<sub>2</sub>O, 0.0066 g L<sup>-1</sup> FeSO<sub>4</sub> · 7 H<sub>2</sub>O, 2.9 g L<sup>-1</sup> CH<sub>3</sub>CO<sub>2</sub>NH<sub>4</sub>, 0.19 g L<sup>-1</sup> pABA, 0.125 g L<sup>-1</sup> L-Cysteine (325)) containing either 49.5 g L<sup>-1</sup> D-(+)-glucose · H<sub>2</sub>O (250 mM) or 41 g L<sup>-1</sup> L-rhamnose (250 mM) and when appropriate spectinomycin (0.65 g L<sup>-1</sup>).

### Plasmid construction

The plasmids used in this study are using the backbone of the pCOSCB plasmids (376) and they are listed in Table 1. The strong constitutive thiolase promoter (ThlP) was used to express all the gene variants and the thiolase terminator (ThlT) was used to terminate transcription. The pAMB1 origin of replication was used to replicate the plasmid and the *aad9* gene was used as a spectinomycin resistance gene to ensure the retention of the plasmid in transformed *C. beijerinckii* NCIMB 8052. To facilitate the assembly of the pCOSCB plasmids, an adapted (377) Golden Gate Assembly method was used. Briefly, the gene of interest was amplified with primers containing overhangs which comprise of the BsaI recognition site, the necessary sequence of the backbone for insertion and a 5' buffer sequence (Table S1). The gene fragment size was verified by running gel electrophoresis and then purified by cutting the gene fragment

from the gel and using the Zymoclean Gel DNA Recovery kit (Zymo Research) by following the instructions of the manufacturer. 2  $\mu\text{L}$  (0.1 - 0.2 pmol  $\mu\text{L}^{-1}$ ) of the purified gene fragment was then mixed with 2  $\mu\text{L}$  (0.01 - 0.02 pmol  $\mu\text{L}^{-1}$ ) of the pCOSCB:EV plasmid and 2  $\mu\text{L}$  of MetaMix stock (10  $\mu\text{L}$  BsaI-HF $\text{v}2$ , 15  $\mu\text{L}$  T4 ligation buffer, 10  $\mu\text{L}$  T4 ligase and 15  $\mu\text{L}$  MQ). The mix solution was then incubated in a thermocycler using the following protocol: 5 min at 37°C, 5 min at 16°C followed by 5 min at 37°C (repeat for 15-30 cycles), 5 min at 37°C, 20 min at 80°C. 1  $\mu\text{L}$  of the solution was then used to transform chemically competent *E. coli* NEB $^{\text{®}}$  5-alpha (NEB) according to the instructions of the manufacturer. Transformed cells were plated on LB agar plates containing spectinomycin (0.1 g  $\text{L}^{-1}$ ) and incubated overnight at 37°C. Obtained colonies were used for plasmid cloning by growing them in 10 mL LB medium containing spectinomycin (0.1 g  $\text{L}^{-1}$ ), incubating overnight at 37°C. Plasmid purification was performed by using the GeneJET Plasmid Miniprep kit (ThermoFischer Scientific), using the manufacturer's instructions. To verify the insert sequence, plasmids were sequenced using Sanger sequencing (Macrogen Europe B.V.).

To obtain the pMCP-L or pMCP-P variants, an identical methodology as the one described above was used. However, the pMCP-L:EV or the pMCP-P:EV (Table 1) were used as backbones for the insertion of the gene of interest.

**Table 1. Plasmids used in this study.**

Plasmid name	Description and relevant characteristics	Reference
pCOSCB:EV	Aad9R, ColE1 ori, pAMB1 ori, Thlp-Thlt	Patinios <i>et al.</i> (2020)
pCOSCB:yEGFP	Aad9R, ColE1 ori, pAMB1 ori, Thlp-yEGFP-Thlt	This study
pCOSCB:WanEat1	Aad9R, ColE1 ori, pAMB1 ori, Thlp-WanEat1-Thlt	Patinios <i>et al.</i> (2020)
pCOSCB:SceAtf1	Aad9R, ColE1 ori, pAMB1 ori, Thlp-SceAtf1-Thlt	This study
pCOSCB:HarmSAAT	Aad9R, ColE1 ori, pAMB1 ori, Thlp-HarmSAAT-Thlt	This study
pMCP-P:EV	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduP EP-Thlt	This study
pMCP-P:yEGFP	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduP EP-yEGFP-Thlt	This study
pMCP-P:yEGFP-SsrA	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduP EP-yEGFP-SsrA-Thlt	This study
pMCP-L:EV	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduL EP-Thlt	This study

pMCP-L:yEGFP	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduL EP-yEGFP-Thlt	This study
pMCP-L:yEGFP-SsrA	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduL EP-yEGFP-SsrA-Thlt	This study
pMCP-L:WanEat1	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduL EP-WanEat1-Thlt	This study
pMCP-L:SceAtf1	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduL EP-SceAtf1-Thlt	This study
pMCP-L:HarmSAAT	Aad9R, ColE1 ori, pAMB1 ori, Thlp-PduL EP-HarmSAAT-Thlt	This study

### Bioinformatic analysis

To identify the BMC operon of *C. beijerinckii* NCIMB 8052, the MicroScope platform was used (378-383). Search terms including “microcompartments”, “propanol”, “propanediol”, “pdu” and “eut” were used. BlastP was used to identify the individual proteins of the *C. beijerinckii* Pdu proteins (NCBI). To identify the EPs of the BMC core enzymes, we again used the MicroScope platform and used the “blast and pattern search” application. The pattern query we used was [LIVAG]-[EKDR]-[EKDRQLIVAG]-[LIVAG]-[TEKDRQ]-[EKDRQ]-[LIVAG]-[LIVAG] or [LIVAG]-[EKDR]-[EKDRQ]-[LIVAG]-[LIVAG]-[EKDRQ]-[EKDRQ]-[LIVAG]-[LIVAG].

### *C. beijerinckii* transformation

To transform *C. beijerinckii* NCIMB 8052, 100  $\mu$ L of heat shocked *C. beijerinckii* spores (1 min at 99°C) were used to inoculate 25 mL of mCGM liquid medium followed by overnight incubation at 37°C. 20 mL of the overnight culture was used to inoculate 180 mL of pre-warmed (37°C) mCGM liquid medium which was then incubated at 37°C until an OD<sub>600</sub> equal to 0.3-0.4 was reached. Following, in an anaerobic tent, the culture was transferred into a 400 mL sterile centrifuge tube which was then sealed with parafilm to limit oxygen contamination. The culture was then centrifuged aerobically at 5,500 x g at 4°C for 10 min. The centrifuged culture was put on ice and transferred again in the anaerobic tent. The supernatant was then discarded and the cell pellet was resuspended with 25 mL of ice-cold anaerobic electroporation buffer (270 mM D-sucrose, 1 mM sodium phosphate buffer pH 7.4, 1 mM MgCl<sub>2</sub>). The resuspended culture was transferred into a 30 mL sterile centrifuge tube which was then sealed with parafilm to limit oxygen contamination. The resuspended culture was then centrifuged aerobically at



## Chapter 4

5,500 x g at 4°C for 10 min followed by discarding the supernatant and resuspending the cell pellet with 1.5 mL of ice-cold anaerobic electroporation buffer. 300 µL of the resuspended cells were used to electroporate (1.25 kV, 25 µF, 100 D) 3-5 µg plasmid DNA using 0.2 cm ice cold electroporation cuvettes. After electroporation, the transformants were recovered at 37°C in 3 mL of anaerobic mCGM for 3-4 h. After recovery, cultures were centrifuged at 5,500 x g for 5 min and the cell pellet was plated on mCGM solid medium containing spectinomycin (0.65 g L<sup>-1</sup>) followed by incubation at 37°C for 72-96 h.

8 colonies (if present) were screened through PCR for the presence of the desired gene and the amplicon was sent for sequencing (Macrogen Europe B.V.) to ensure the correctness of the construct. To perform colony PCR, colonies were resuspended in 50 µL PBS buffer (pH 7.4) and boiled for 10 min at 99°C. 1 µL of the boiled solution was then used as template for PCR using the Q5<sup>®</sup> High-Fidelity DNA Polymerase (NEB).

Glycerol stock of correct colonies was made by mixing 6 mL of a late log phase bacterial culture to 4 mL glycerol-phosphate solution (10.7 mM K<sub>2</sub>HPO<sub>4</sub>, 9.3 mM KH<sub>2</sub>PO<sub>4</sub>, pH7, 1:1 (v/v) glycerol). The vial was reduced by adding one drop of titanium citrate solution (100 mM) and stored at -80°C.

### Microscopy

1 mL of late exponential or early stationary phase culture was recovered and the cells were collected through centrifugation at 5500 x g for 5 min. The supernatant was decanted and the cells were resuspended in 1 mL PBS buffer (pH 7.4). The cells were then collected through centrifugation at 5500 x g for 5 min and the supernatant was decanted. The cells were finally resuspended in 500 µL PBS buffer (pH 7.4). 10 µL of the cell suspension was diluted in 4 mL PBS buffer (pH 7.4) and filtered through a 0.2 µM Millipore<sup>®</sup> Filter Membrane (Merck) to fix the cells. Approximately 15 µL VECTASHIELD<sup>®</sup> Antifade Mounting Medium with DAPI (Vector technologies) was used to stain the cells. Cells were viewed using a Nikon Eclipse Ti2 Confocal Fluorescence Microscope and the pictures were captured using its associated software. GFP emission was at 509 nm whereas for DAPI the emission was at 457 nm.

### Analytics

*C. beijerinckii* NCIMB 8052 cells were grown anaerobically in 25 mL GAPES liquid medium containing, when necessary, spectinomycin (0.65 g L<sup>-1</sup>) for 72 hours at 37°C. For the analyses of alcohol and ester products, 200 µL of the culture was recovered every 24 h and centrifuged

at 21,300 x g for 5 min after which 100  $\mu$ L of the supernatant was transferred in a 10-mL gas tight vial. The samples were analyzed on a Shimadzu GC 2010 gas chromatograph equipped with an HS-20 autosampler (Shimadzu). The sample vials were first heated at 60°C for 6 min to allow evaporation of the volatile products to the headspace of the vial. 1 mL of the headspace was then injected on a DB-WAX UI column (30 m length, 0.53 mm inner diameter, 1  $\mu$ m film thickness, Agilent) with a split ratio of 1:20. The column temperature was kept at 70°C for 1 min, then the temperature was increased to 125°C at a rate of 50°C min<sup>-1</sup> followed by an increase to 230°C at a rate of 70°C min<sup>-1</sup> after which the temperature was kept at 230°C for 1 min.

For the analyses of acids and sugars, a Shimadzu LC-2030 plus with a Shimadzu RID-20A detector was used. A Shodex SUGAR SH1821 column with a temperature of 65°C and flow time of 25 mins was used to separate metabolites. H<sub>2</sub>SO<sub>4</sub> was added to all samples to a final concentration of 0.01 M. References of D-glucose, L-rhamnose, lactate, acetate, propionate and butyrate were used to create a calibration curve.

All the data presented in this study are averages of biological triplicates and the standard deviation is presented as the error bar in all figures.

### ***In vitro* Eat1 alcoholysis assay**

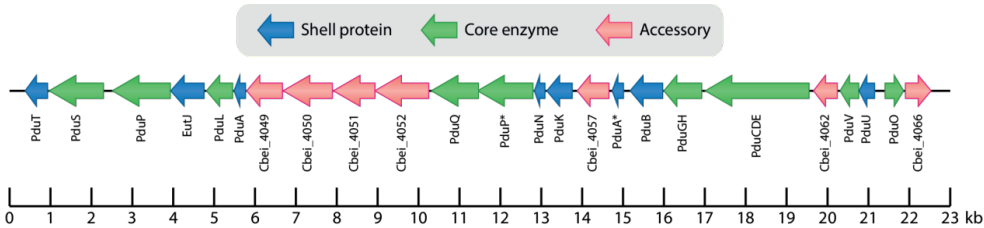
The *in vitro* alcoholysis assay was performed as described by Patinios *et al.* (2020). The reaction was performed in a total volume of 250  $\mu$ L phosphate buffer (50 mM KPi, 150 mM NaCl, pH 7.5). The solution was supplemented with 2.5 mM of ethyl propionate and 2.5 mM propanol as substrates and heated at 30°C for 10 minutes. The reaction was initiated by adding 13.98  $\mu$ g mL<sup>-1</sup> purified WanEat1 protein and was terminated by the addition of 250  $\mu$ L stop solution (0.1 N H<sub>2</sub>SO<sub>4</sub>). The stop solution contained 10 mM of acetone as internal standard. Furthermore, the substrates and the products were extracted by adding 250  $\mu$ L n-hexane, followed by vortexing and allowing for the extraction to complete for 15 mins. The final n-hexane extract was analyzed on a Shimadzu GC 2010 gas chromatograph as previously described (376).

## Results

### *C. beijerinckii* NCIMB 8052 has a GRM3 Pdu BMC operon

*Clostridium phytofermentans* was the first described *Clostridium* species to have BMCs (384). *C. phytofermentans* has a glycyl radical enzyme-containing microcompartment 5 (GRM5) locus characterized by the presence of genes that enable it to metabolize fucose and rhamnose, via the toxic propionaldehyde intermediate. Another *Clostridium* species, *C. beijerinckii* DSM 6423, was recently described to have a GRM3 locus as it contains the main enzymes to metabolize 1,2-propanediol to propanol and propionate and contains several ancillary genes such as an *EutJ* homolog, a flavoprotein-, an *S*-adenosylmethionine synthetase-, signaling protein- and a peptidase encoding-gene, which are characteristic genes found in GRM3 loci (375). Based on the available data for other *Clostridium* species, we investigated the presence of BMC related genes and proteins in *C. beijerinckii* NCIMB 8052.

Our bioinformatic analyses revealed that the *C. beijerinckii* NCIMB 8052 genome contains a BMC operon that is almost identical to that of *C. beijerinckii* DSM 6423 (Fig. 3). The operon belongs to the GRM3 locus as it has all the ancillary genes which are characteristic for this locus (Table 2). The genes encoding for the shell proteins PduA, A\*, B, K, N, T and U are present in the operon along with the *EutJ* homolog. Also, the complete pathway for the conversion of 1,2-propanediol to propanol and propionate is present (Fig. 2 and 3 and Table 2). All the homologs between the two strains share a protein sequence similarity between 93 and 100%. However, the *C. beijerinckii* DSM 6423 BMC operon contains a truncated (240 nts) *PduS* whereas the *C. beijerinckii* NCIMB 8052 BMC operon contains a full length (1326 nts) *PduS*. The *C. beijerinckii* DSM 6423 *PduS* has 74.36% sequence similarity to the C-terminus of the *C. beijerinckii* NCIMB 8052 *PduS*. Moreover, *PduT* is absent from the *C. beijerinckii* DSM 6423 BMC operon. The *C. beijerinckii* NCIMB 8052 BMC operon is split into two parts where Cbei\_4043 to Cbei\_4064 are encoded together in the same direction whereas Cbei\_4065 and Cbei\_4066 are encoded in the opposite direction. The two parts are separated by a 255 nt sequence which presumably contains a bidirectional promoter.



**Figure 3.** GRM3 Pdu BMC operon of *C. beijerinckii* NCIMB 8052. The ruler and gene length are to scale. Accessory proteins are labelled with their Cbei number. BMC proteins are labelled with their Pdu names.

**Table 2.** Gene and protein composition of the BMC locus in *C. beijerinckii* NCIMB 8052 and DSM 6423.

<i>C. beijerinckii</i> NCIMB 8052 gene	<i>C. beijerinckii</i> DSM 6423 gene homologue	BMC ID	Protein similarity between the two <i>C.</i> <i>beijerinckii</i> strains (%)	Protein function
Cbei_4043	N/A	PduT	N/A	Trimer (pseudohexamer) shell protein
Cbei_4044	CIBE_4883	PduS	74.36**	NADH dehydrogenase/Cobalamin reductase
Cbei_4045	CIBE_4884	PduP	96.82	Propionaldehyde dehydrogenase
Cbei_4046	CIBE_4885	EutJ	97.06	Ethanolamine utilization protein/Putative chaperonin
Cbei_4047	CIBE_4886	PduL	100	Phosphate propanoyl transferase
Cbei_4048	CIBE_4887	PduA	100	Hexamer shell protein
Cbei_4049	CIBE_4888	-	99.65	Hypothetical membrane protein
Cbei_4050	CIBE_4889	-	97.74	Methionine adenosyltransferase
Cbei_4051	CIBE_4890	-	97.32	Response regulator receiver protein
Cbei_4052	CIBE_4891	-	99.54	Signal transduction histidine kinase
Cbei_4053	CIBE_4892	PduQ	98.69	Propanol dehydrogenase
Cbei_4054	CIBE_4893	PduP*	98.22	Propionaldehyde dehydrogenase
Cbei_4055	CIBE_4894	PduN	100	Pentamer shell protein
Cbei_4056	CIBE_4895	PduK	93.72	Hexamer shell protein
Cbei_4057	CIBE_4896	-	97.67	Protein of unknown function
Cbei_4058	CIBE_4897	PduA*	100	Hexamer shell protein
Cbei_4059	CIBE_4898	PduB	99.62	Trimer shell protein

Cbei_4060	CIBE_4899	PduGH	98.37	Glycyl-radical activating family protein
Cbei_4061	CIBE_4900	PduCDE	98.70	Glycerol dehydratase
Cbei_4062	CIBE_4901	-	97.38	PfpI family intracellular peptidase
Cbei_4063	CIBE_4902	PduV	98.00	Conserved protein of predicted structural function
Cbei_4064	CIBE_4903	PduU	98.41	Hexamer shell protein
Cbei_4065	CIBE_4904	PduO	97.33	Adenosyltransferase
Cbei_4066	CIBE_4905	-	99.52	Cytidylate kinase-like family protein

N/A: not applicable.

\*: indicates the homology with another protein in the same operon (e.g. PduA\* is homolog of PduA).

\*\*: shows the percentage similarity between the 80 aa long CIBE\_4883 and the C-terminus of the 442 aa long Cbei\_4044.

### ***C. beijerinckii* NCIMB 8052 can metabolize L-rhamnose and produces propanol and propionate as terminal products**

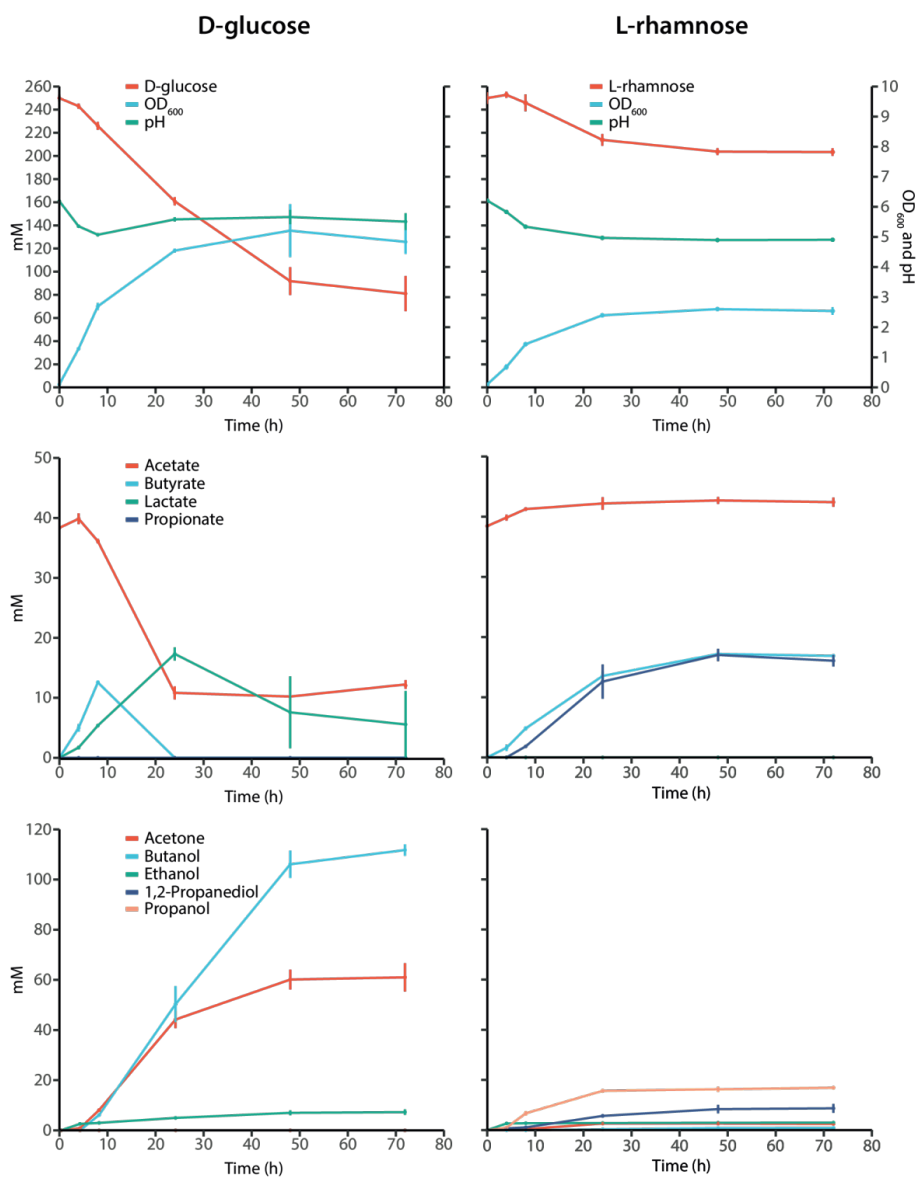
In a previous study by Diallo *et al.* (2019), growth of *C. beijerinckii* DSM 6423 on L-rhamnose and its conversion into propanol and propionate was demonstrated (375). Based on this study, and the presence of the BMC operon in *C. beijerinckii* NCIMB 8052, we assumed that *C. beijerinckii* NCIMB 8052 is also capable of utilizing L-rhamnose through the formation of BMCs. To demonstrate this, we grew *C. beijerinckii* NCIMB 8052 in GAPES medium containing L-rhamnose (250 mM) or D-glucose (250 mM) for 72 h during which we analyzed the fermentation products, the optical density (OD<sub>600</sub>) and the pH of the culture. Ammonium acetate was also included in the medium (~40 mM).

Acetate, butyrate and propionate were produced when *C. beijerinckii* NCIMB 8052 was grown on L-rhamnose (Fig. 4). Propionate ( $16.06 \pm 0.97$  mM) and butyrate ( $16.87 \pm 0.28$ ) were the main acid metabolites at the end of the fermentation. Acetate was not consumed, rather increased by approximately 4 mM. Propanol was the main alcohol product ( $16.91 \pm 0.56$  mM) followed by 1,2-propanediol ( $8.69 \pm 1.78$  mM). We were also able to detect the production of acetone ( $2.47 \pm 0.13$  mM), ethanol ( $2.97 \pm 0.05$  mM) and butanol ( $0.87 \pm 0.23$  mM), although at very low levels. A maximum OD<sub>600</sub> equal to  $2.6 \pm 0.06$  was reached after 48 h of cultivation and a pH crash was already observed at 24 h of cultivation (Fig. 4). The low OD<sub>600</sub> may be the result of the acid crash followed by the inability of the bacterial cells to fully utilize L-rhamnose,

as only 19% (47 mM) of the total L-rhamnose present at the start of the fermentation was consumed.

The typical ABE products were observed when *C. beijerinckii* NCIMB 8052 was grown in medium containing D-glucose as the main carbon source (Fig. 4). Acetone ( $61.02 \pm 5.66$  mM), ethanol ( $7.27 \pm 1.13$  mM) and butanol ( $111.76 \pm 2.25$  mM) were the main alcohol products. The maximum OD<sub>600</sub> was  $5.21 \pm 0.88$ , which is double of the maximum OD<sub>600</sub> when grown on L-rhamnose ( $2.6 \pm 0.06$ ). An initial drop in the pH ( $5.07 \pm 0.02$ ) was the result of the production of acetate, butyrate and lactate, followed by an increase in the pH ( $5.51 \pm 0.28$ ) at the end of the fermentation, which was the result of the consumption of the acids and the production of ABE.

In general, our results are in agreement to that of Diallo *et al.* (2019), and indicate the production of BMC metabolites (propanol, 1,2-propanediol and propionate) by *C. beijerinckii* NCIMB 8052 when grown on L-rhamnose.



**Figure 4.** Fermentation profile of the WT *C. beijerinckii* NCIMB 8052 strain when grown in medium containing D-glucose (left) or L-rhamnose (right). WT *C. beijerinckii* NCIMB 8052 was grown in GAPES medium containing 250 mM of either D-glucose or L-rhamnose as the main carbon source. Error bars indicate the standard deviation. (n=3).

## GFP encapsulation in the BMCs of *C. beijerinckii* NCIMB 8052 through the use of N-terminal EPs

Enzyme encapsulation in BMCs is generally achieved by conserved N-terminal encapsulation peptides (EPs) that interact with shell proteins, mediating their encapsulation into the lumen of the BMCs (365,385,386). Based on the conserved N-terminal EP sequence identified by other studies (385,387-389), we performed bioinformatic analysis to unravel the EP sequences of the *C. beijerinckii* NCIMB 8052 BMC core enzymes.

As expected, multiple hits were obtained, including several genes from the BMC operon (Table 3). Five of the obtained hits (PduO, PduS, PduP, PduL and PduQ) corresponded to the N-terminal sequence of the respective protein; a characteristic of EPs. All five proteins are catalytic proteins and they are expected to be in the lumen of BMCs, justifying the presence of an N-terminal EP at their corresponding protein sequence. In contrast, the rest of the core proteins (PduV, PduP\*, PduGH and PduCDE) did not have a hit for the N-terminal EP sequence. PduV also did not have an obvious EP, although previous studies showed its association with BMCs (390). Intriguingly, an EP sequence (IEVLKKII) was predicted at the C-terminus of PduGH, which may indicate that its encapsulation into the BMC relies on a C-terminal instead of an N-terminal EP.

To visualize the EP of the core proteins, we attempted to create 3D models of the core proteins using the AlphaFold2 protein structure prediction software (391). PduP and PduL clearly displayed an N-terminal  $\alpha$ -helix sequence extending outside of the main protein structure (Fig. S1) (369,392). PduS also displayed an N-terminal  $\alpha$ -helix sequence, although much shorter than the ones of PduL and PduP. A previous study showed that the *Citrobacter freundii* PduS interacts with the PduT shell protein through its N-terminal sequence, suggesting that PduS has an N-terminal EP (393). In accordance with our bioinformatics prediction, the 3D structure of PduO clearly shows an  $\alpha$ -helix extension at the N-terminus of the protein. To our knowledge, there is no study that shows the encapsulation of PduO through an N-terminal EP. However, PduO interacts with PduS (394) and it has been suggested that PduO is “piggybacking” on PduS for its encapsulation. PduV displayed both an N- and C-terminal extension that may resemble EPs. The N-terminus of PduV was previously shown to be important in encapsulation, although a larger (48 aa) sequence was used (390). N- and C-terminal extensions that resemble EPs were also apparent for PduCDE, although the C-terminus of PduCDE was predicted to form interactions with the main protein structure. PduQ did not show the characteristic  $\alpha$ -helix



extension at the N-terminus, rather, an  $\alpha$ -helix extension was apparent at the C-terminus of the protein. C-terminal EPs have been described before, demonstrating that EPs are not necessarily found at the N-terminus of the core proteins (386). Moreover, the N-terminus of PduQ is predicted as an unstructured, distorted peptide, possibly suggesting that its encapsulation may be mediated through its C-terminal sequence. Lastly, based on the predicted PduGH structure, the C-terminus of the protein has a characteristic exposed  $\alpha$ -helix whereas its N-terminus has a distorted region. These data suggest that PduQ and PduGH may be encapsulated through C-terminal peptides whereas the rest of the Pdu enzymes have N-terminal EPs that enable their encapsulation.

**Table 3. Predicted EP at the N-terminus of *C. beijerinckii* NCIMB 8052 BMC core proteins. The predicted EP sequence is underlined and bold.**

<i>C. beijerinckii</i> NCIMB 8052 gene	BMC ID	Predicted N-terminal encapsulation peptide sequence (display of the first 25 aa)
Cbei_4044	PduS	MEKSF <u><b>DLIKDAG</b></u> IIGAGGAGFP
Cbei_4045	PduP	MDVDVVL <u><b>VEKLVRQA</b></u> IEEVKNKNLL
Cbei_4047	PduL	MHSEDV <u><b>VKLITKIV</b></u> VDKIKALENYK
Cbei_4053	PduQ	MVFQL <u><b>GRKGEEVL</b></u> KSVYIKTKIYSG
Cbei_4065	PduO	MINEE <u><b>LIKKLSSD</b></u> LIISSIKTPKLT

Green: hydrophobic uncharged residues, like F I L M V W A and P; Red: acidic residues, like D E and C-terminal -COOH; Blue: basic residues, like R K H and N-terminal -NH<sub>2</sub>; Black: other residues, like G S T C N Q and P.

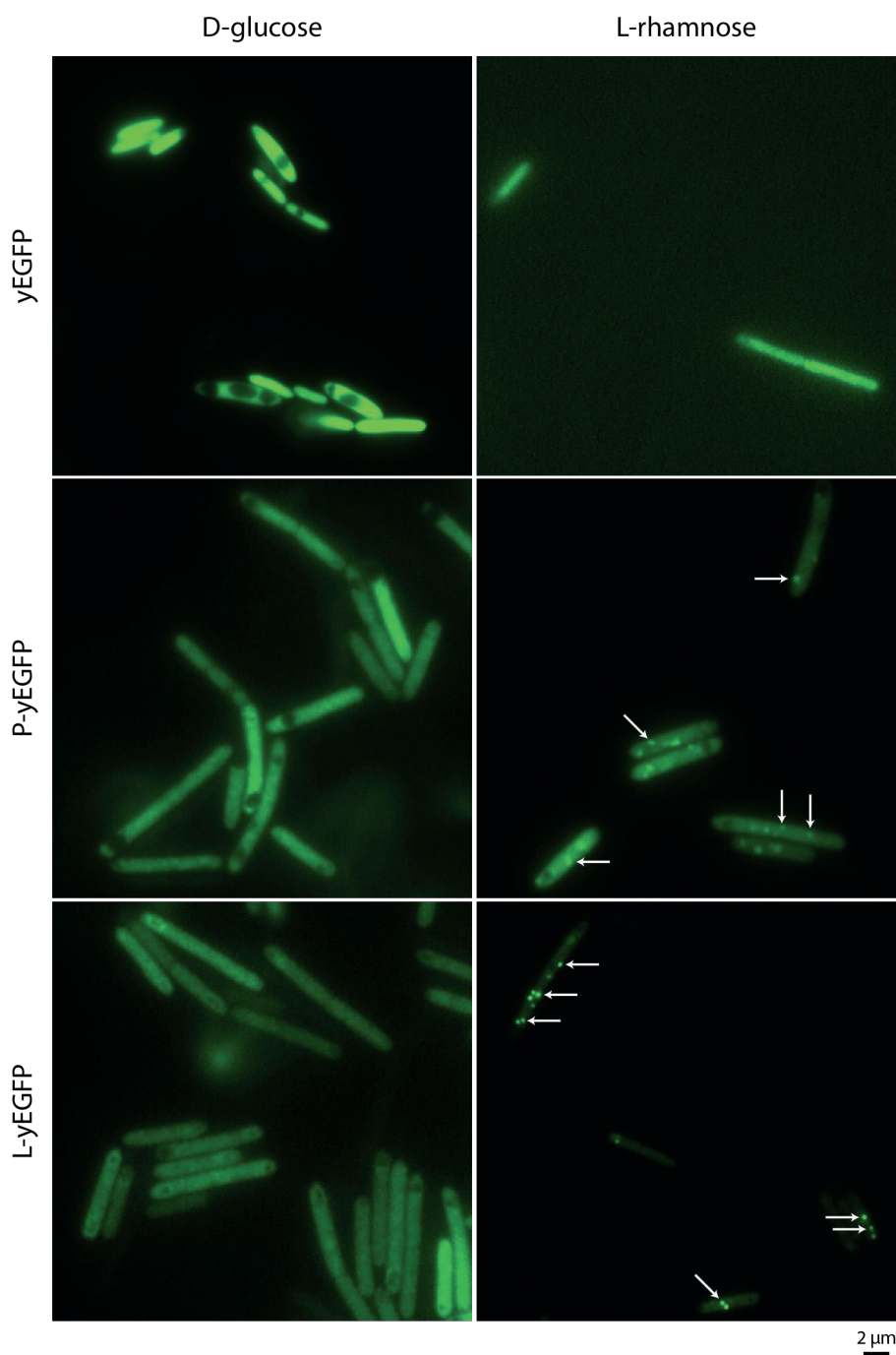
The results of our bioinformatic analysis and protein structure prediction revealed distinct N-terminal EPs at the PduP and PduL proteins. To investigate the capacity of *C. beijerinckii* NCIMB 8052 BMCs to encapsulate exogenously expressed proteins, we constructed a library of plasmids that constitutively express yEGFP with or without the presence of the PduP (MDVDVVLVEKLVRQAIEEVK) or the PduL (MHSEDVVKLITKIVVDKIK) EPs at its N-terminus and expressed these variants in conditions that induce (L-rhamnose) or prevent (D-glucose) BMC formation. yEGFP variants that have the PduP EP at their N-terminus were named P-yEGFP whereas variants that have the PduL EP at their N-terminus were named L-yEGFP.

According to our expectations, the constitutive expression of yEGFP in *C. beijerinckii* NCIMB 8052 showed full cell fluorescence regardless which carbon source (L-rhamnose or D-glucose) was present in the medium (Fig. 5, S2 and S3). Full cell fluorescence was observed in transformants expressing the P-yEGFP or L-yEGFP variants and grown in medium containing

D-glucose as the sole carbon source (Fig. 5, S4 and S6). In contrast, when transformants carrying the P-yEGFP and L-yEGFP variants were grown in medium containing L-rhamnose as the sole carbon source, green agglomerates were observed (Fig. 5, S5 and S7). The agglomerates were distinct for L-yEGFP and little cytosolic fluorescence was observed. The P-yEGFP variant also showed agglomerates, however, strong cytosolic fluorescence inhibited the distinct visualization of the agglomerates. The agglomerates observed for the P-yEGFP and L-yEGFP are in agreement with other reports in literature where such agglomerates were observed when the PduP, PduL or other EPs were fused at the N-terminus of GFP (369,370,388,390,395,396). These results indicate successful encapsulation of the P-yEGFP and L-yEGFP variants into the lumen of the *C. beijerinckii* NCIMB 8052 BMCs, although with residual cytosolic yEGFP.

To address the presence of residual cytosolic yEGFP molecules, we attached the *E. coli* SsrA degradation peptide (AANDENYALAA) at the C-terminus of the yEGFP variants (397,398). This approach has been used in other studies to eliminate cytosolic yEGFP (decreased half-life due to increased degradation by the proteasome) and allows only the visualization of BMC-encapsulated yEGFP, i.e. physically separated from the cytoplasmic proteasome (369,399-401). We have chosen the *E. coli* SsrA tag as it was previously shown to be recognized by the proteasome of other *Clostridium* species (402,403). yEGFP variants that have the SsrA tag at their C-terminus were named accordingly (e.g. yEGFP-SsrA, P-yEGFP-SsrA and L-yEGFP-SsrA; also see Table 1).

As expected, absence of fluorescence was observed in the yEGFP-SsrA transformants in both growth conditions (Fig. S8). Surprisingly, the L-GFP-SsrA and P-GFP-SsrA variants, did not show fluorescence neither in BMC favoring or disfavoring conditions (Fig. S9 and S10). Although we cannot pinpoint the reason why we did not observe encapsulation of the P-yEGFP-SsrA and L-yEGFP-SsrA variants in the lumen of BMCs, we decided to continue with our experiments since the L-yEGFP variant showed distinct agglomerates, indicating (at least some) encapsulation of the target protein.



**Figure 5.** Encapsulation of yEGFP into the lumen of *C. beijerinckii* NCIMB 8052 BMCs. *C. beijerinckii* NCIMB 8052 cells were transformed with the yEGFP, P- yEGFP and L- yEGFP variants and grown in BMC favoring (L-rhamnose) and disfavoring (D-glucose) conditions. White arrows indicate BMCs.

## Attempting propyl propionate production in the BMCs of *C. beijerinckii* NCIMB 8052

The rhamnose metabolism of *C. beijerinckii* involves the production of 1,2-propanediol, which is further metabolized in the lumen of the BMCs (Fig. 2A). 1,2-propanediol is converted into the toxic propionaldehyde, which presumably is the main reason that this pathway operates inside the BMC. Propionaldehyde is then converted into propanol and propionyl-CoA, with the latter being further converted into propionate (375). Thus, *C. beijerinckii* NCIMB 8052 BMCs contain both propionyl-CoA and propanol, which can serve as the substrates for an AAT for the production of propyl propionate (Fig. 2B).

From a library of characterized AATs (Atf1/2, Eht1, Eeb1, Eat1, SAAT, VAAT), SAAT was previously shown to produce propyl propionate by combining propanol and propionyl-CoA in *E. coli* (37). Therefore, SAAT was chosen as the best candidate for the production of propyl propionate in *C. beijerinckii* NCIMB 8052 BMCs. In addition, Eat1 was previously shown to accept C2 and C4 acyl donors (376). Since propionyl-CoA has a C3 acyl moiety, we tested the potency of Eat1 to accept propionate and transfer it to an alcohol donor. To test this in a fast and convenient fashion, we performed *in vitro* alcoholysis assays between ethyl propionate (acyl donor) and propanol (acyl acceptor) as previously described by Patinios *et al.* (2020). Indeed, Eat1 could produce propyl propionate through alcoholysis (Table S2) and was selected as an AAT candidate for the production of propyl propionate in *C. beijerinckii* NCIMB 8052 BMCs. In contrast to SAAT and Eat1, Atf1 is not capable of accepting propionyl-CoA (37) and was used as a control in this study.

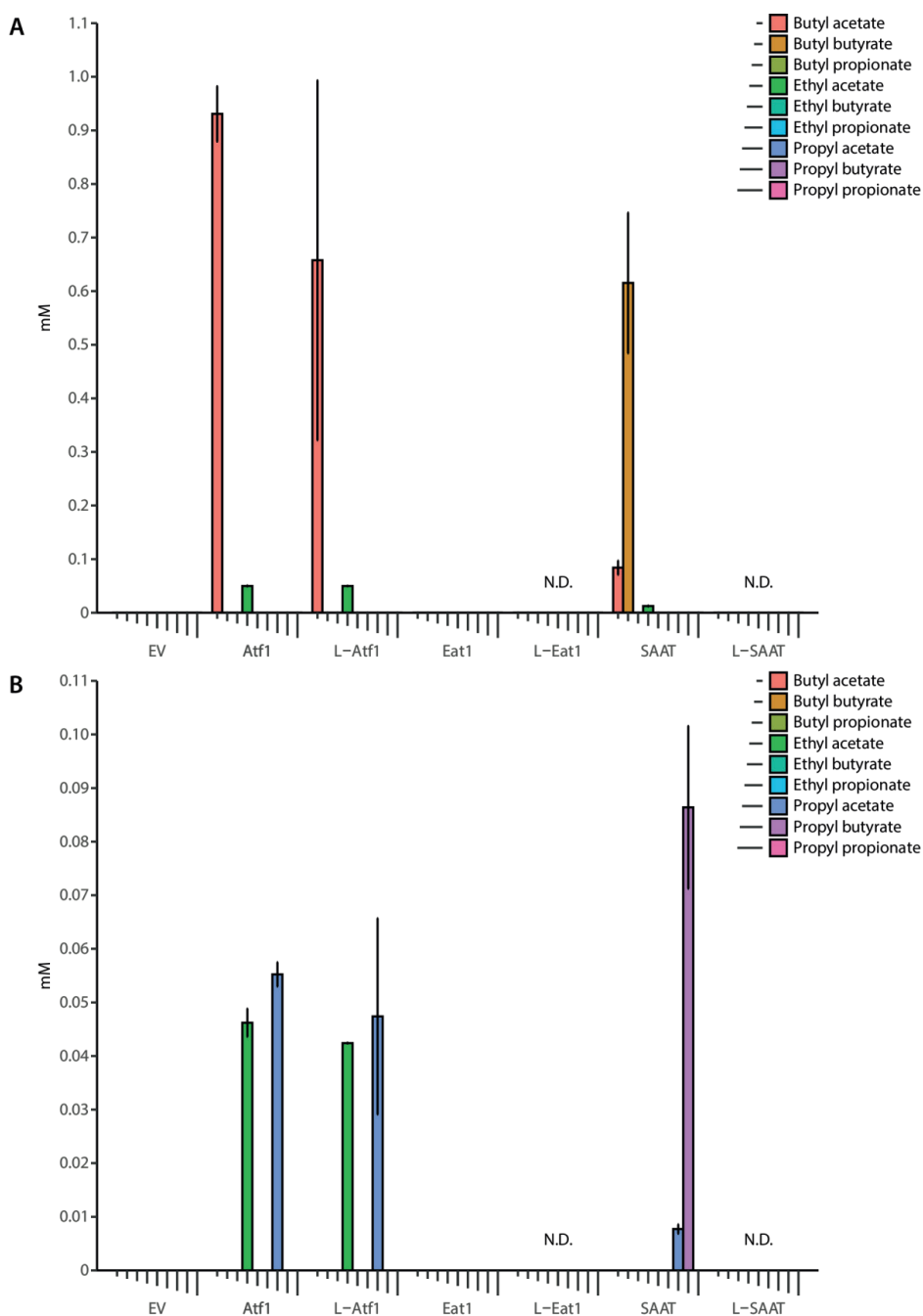
In total, three AATs (SAAT, Eat1 and Atf1) were chosen to assess the production of propyl propionate by directing them into the lumen of *C. beijerinckii* NCIMB 8052 BMCs. To this end, the PduL N-terminal EP was attached to the N-terminus of SAAT, Eat1 and Atf1 to create L-SAAT, L-Eat1 and L-Atf1, respectively, and were transformed in *C. beijerinckii* NCIMB 8052 cells. The PduL N-terminal EP was chosen based on our previous results for successful encapsulation of yEGFP (Fig. 5). As a control, wild-type versions (i.e. without the N-terminal EP) of SAAT, Eat1 and Atf1 were expressed in *C. beijerinckii* NCIMB 8052. Since the SsrA degradation tag did not improve the encapsulation efficiency of yEGFP, it was excluded from our setup. All constructs were expressed under the constitutive ThlP promoter and an empty vector (i.e. without expressing any gene) was also used as a control.

Transformed *C. beijerinckii* NCIMB 8052 containing the empty vector control did not show production of esters, regardless of the growth medium (L-rhamnose or D-glucose) (Fig. 6). Cultures expressing Atf1 or L-Atf1 produced butyl acetate ( $0.93 \pm 0.05$  mM and  $0.66 \pm 0.34$  mM, respectively) and ethyl acetate ( $0.05 \pm 0.01$  mM and  $0.05 \pm 0.01$  mM, respectively) when grown on D-glucose. Propyl acetate ( $0.06 \pm 0.01$  mM and  $0.05 \pm 0.02$  mM, respectively) and ethyl acetate ( $0.05 \pm 0.01$  mM and  $0.04 \pm 0.01$  mM, respectively) were produced when grown on L-rhamnose. As expected, since Atf1 cannot accept propionyl-CoA as substrate, there was no production of propionate esters when transformed *C. beijerinckii* NCIMB 8052 expressing Atf1 or L-Atf1 was grown on L-rhamnose.

Cultures expressing Eat1 did not show ester production regardless of the growth medium (L-rhamnose or D-Glucose). The inability of Eat1 to produce esters in *C. beijerinckii* NCIMB 8052 was also observed in our previous study (376). Although the AAT activity of Eat1 could not be observed, its alcoholysis activity was still apparent, indicating that Eat1 was functionally expressed. It is known that Eat1 has very low affinity for acetyl-CoA (and probably other acyl-CoAs) which may explain the inability of Eat1 to produce esters in *C. beijerinckii* NCIMB 8052 (35), despite being active. However, studies in *E. coli* have shown that Eat1 is capable of using acetyl-CoA and ethanol efficiently and produce ethyl acetate at 72% of the maximum pathway yield (339,404). We believe that the low acyl-CoA availability in *C. beijerinckii* NCIMB 8052 in combination with the low affinity of Eat1 to acyl-CoAs prevents the production of esters. To overcome this limitation, we attempted to encapsulate Eat1 in the lumen of BMCs where both the acyl donor (propionyl-CoA) and the acyl acceptor (propanol) are present in close proximity (Fig. 2B). To this end, we fused the PduL EP to the N-terminus of Eat1, and created L-Eat1. Unfortunately, after multiple attempts, we were unable to obtain transformants carrying the L-Eat1 variant. At this point, we could not identify the reason for not obtaining transformants but a possible toxic combination between the PduL EP and the Eat1 protein is possible. Future experiments should assess the change of the PduL EP with other EPs and even the truncation of the N-terminus of Eat1 and its replacement with the PduL EP.

Transformants expressing the SAAT variant produced ethyl acetate ( $0.01 \pm 0.01$ ), butyl acetate ( $0.08 \pm 0.01$ ) and butyl butyrate ( $0.62 \pm 0.13$ ) when grown in medium containing D-glucose as the main carbon source, or propyl acetate ( $0.01 \pm 0.01$ ) and propyl butyrate ( $0.09 \pm 0.02$ ) when grown in medium containing L-rhamnose as the main carbon source. Production of butyl butyrate by expressing SAAT was also observed in *C. saccharoperbutylacetonicum*, which reached  $1.6 \text{ g L}^{-1}$  after using extensive metabolic engineering approaches (32). Since, SAAT

was functional in *C. beijerinckii* NCIMB 8052 and was previously shown to produce propyl propionate (37), we attached the PduL EP at its N-terminus (creating L-SAAT) in order to localize it in the *C. beijerinckii* NCIMB 8052 BMCs to produce propyl propionate. Unfortunately, after multiple attempts we could not obtain transformants carrying L-SAAT. Similar to the failed attempts to transform L-Eat1 in *C. beijerinckii* NCIMB 8052 BMCs, we cannot reason why L-SAAT was not transformed successfully.



**Figure 6.** Ester products of *C. beijerinckii* NCIMB 8052 transformed with the AAT variants. (A) Ester products when transformants were grown in medium containing D-glucose (A) or L-rhamnose (B) as the main carbon source. Empty vector (EV) was used as control. Error bars indicate the standard deviation. N.D.: not determined. (n=3).

## Discussion and conclusion

In this study, we describe the production of BMCs and BMC-metabolites in *C. beijerinckii* NCIMB 8052. We show that *C. beijerinckii* NCIMB 8052 has a GRM3 Pdu BMC operon composed by the necessary genes and enzymes to metabolize L-rhamnose to 1,2-propanediol and subsequently to the terminal propanol and propionate products. The *C. beijerinckii* NCIMB 8052 BMC operon is split into two parts where PduO (adenosyltransferase) and the Cbei\_4066 accessory gene are encoded in a different direction from the rest of the Pdu genes. The two parts of the operon are separated by a 255 bp intergenic sequence, which presumably contains a bidirectional promoter, the activity of which is likely controlled by the L-rhamnose catabolism (most likely by 1,2-propanediol). The GRM3 subtype is characterized by the B<sub>12</sub>-independent dehydration of 1,2-propanediol, although the B<sub>12</sub>-recycling enzymes PduS (cobalamin reductase) and PduO (adenosyltransferase) are present in the GRM3 operon (405). There is no solid experimental explanation why PduS and PduO are present in the operon but a current hypothesis suggests that PduS may act as an NADH:flavodoxin oxidoreductase which extracts electrons from NADH to reduce the PduGH (activating enzyme) and support the activation of PduCDE (glycyl radical enzyme) (405). Whether PduS and PduO are important for the GRM3 Pdu BMC metabolism, whether they are important for assembly or whether they are just relics of evolution can be assessed only by the inactivation/elimination of these genes from the operon.

The growth of *C. beijerinckii* NCIMB 8052 on L-rhamnose is impaired compared to growth on D-glucose. The impaired growth could be due to a pH/acid crash since the bacterial cells were not able to assimilate the produced acids (acetate, propionate and butyrate) and convert them into alcohols to alleviate the pH crash. Moreover, the consumption of L-rhamnose and the generation of products (acids and alcohols) terminated at around 24 h of fermentation, supporting the observed impaired growth. A similar observation was made by Diallo *et al.* (2019) when *C. beijerinckii* DSM 6423 was grown on L-rhamnose (375). The authors excluded the possibility of an acid crash since pH-controlled conditions did not enhance the growth and the L-rhamnose conversion. We took a different approach to test this hypothesis by removing ammonium acetate from the growth medium and consequently reducing the acid pressure (data not shown). However, this strategy did not enhance growth or the consumption of L-rhamnose, further supporting that the impaired growth is not the result of an acid crash. Diallo *et al.* (2019) further suggested that the impaired growth may be based on the lower ATP yield (38% lower) when *C. beijerinckii* is grown on L-rhamnose instead of on D-glucose. Also, the protein shell



and the core enzymes of the BMCs were suggested to impose a large energy burden on the organism which result in lower growth (375). Although both suggestions are valid, we would expect suppression of the growth rate rather than suppression of the final optical density, product formation and substrate utilization. The reason for the observed phenotype is obscure but its effect is vital for biotechnological applications as it can hinder product titer, yield and productivity. Future endeavors should consider decoupling the BMC metabolism from any regulation (i.e., the 1,2-propanediol-dependent promoter) in order to expose the cells to optimal for growth and production conditions (e.g. on D-glucose).

Several studies showed encapsulation of foreign proteins into the lumen of BMCs through the attachment of EPs at the N-terminus of the target protein (365,385-389). Similarly, we assessed the encapsulation of yEGFP in the lumen of the *C. beijerinckii* NCIMB 8052 BMCs by attaching the 20 aa EP of the PduP or PduL enzymes at the N-terminus of yEGFP. Both strategies showed agglomerates in BMC inducing conditions (L-rhamnose), although PduL showed distinct agglomerates and less cytosolic yEGFP. To avoid the presence of cytosolic yEGFP, we attached the SsrA degradation peptide at the C-terminus of yEGFP. This strategy ensures the degradation of cytosolic yEGFP and prevents the degradation of BMC-encapsulated yEGFP (369,399-401). Unfortunately, both the P-yEGFP-SsrA and L-yEGFP-SsrA constructs, did not show the expected distinct agglomerates that would indicate BMC encapsulation. At this stage we cannot pinpoint the reason why we did not observe encapsulation of the P-yEGFP-SsrA and L-yEGFP-SsrA variants in the lumen of BMCs. However, we can speculate that the *E. coli* SsrA degradation tag is rapidly recognized by the *C. beijerinckii* NCIMB 8052 proteasome which in turn rapidly degrades the P-yEGFP-SsrA and L-yEGFP-SsrA variants, preventing their encapsulation in the lumen of BMCs. Future studies may consider the use of different degradation tags and even the use of the (predicted) endogenous *C. beijerinckii* NCIMB 8052 SsrA tag (AEDNFALAA) derived from the Cbei\_R0066 tmRNA (Table S3).

Although the attachment of the SsrA degradation tag did not eliminate the presence of cytosolic yEGFP, we could differentiate distinct agglomerates when the PduL EP was used. We believe that the agglomerates represent the encapsulation of yEGFP in the lumen of BMCs as they are present only in inducing conditions (L-rhamnose) and the obtained images align very well with data presented in other reports (369,399-401). Future studies should assess successful protein encapsulation by BMCs by fusing, for example, one of the shell proteins to a fluorescent protein (e.g. RFP) and check its colocalization with the encapsulated yEGFP.

Following the encapsulation of yEGFP in the lumen of the *C. beijerinckii* NCIMB 8052 BMCs, we used the 20 aa PduL EP to encapsulate various AATs in the lumen of the BMCs in an attempt to produce propyl propionate. Unfortunately, the two most important variants, L-Eat1 and L-SAAT, could not be transformed in *C. beijerinckii* NCIMB 8052 cells. We do not know the cause of this but potential toxicity of the EP-AAT variant may be the reason since the WT (i.e. without the attachment of the N-terminal EP) Eat1 and SAAT variants were successfully transformed in *C. beijerinckii* NCIMB 8052. Also, the L-Atf1 variant was successfully transformed in *C. beijerinckii* NCIMB 8052 which indicates that toxicity may be present only with the L-Eat1 and L-SAAT variants. To avoid toxicity one may change the EP fused to the target AAT enzymes. This approach may require the generation of a library of EPs corresponding to the different EPs of the Pdu core enzymes. Also, a different approach may include the fusion of the AAT protein to the C-terminus of one of the Pdu core enzymes or to the C-terminus of the L- or P-yEGFP, to facilitate encapsulation of the enzyme in the lumen of BMCs.

Production of esters was observed by the Atf1, L-Atf1 and SAAT variants although at very low quantities: <1 mM when grown on D-glucose and <0.1 mM when grown on L-rhamnose. The low production of esters is likely the result of the inhibited growth of transformed strains as low OD<sub>600</sub> and low production of the typical ABE (on D-glucose) or propanol (on L-rhamnose) products was observed (data not shown). Growth inhibition may be the result of the use of antibiotics to retain the plasmid in the transformed strain but also due to the energy burden that the replication and expression of the plasmid might have on the cells. Future studies should introduce the AAT variants in *C. beijerinckii* NCIMB 8052 through chromosomal integration in by using, for example, CRISPR-Cas technologies (65,67,406-411).

In summary, this study presents the first attempt to visualize and repurpose the BMCs of *C. beijerinckii* NCIMB 8052 and will serve as a basis for future studies that aim to characterize and engineer the BMCs of clostridia species.

## Acknowledgements

We would like to thank Dr. Burak Avci for his help with the microscopy-related work of this study and the fruitful discussions surrounding BMCs.

## Supplementary information

**Supplementary table 1. Oligonucleotides used in this study.**

Oligo ID	Oligo sequence (5' to 3')	Description
BG24714	AAAATTTTAGGAGGTCAAAC ATGTCTAAAGGTGAAGAATT ATTC	Amplification of yEGFP for insertion into pCOSCB:EV through Gibson Assembly, forward
BG24715	AGAGTTATTTTAAACAATACT TTTATTTGTACAATTCATCA ATACC	Amplification of yEGFP for insertion into pCOSCB:EV through Gibson Assembly, reverse
BG16270	ATGAATGAAATCGATGAGAA AAATCAG	Amplification of SeeAtf1 for insertion into pCOSCB:EV through Gibson Assembly, forward
BG16271	AGAGTTATTTTAAACAATACT TTCTAAGGGCCTAAAAGGAG AG	Amplification of SeeAtf1 for insertion into pCOSCB:EV through Gibson Assembly, reverse
BG25564	GATCAGGTCTCACAAACATG GAAAAAATAGAAGTGTCTAT TAATTCTAAAC	Amplification of HarmSAAT for insertion into pCOSCB:EV through Golden Gate, forward
BG25565	GATCAGGTCTCGACTTTCTAT ATAAGAGTTTTTGGTGAAGCA AG	Amplification of HarmSAAT for insertion into pCOSCB:EV through Golden Gate, reverse
BG24828	GATCAGGTCTCAGAAATCTAA AGGTGAAGAATTATTCAGT	Amplification of yEGFP for insertion into pMCP-P:EV through Golden Gate, forward
BG24826	GATCAGGTCTCACTTTTATT TGTACAATTCATCAATACCAT G	Amplification of yEGFP for insertion into pMCP-P:EV through Golden Gate, reverse
BG24828	GATCAGGTCTCAGAAATCTAA AGGTGAAGAATTATTCAGT	Amplification of yEGFP-SsrA for insertion into pMCP-P:EV through Golden Gate, forward
BG24827	GATCAGGTCTCACTTTTATG CTGCTAATGCATAATTTTCAT CATTGCTGCTTTGTACAATT CATCAATACCATG	Amplification of yEGFP-SsrA for insertion into pMCP-P:EV through Golden Gate, reverse
BG24825	GATCAGGTCTCAAGCATCTAA AGGTGAAGAATTATTCAGT	Amplification of yEGFP for insertion into pMCP-L:EV through Golden Gate, forward
BG24826	GATCAGGTCTCACTTTTATT TGTACAATTCATCAATACCAT G	Amplification of yEGFP for insertion into pMCP-L:EV through Golden Gate, reverse
BG24825	GATCAGGTCTCAAGCATCTAA AGGTGAAGAATTATTCAGT	Amplification of yEGFP-SsrA for insertion into pMCP-L:EV through Golden Gate, forward

BG24827	GATCAGGTCTCACTTTTATG CTGCTAATGCATAATTTTCAT CATTTGCTGCTTTGTACAATT CATCAATACCATG	Amplification of yEGFP-SsrA for insertion into pMCP-L:EV through Golden Gate, reverse
BG25559	GATCAGGTCTCAAGCATTTTT CACAAAAGTACTAAATAACC	Amplification of WanEat1 for insertion into pMCP-L:EV through Golden Gate, forward
BG22800	GATCAGGTCTCGCTTTTAAA CAGTGATTTCTTTGTTTTG	Amplification of WanEat1 for insertion into pMCP-L:EV through Golden Gate, reverse
BG25558	GATCAGGTCTCAAGCAAATG AAATCGATGAGAAAAATC	Amplification of SceAtf1 for insertion into pMCP-L:EV through Golden Gate, forward
BG22798	GATCAGGTCTCGCTTTCTAAG GGCCTAAAAGGAG	Amplification of SceAtf1 for insertion into pMCP-L:EV through Golden Gate, reverse
BG25560	GATCAGGTCTCAAGCAGAAA AAATAGAAGTGTCTATTAATT C	Amplification of HarmSAAT for insertion into pMCP-L:EV through Golden Gate, forward
BG25561	GATCAGGTCTCGCTTTCTATA TAAGAGTTTTTGGTGAAG	Amplification of HarmSAAT for insertion into pMCP-L:EV through Golden Gate, reverse

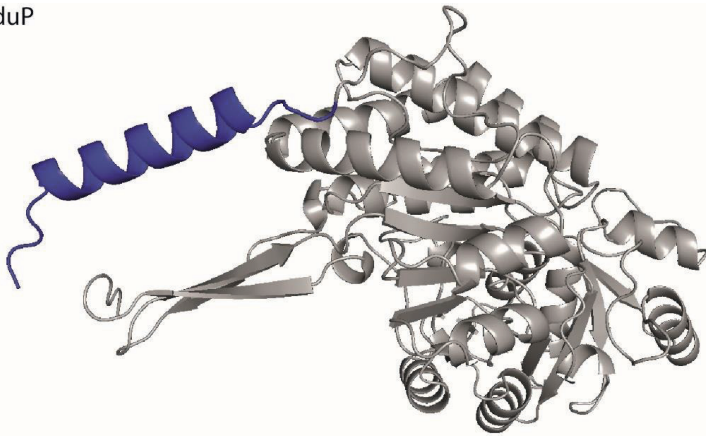
**Supplementary table 2.** *In vitro* Eat1 alcoholysis of ethyl propionate and 1-propanol to produce propyl propionate.

	Substrates		Product	Specific activity (U mg <sup>-1</sup> )
	Ethyl propionate (mM)	1-propanol (mM)	Propyl propionate (mM)	
<b>Initial</b>	2.5	2.5	-	-
<b>After reaction</b>	1.3	1.8	1.5	5.36

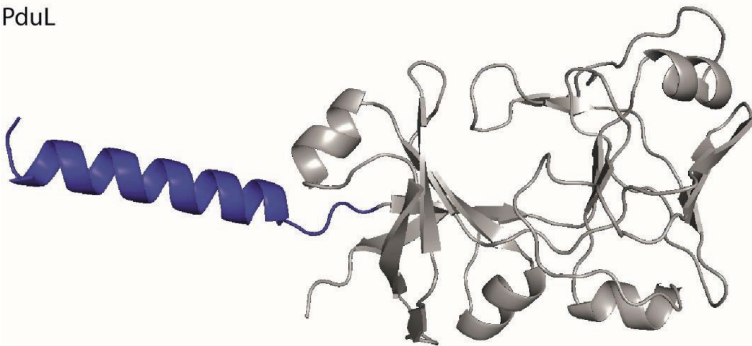
**Supplementary table 3.** The tmRNA sequence (5' to 3') is shown along with the predicted SsrA degradation tag. The tmRNA region that encodes for the degradation tag is underlined.

tmRNA (Cbei_R0066)	GGGGCGTTTTGGCTTCGACGGGGTAAGAAGGGTTTGTTAAGCGAGTCGAGGGA CGCATGGAGCCTCGTTAAAAAAGTATGCACTAAATGTAAACGCAGAAGATAATTT <u>TGCATTAGCAGCTTAGTTTAAATAGCTGTTTCATCAGCCAGGTTGCCTACGGCTTG</u> GATCACTGGTGTCAATTAAGTGGGAAACGAAGCCTAGCAAAGCTTTGAGCTAGAG GGGTATTTCATGAAGCTACTGAGAAGTATAGCCTGTCTAAGGGCGATACTTTGAGG GAATTTTAAACTTAGACTGCACTCGGAGAAAGATAAATCAATCTACTTTCGGAC ACGGGTTCGACACCCGTCGCCTCCACC
Predicted degradation tag	AEDNFALAA

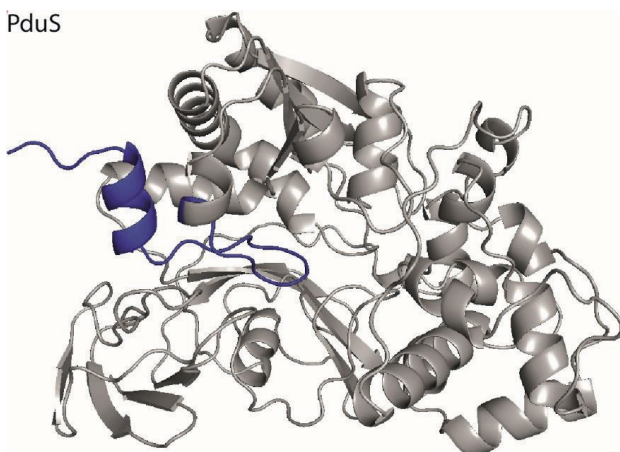
PduP



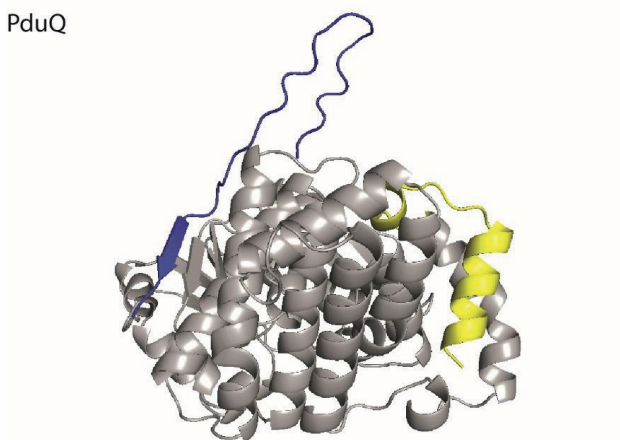
PduL



PduS

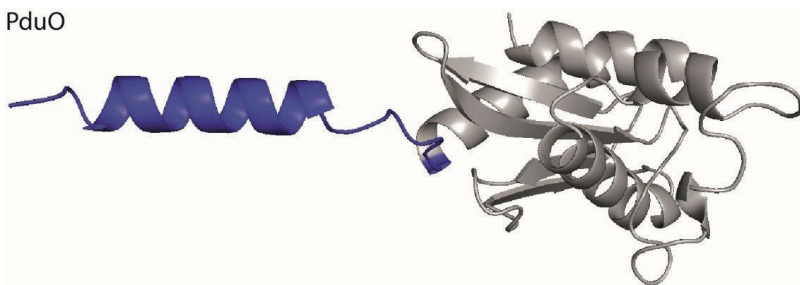


PduQ

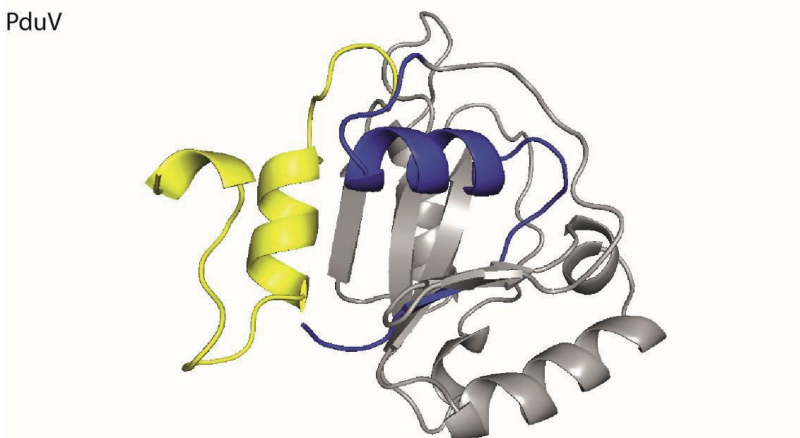


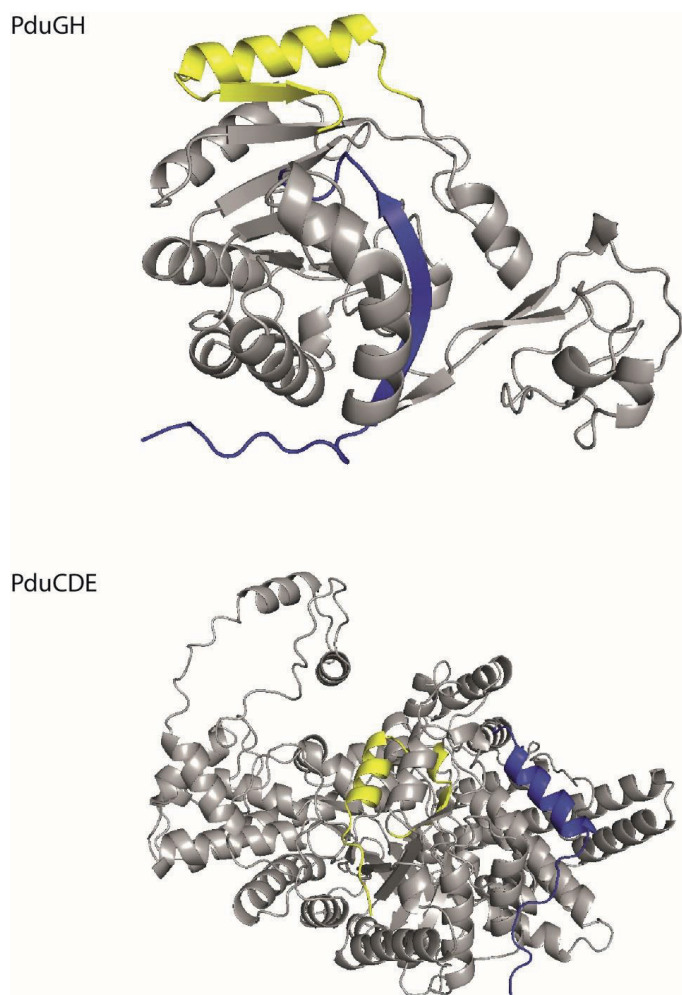
## Chapter 4

PduO



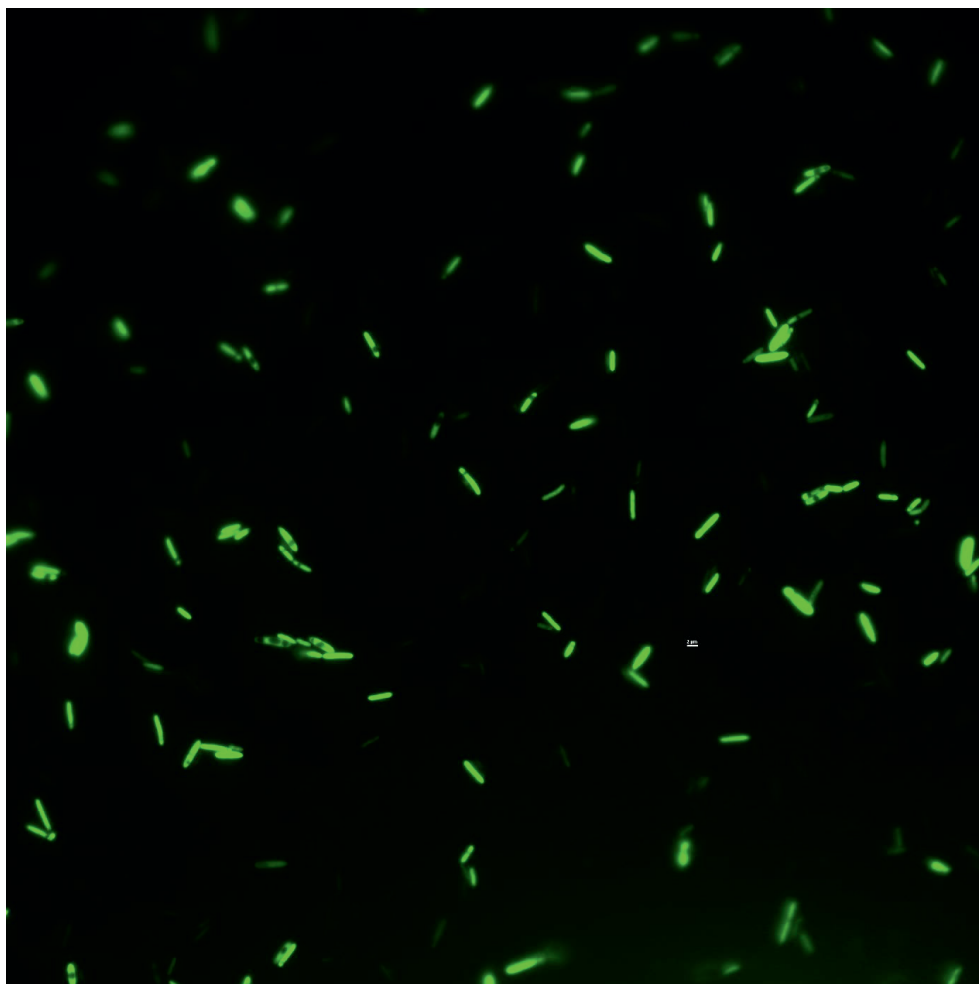
PduV



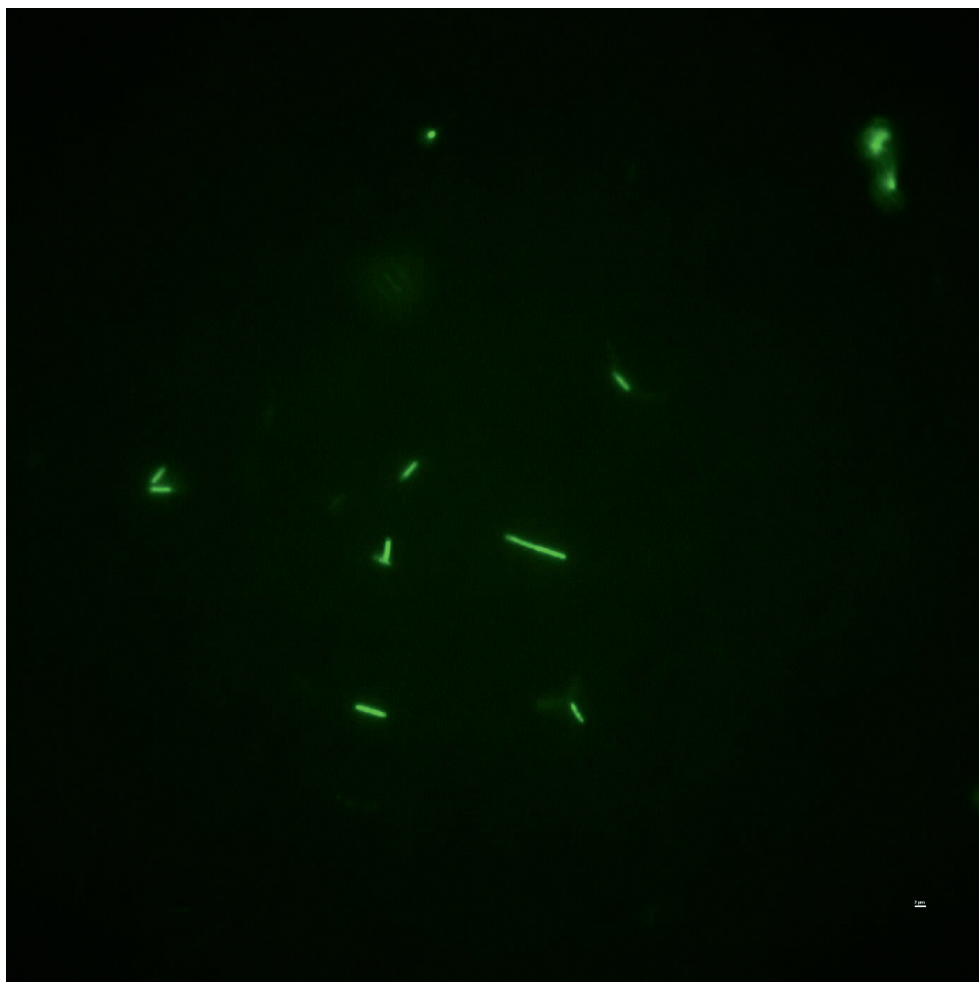


**Supplementary figure 1. Predicted 3D structures of the *C. beijerinckii* NCIMB 8052 BMC core enzymes.** The structures of the BMC core enzymes (PduL, PduP, PduS, PduV, PduO, PduQ, PduGH and PduCDE) were predicted using AlphaFold2 (391). The 25 first amino acids of each protein are highlighted with blue and the core protein is highlighted with grey. The C-terminus of PduQ, PduV, PduGH, PduCDE is highlighted with yellow to indicate the presence of an  $\alpha$ -helix at the C-terminus of the protein.

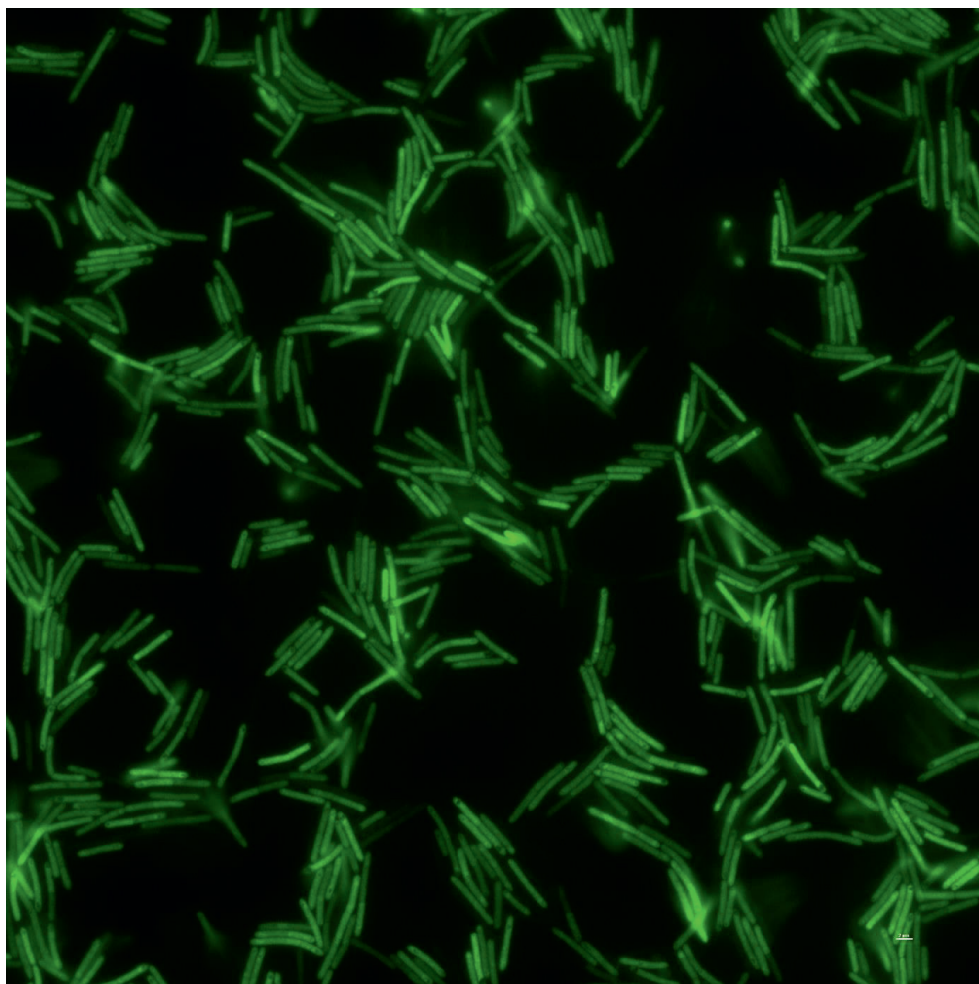




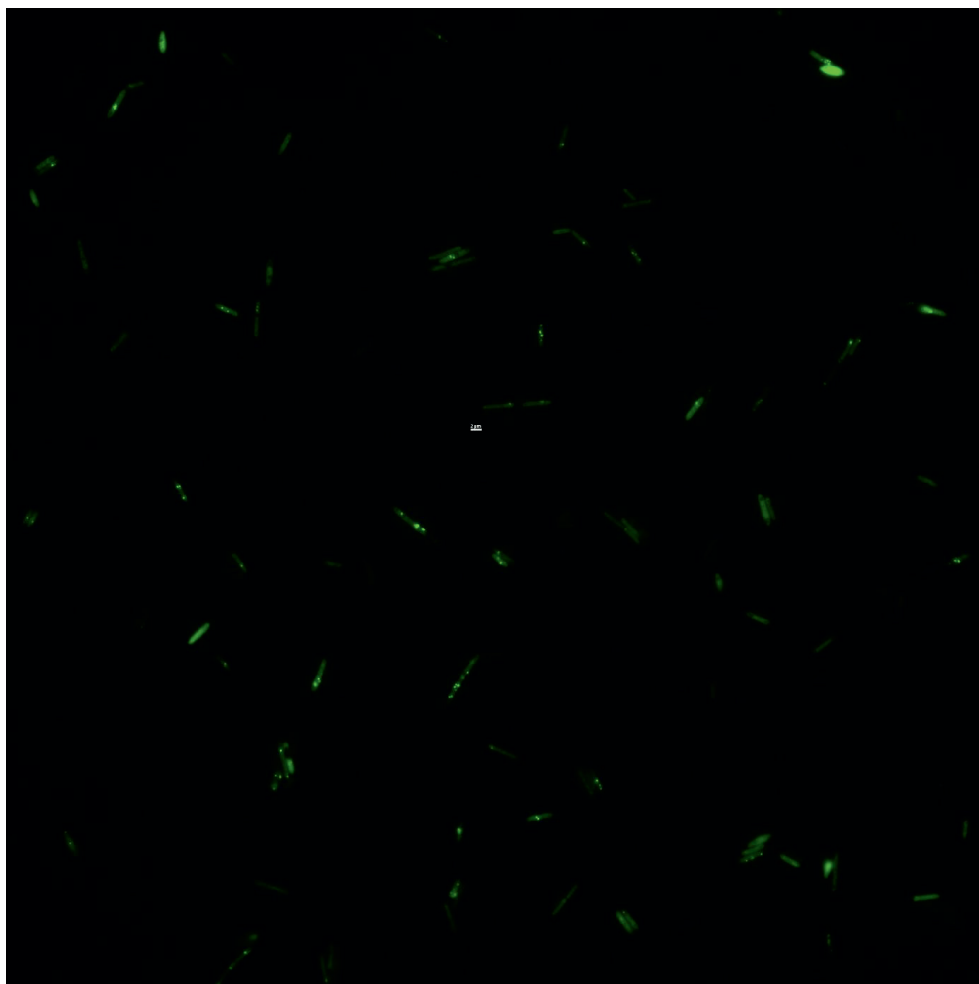
Supplementary figure 2. Raw picture of *C. beijerinckii* NCIMB 8052 carrying the yEGFP variant and grown in medium containing D-glucose as the main carbon source.



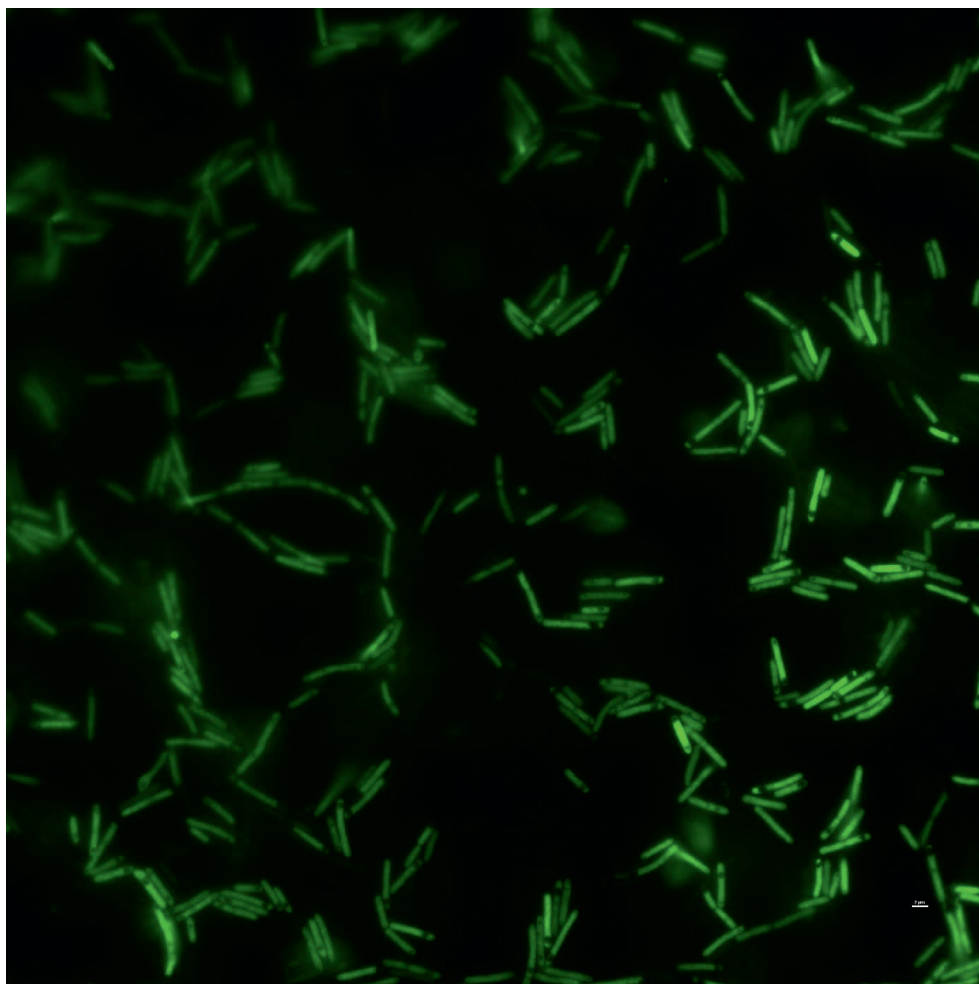
Supplementary figure 3. Raw picture of *C. beijerinckii* NCIMB 8052 carrying the yEGFP variant and grown in medium containing L-rhamnose as the main carbon source.



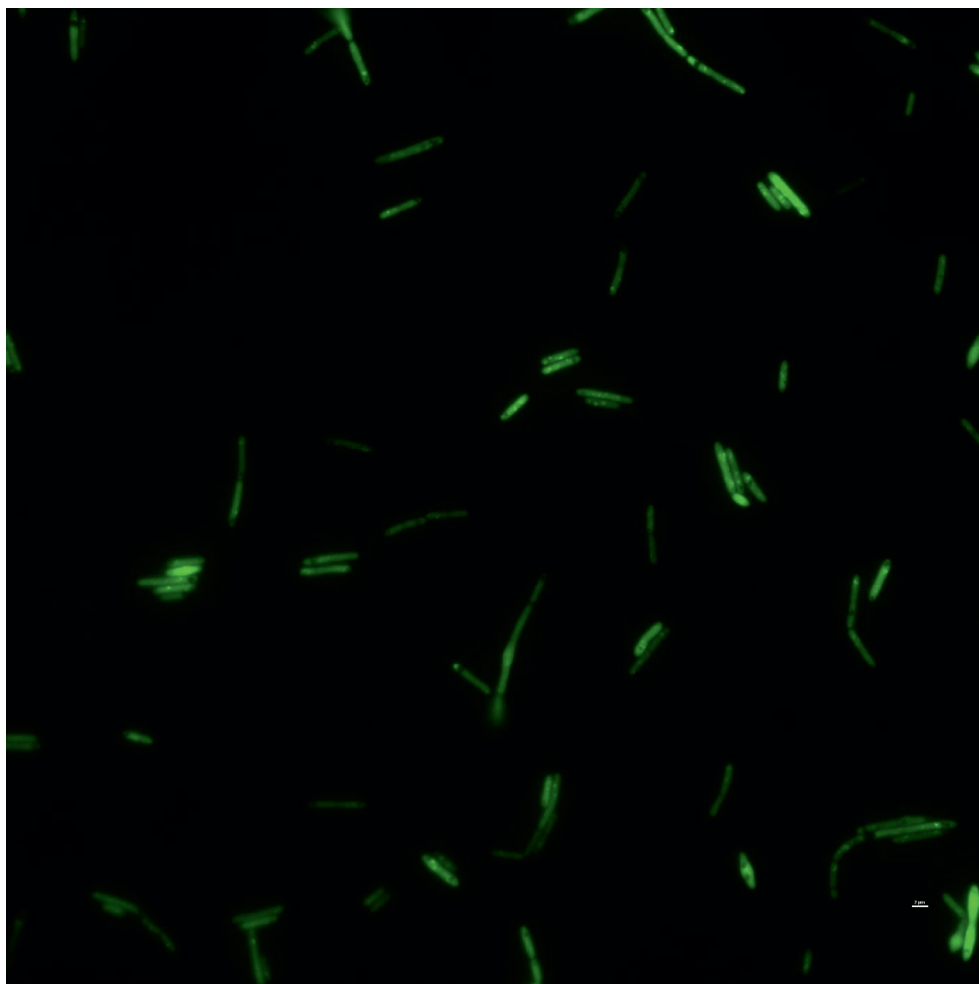
Supplementary figure 4. Raw picture of *C. beijerinckii* NCIMB 8052 carrying the L-yEGFP variant and grown in medium containing D-glucose as the main carbon source.



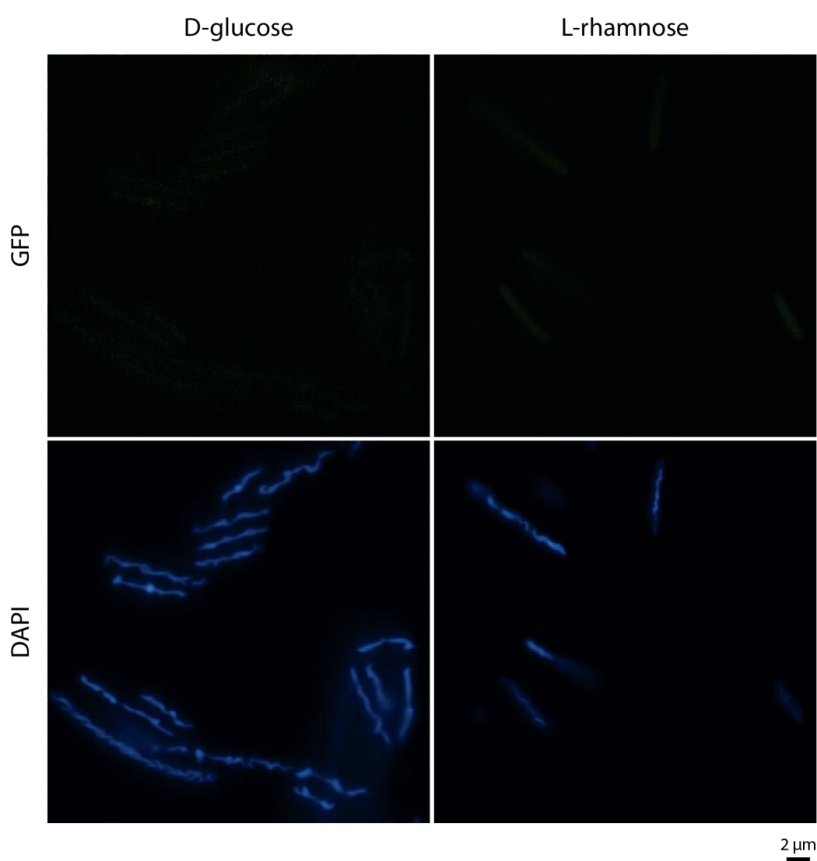
Supplementary figure 5. Raw picture of *C. beijerinckii* NCIMB 8052 carrying the L-yEGFP variant and grown in medium containing L-rhamnose as the main carbon source.



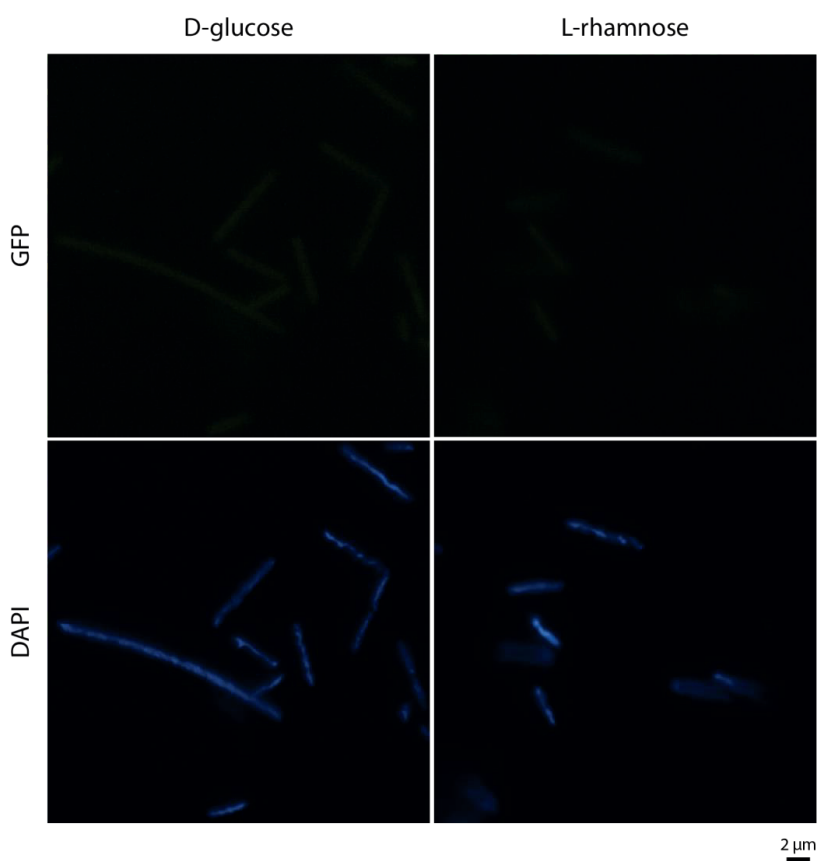
Supplementary figure 6. Raw picture of *C. beijerinckii* NCIMB 8052 carrying the P-yEGFP variant and grown in medium containing D-glucose as the main carbon source.



Supplementary figure 7. Raw picture of *C. beijerinckii* NCIMB 8052 carrying the P-yEGFP variant and grown in medium containing L-rhamnose as the main carbon source.

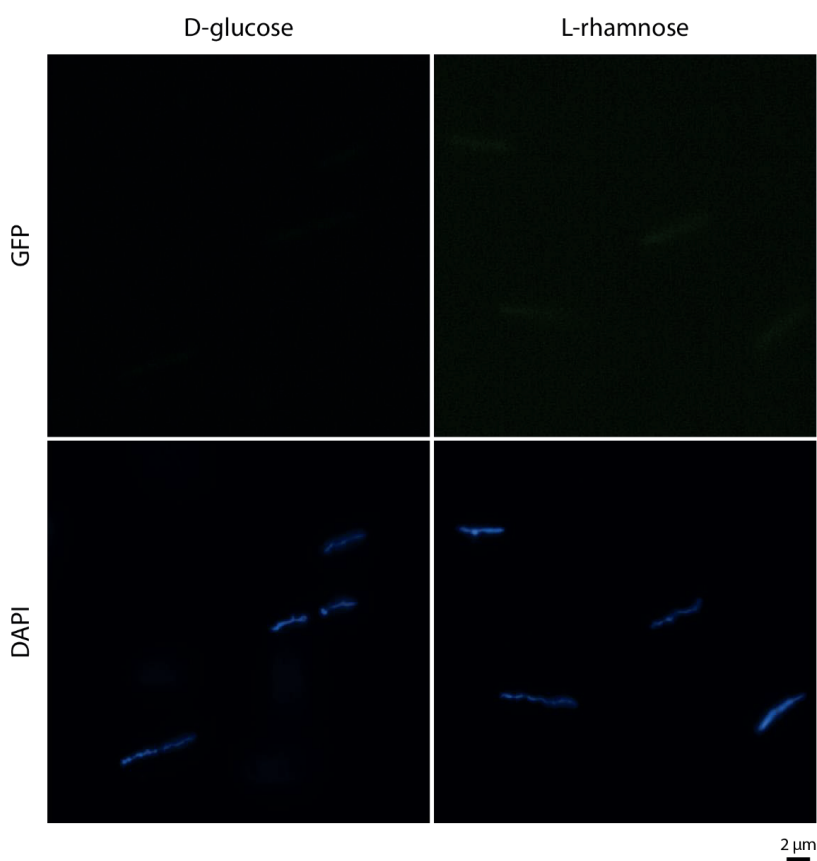


Supplementary figure 8. *C. beijerinckii* NCIMB 8052 carrying the yEGFP-SsrA variant and grown in medium containing D-glucose or L-rhamnose as the main carbon source. DAPI was used to stain the bacterial DNA.



Supplementary figure 9. *C. beijerinckii* NCIMB 8052 carrying the P-yEGFP-SsrA variant and grown in medium containing D-glucose or L-rhamnose as the main carbon source. DAPI was used to stain the bacterial DNA.

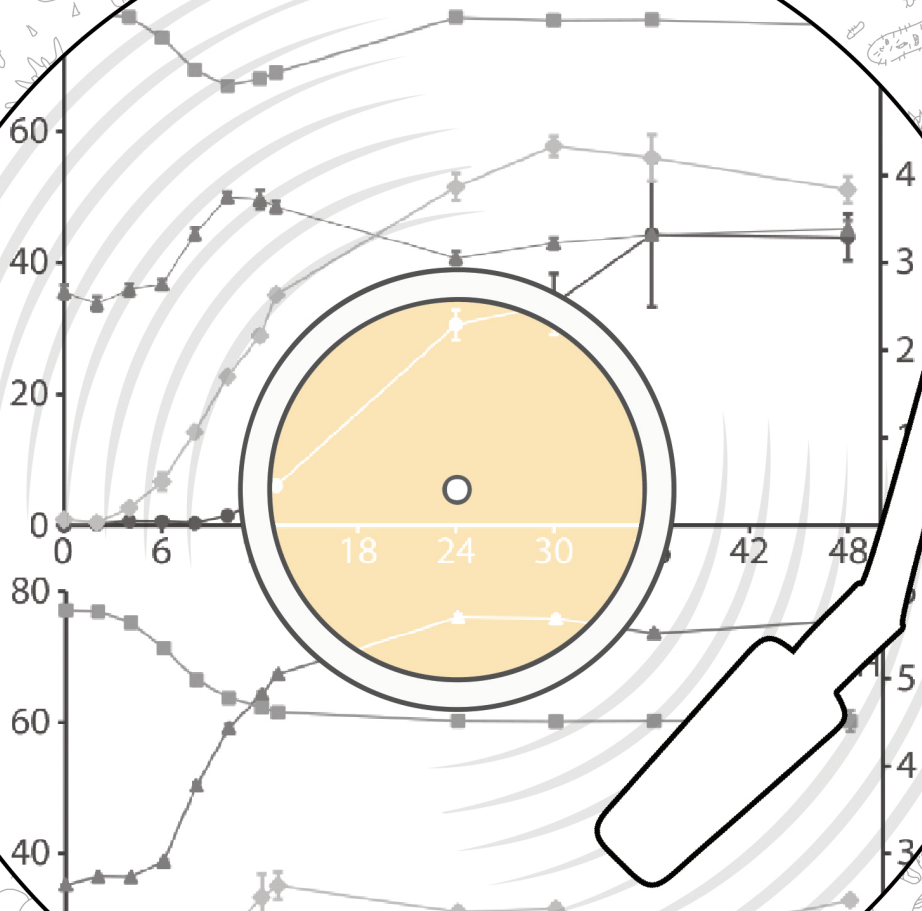




Supplementary figure 10. *C. beijerinckii* NCIMB 8052 carrying the L-yEGFP-SsrA variant and grown in medium containing D-glucose or L-rhamnose as the main carbon source. DAPI was used to stain the bacterial DNA.



# CHAPTER 5



## Chapter 5

# Multiplex genome engineering in *Clostridium beijerinckii* using CRISPR-Cas12a

Constantinos Patinios<sup>1,2,a</sup>, Stijn Tobias de Vries<sup>1,a</sup>, Mamou Diallo<sup>1,3,a</sup>, Lucrezia Lanza<sup>1</sup>, Pepijn L. J. V. Q. Verbrugge<sup>1</sup>, Ana M López-Contreras<sup>3</sup>, John van der Oost<sup>1</sup>, Ruud A Weusthuis<sup>2</sup>, Servé WM Kengen<sup>1</sup>

<sup>1</sup>Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands.

<sup>2</sup>Bioprocess Engineering, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands.

<sup>3</sup>Bioconversion Group, Wageningen Food and Biobased Research, Bornse Weilanden 9, 6708WG Wageningen, the Netherlands

<sup>a</sup>These authors have contributed equally to the work

Manuscript in preparation

### Abstract

*Clostridium* species are re-emerging as biotechnological workhorses for industrial acetone-butanol-ethanol production. This revival is largely due to major scientific advances in both fermentation technologies and in genome engineering. Re-programming of metabolic pathways by genome editing has been revolutionized by the development of a wide range of CRISPR-Cas tools. Here, we describe genome engineering by FnCas12a in *Clostridium beijerinckii* NCIMB 8052. By controlling the expression of FnCas12a with the strict xylose-inducible promoter, we achieved efficient (25-100%) single-gene knock-out of five *C. beijerinckii* genes (*Spo0A*, *Upp*, *Cbei\_1291*, *Cbei\_3238*, *Cbei\_3832*). Moreover, we used multiplex genome engineering to simultaneously knock-out the *Spo0A* and *Upp* genes in a single step with an efficiency of 18%. In agreement with earlier reports, we showed that the spacer sequence and its position in the CRISPR array can affect the editing efficiency outcome. Overall, we successfully developed an efficient genome editing system for the industrially relevant organism *Clostridium beijerinckii*.

## Introduction

*Clostridium beijerinckii* NCIMB 8052 is a gram-positive, spore-forming, anaerobic bacterium, that is a member of the acetone-butanol-ethanol (ABE) producing *Clostridium* species. ABE fermentation has a considerable industrial history as it played a major role during the first decades of the twentieth century. Especially during World War I, *Clostridium* played a key role in the production of acetone (acetone was used to produce cordite/gunpowder), butanol, lacquer solvents and jet fuel. In the period between 1920-1940, the production of ABE and related compounds by bacterial fermentation was outcompeted by petrochemical processes (20-23). Due to economic and environmental reasons and due to major advances in biotechnology, sustainable ABE fermentation (i.e. use of renewable sources) using *Clostridium* species has re-emerged (24).

To take advantage of the industrial potential of *Clostridium* species, several genome engineering tools have been developed (407). The genome engineering tools of *Clostridium* can be broadly divided into two: the ones which rely on homologous recombination (HR) based allelic exchange, and the ones which depend on group II intron-retargeting mutagenesis (TargeTron/ClosTron) (412,413). Whilst the group II intron-retargeting mutagenesis generally allows quick and efficient genome engineering, it has several disadvantages. These drawbacks include the inability to target genes smaller than 400 bp, it interrupts rather than deletes the target of interest and, the intron may be spliced back out by its associated intron-encoding protein (410). On the other hand, HR-based techniques enable the generation of scarless and complete deletion mutants and they do not rely on the integration of antibiotics in the genome of the target organism. For high accuracy, however, HR should be combined with an efficient counterselection mechanism such as CRISPR-Cas (61). Following this requirement, CRISPR-Cas in combination with HR has been used to achieve high editing efficiencies in various clostridia species as summarized by McAllister and Song (2019).

To date, several CRISPR-Cas-based tools have been developed for *C. beijerinckii* (65,67,406,408-410,414). Most of the developed CRISPR-Cas tools rely on the Cas9 nuclease and only a small fraction is based on the Cas12a nuclease. The preference towards Cas9 is probably due to a first-comer effect and due to its successful application in various organisms (415). However, Cas12a has distinctive advantageous features over Cas9 including smaller size (Cas12a: ~1300 a.a., Cas9: ~1400 a.a.) and recognition of a T-rich 5'-(T)TTV-3' PAM site at the 5' end of the protospacer sequence which increases the number of target sites in AT-rich organisms like clostridia (~30% GC-content) (416). In addition, unlike Cas9, Cas12a can process its own precursor-crRNA array due to the presence of a specific RNase domain, a feature that makes Cas12a ideal for multiplex genome engineering (416).

## Chapter 5

In this study, we used the FnCas12a nuclease to create single- and multi-gene deletions in *C. beijerinckii* NCIMB 8052. Depending on the target gene, single gene knock-out efficiencies varied from 25% to 100%. Multiplex deletion was also achieved by the simultaneous knock-out of two genes with an efficiency of 18%. Spacer sequence and position in the CRISPR array affected the multiplex knock-out efficiency, revealing potential limitations and room for improvement. This is the first report that demonstrates multiplex genome engineering in *C. beijerinckii* NCIMB 8052, hence contributing to the ongoing advances of genome editing of *Clostridium* and related organisms.

## Materials and methods

### Microbial strains and growth conditions

Table 1 shows all the strains used or generated in this study. *Escherichia coli* NEB® 5-alpha was used for plasmid assembly and cloning following the manufacturer's instructions (NEB). Transformed *E. coli* cells were grown at 37°C in LB liquid medium (10 g L<sup>-1</sup> tryptone, 5 g L<sup>-1</sup> yeast extract, 10 g L<sup>-1</sup> NaCl) or on LB agar plates (LB liquid medium, 15 g L<sup>-1</sup> bacteriological agar) containing spectinomycin (0.1 g L<sup>-1</sup>).

Transformed *Clostridium beijerinckii* NCIMB 8052 cells were grown anaerobically at 37°C in modified clostridial growth medium containing glucose as the main carbon source (mCGM-G: 5 g L<sup>-1</sup> yeast extract, 0.75 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.75 g L<sup>-1</sup> K<sub>2</sub>HPO<sub>4</sub>, 0.4 g L<sup>-1</sup> MgSO<sub>4</sub> · 7H<sub>2</sub>O, 0.01 g L<sup>-1</sup> MnSO<sub>4</sub> · H<sub>2</sub>O, 0.01 g L<sup>-1</sup> FeSO<sub>4</sub> · 7H<sub>2</sub>O, 1 g L<sup>-1</sup> NaCl, 2 g L<sup>-1</sup> L-asparagine, 2 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.125 g L<sup>-1</sup> L-cysteine, 13.753 g L<sup>-1</sup> D-(+)-glucose · H<sub>2</sub>O) or on mCGM-G agar (1 g L<sup>-1</sup> yeast extract, 2 g L<sup>-1</sup> tryptone, 0.5 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 1 g L<sup>-1</sup> K<sub>2</sub>HPO<sub>4</sub>, 0.1 g L<sup>-1</sup> MgSO<sub>4</sub> · 7 H<sub>2</sub>O, 0.01 g L<sup>-1</sup> MnSO<sub>4</sub> · H<sub>2</sub>O, 0.015 g L<sup>-1</sup> FeSO<sub>4</sub> · 7 H<sub>2</sub>O, 0.01 g L<sup>-1</sup> CaCl<sub>2</sub>, 0.002 g L<sup>-1</sup> CoCl<sub>2</sub>, 0.002 g L<sup>-1</sup> ZnSO<sub>4</sub>, 2 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 55 g L<sup>-1</sup> D-(+)-glucose · H<sub>2</sub>O, 12 g L<sup>-1</sup> agar) supplemented with 0.65 g L<sup>-1</sup> spectinomycin.

For fermentation assays and knock-out generation, *C. beijerinckii* NCIMB 8052 cells were grown in GAPES medium (2.5 g L<sup>-1</sup> yeast extract, 1 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.61 g L<sup>-1</sup> K<sub>2</sub>HPO<sub>4</sub>, 1 g L<sup>-1</sup> MgSO<sub>4</sub> · 7H<sub>2</sub>O, 0.0066 g L<sup>-1</sup> FeSO<sub>4</sub> · 7 H<sub>2</sub>O, 2.9 g L<sup>-1</sup> C<sub>2</sub>H<sub>7</sub>NO<sub>2</sub>, 0.19 g L<sup>-1</sup> pABA, 0.125 g L<sup>-1</sup> L-cysteine, 65.6 g L<sup>-1</sup> D-(+)-glucose · H<sub>2</sub>O) supplemented with 0.65 g L<sup>-1</sup> spectinomycin (325).

To induce the expression of *FnCas12a*, transformed *C. beijerinckii* NCIMB 8052 cells were plated on mCGM-X agar (containing 40 g L<sup>-1</sup> xylose instead of glucose as the carbon source) supplemented with 0.65 g L<sup>-1</sup> spectinomycin.



**Table 1. Strains used or generated in this study.**

Strain	Genotype	Source
<i>Escherichia coli</i> NEB® 5-alpha	<i>fhvA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	NEB
<i>Clostridium beijerinckii</i> NCIMB 8052	Wild type	(327)
<i>C. beijerinckii</i> NCIMB 8052 <i>ΔSpo0A</i>	<i>ΔSpo0A (Cbei_1712)</i>	This study
<i>C. beijerinckii</i> NCIMB 8052 <i>ΔUpp</i>	<i>ΔUpp (Cbei_0408)</i>	This study
<i>C. beijerinckii</i> NCIMB 8052 <i>ΔSpo0A, ΔUpp</i>	<i>ΔSpo0A (Cbei_1712), ΔUpp (Cbei_0408)</i>	This study
<i>C. beijerinckii</i> NCIMB 8052, <i>ΔCbei_1291</i>	<i>ΔCbei_1291</i>	This study
<i>C. beijerinckii</i> NCIMB 8052, <i>ΔCbei_3238</i>	<i>ΔCbei_3238</i>	This study
<i>C. beijerinckii</i> NCIMB 8052, <i>ΔCbei_3932</i>	<i>ΔCbei_3932</i>	This study

### Plasmid construction and transformation

The plasmids used in this study are shown in Table 2. Unless otherwise specified, all plasmids were assembled through NEBuilder® HiFi DNA Assembly (NEB). The basic backbone plasmid pCOMA\_NT was constructed by amplifying the pCB102 ori, colE1 ori and aad9 from pWUR100S (pS), FnCas12a from pY002 and the XylR-XylBP from pE\_X\_cas9. The crRNA was ordered as a synthetic DNA fragment (Twist Bioscience).

To introduce the homology arms into the pCOMA\_NT-crRNA plasmid series, 1 µg pCOMA\_NT-crRNA was linearized using AccI (NEB). The linear backbone was then dephosphorylated using shrimp alkaline phosphatase (rSAP, NEB) following the manufacturer's instructions. rSAP and residual AccI were deactivated by incubating the solution at 80°C for 20 min. The linearized backbone was further purified using the DNA clean and concentrator kit (Zymo Research). The homology arms were amplified by PCR using *C. beijerinckii* NCIMB 8052 genomic DNA as template and the oligonucleotides listed in Table S1. The correct size of the homology arms was confirmed through gel electrophoresis. Following, the amplicons were gel purified using the zymoclean gel DNA recovery kit (Zymo Research) and introduced into the linearized pCOMA\_NT-crRNA through NEBuilder® HiFi DNA Assembly (NEB) following the instructions from the manufacturer. 5 µL of the assembly was used to transform *E. coli* NEB® 5-alpha cells

(NEB). Transformed cells were plated on LB agar containing spectinomycin ( $0.1 \text{ g L}^{-1}$ ) and incubated at  $37^\circ\text{C}$  overnight. Obtained colonies were screened through PCR and the obtained plasmids were sequenced for the correctness of the homology arm insert using Sanger sequencing (Macrogen Europe B.V.).

Single or double targeting spacers were introduced through Golden Gate assembly using an adapted protocol (377). The “spacer insertion” and “spacer selection” protocols are detailed in the supplementary information. Briefly,  $1 \mu\text{L}$  of each of the two complementary oligonucleotides ( $100 \mu\text{M}$  each; Table S1) were mixed with  $1 \mu\text{L}$  NaCl ( $1 \text{ M}$ ) and  $47 \mu\text{L}$  MQ water and incubated at  $95^\circ\text{C}$  for 5 min. Following, the solution was allowed to cool down at room temperature for at least 2 h to achieve annealing of the complementary oligonucleotides. The annealed oligonucleotides were then diluted 10 times and  $2 \mu\text{L}$  of the diluted oligonucleotides was mixed with  $2 \mu\text{L}$  ( $0.01 - 0.02 \text{ pmol } \mu\text{L}^{-1}$ ) of the appropriate pCOMA\_NT-crRNA plasmid series and  $2 \mu\text{L}$  of MetaMix stock ( $10 \mu\text{L}$  BsaI-HF®v2,  $15 \mu\text{L}$  T4 ligation buffer,  $10 \mu\text{L}$  T4 ligase and  $15 \mu\text{L}$  MQ). The mix solution was then incubated in a thermocycler using the following protocol: 5 min at  $37^\circ\text{C}$ , 5 min at  $16^\circ\text{C}$  followed by 5 min at  $37^\circ\text{C}$  (repeat for 15-30 cycles), 5 min at  $37^\circ\text{C}$ , 20 min at  $80^\circ\text{C}$ .  $1 \mu\text{L}$  of the solution was then used to transform chemically competent *E. coli* NEB® 5-alpha (NEB) according to the instructions of the manufacturer. Transformed cells were plated on LB agar plates containing spectinomycin ( $0.1 \text{ g L}^{-1}$ ) and incubated overnight at  $37^\circ\text{C}$ . Obtained colonies were used for plasmid cloning by growing them in  $10 \text{ mL}$  LB medium containing spectinomycin ( $0.1 \text{ g L}^{-1}$ ), incubating overnight at  $37^\circ\text{C}$ . Plasmid purification was performed by using the GeneJET Plasmid Miniprep kit (ThermoFischer Scientific) and following the manufacturer’s instructions. To verify the spacer(s) sequence, plasmids were sequenced using Sanger sequencing (Macrogen Europe B.V.).

*C. beijerinckii* was transformed as previously described (417). In detail,  $100 \mu\text{L}$  of heat-shocked *C. beijerinckii* spores (1 min at  $99^\circ\text{C}$ ) were used to inoculate  $25 \text{ mL}$  of mCGM liquid medium followed by overnight incubation at  $37^\circ\text{C}$ .  $20 \text{ mL}$  of the overnight culture was used to inoculate  $180 \text{ mL}$  of pre-warmed ( $37^\circ\text{C}$ ) mCGM liquid medium which was then incubated at  $37^\circ\text{C}$  until an  $\text{OD}_{600}$  equal to 0.3-0.4 was reached. Following, in an anaerobic tent, the culture was transferred into a  $400\text{-mL}$  sterile centrifuge tube which was then sealed with parafilm to limit oxygen entrance. The culture was then centrifuged aerobically at  $6000 \text{ rpm}$  at  $4^\circ\text{C}$  for 10 min. The centrifuged culture was put on ice and transferred again in the anaerobic tent. The supernatant was then discarded and the cell pellet was resuspended with  $25 \text{ mL}$  of ice-cold anaerobic electroporation buffer ( $270 \text{ mM}$  D-sucrose,  $1 \text{ mM}$  sodium phosphate buffer pH 7.4,  $1 \text{ mM}$   $\text{MgCl}_2$ ). The resuspended culture was transferred into a  $30\text{-mL}$  sterile centrifuge tube which was then sealed with parafilm to limit oxygen entrance. The resuspended culture was then centrifuged aerobically at  $6000 \text{ rpm}$  at  $4^\circ\text{C}$  for 10 min

## Chapter 5

followed by discarding the supernatant and resuspending the cell pellet with 1.5 mL of ice-cold anaerobic electroporation buffer. 300 µL of the resuspended cells were used to electroporate (1.25 kV, 25 µF, 100 D) 3-5 µg plasmid DNA using 0.2 cm ice-cold electroporation cuvettes. After electroporation, the transformants were recovered at 37°C in 3 mL of anaerobic mCGM for 3-4 h. After recovery, cultures were centrifuged at 6000 rpm for 5 min and the cell pellet was plated on mCGM solid medium containing spectinomycin (0.65 g L<sup>-1</sup>) followed by incubation at 37°C for 72-96 h. Obtained colonies were screened for the presence of FnCas12a, the crRNA and the homology arms through colony PCR. Correct transformants were subcultured in mCGM and stored as vegetative cells in 20% glycerol at -80°C until use.

**Table 2. Plasmids used in this study.**

Plasmid name	Relevant characteristics	Reference
pY002	p15A ori, Lacp-WT FnCas12a, TetR, CmR	(416)
pWUR100S (pS)	pCB102 ori, colE1 ori, Aad9R	(418)
pE_X_cas9	colE1 ori, pAMβ1 ori, 2 µ ori, AmpR, ErmR, URA3	(418)
pCOMA_NT-crRNA	pCB102 ori, colE1 ori, ThIP-NT-crRNA-ThiT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R	This study
pCOMA_Spo0A-crRNA	pCB102 ori, colE1 ori, ThIP-Spo0A-crRNA-ThiT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R	This study
pCOMA_NT-crRNA_Spo0AHA	pCB102 ori, colE1 ori, ThIP-NT-crRNA-ThiT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Spo0A	This study
pCOMA_Spo0A-crRNA_Spo0AHA	pCB102 ori, colE1 ori, ThIP-Spo0A-crRNA-ThiT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Spo0A	This study
pCOMA_NT-crRNA_UppHA	pCB102 ori, colE1 ori, ThIP-NT-crRNA-ThiT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Upp	This study
pCOMA_Upp1-crRNA_UppHA	pCB102 ori, colE1 ori, ThIP-Upp1-crRNA-ThiT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Upp	This study
pCOMA_Cbei_1291-crRNA_Cbei_1291HA	pCB102 ori, colE1 ori, ThIP- Cbei_1291-crRNA-ThiT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Cbei_1291	This study

pCOMA_ Cbei_3238-crRNA_Cbei_3238HA	pCB102 ori, colE1 ori, ThlP- Cbei_3238-crRNA-ThlT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Cbei_3238	This study
pCOMA_ Cbei_3932-crRNA_Cbei_3932HA	pCB102 ori, colE1 ori, ThlP- Cbei_3932-crRNA-ThlT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Cbei_3932	This study
pCOMA_NT-crRNA_Spo0AHA_UppH A	pCB102 ori, colE1 ori, ThlP-NT-crRNA-ThlT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Spo0A, 500 bp homologous arms for Upp	This study
pCOMA_Spo0A-Upp1-crRNA_Spo0AHA_UppH A	pCB102 ori, colE1 ori, ThlP-Spo0A-Upp1-crRNA-ThlT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Spo0A, 500 bp homologous arms for Upp	This study
pCOMA_Upp1-Spo0A-crRNA_Spo0AHA_UppH A	pCB102 ori, colE1 ori, ThlP-Upp1-Spo0A-crRNA-ThlT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Spo0A, 500 bp homologous arms for Upp	This study
pCOMA_Spo0A-Upp2-crRNA_Spo0AHA_UppH A	pCB102 ori, colE1 ori, ThlP-Spo0A-Upp2-crRNA-ThlT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Spo0A, 500 bp homologous arms for Upp	This study
pCOMA_Upp2-Spo0A-crRNA_Spo0AHA_UppH A	pCB102 ori, colE1 ori, ThlP-Upp2-Spo0A-crRNA-ThlT, XylRP-XylR, XylBP-FnCas12a-FdxT, Aad9R, 500 bp homologous arms for Spo0A, 500 bp homologous arms for Upp	This study

### *C. beijerinckii* NCIMB 8052 knock-out generation

*C. beijerinckii* cells transformed with the pCOMA plasmid series were grown in 25 mL selective GAPES medium for 48-96 h to allow for homologous recombination to occur. 100 µL of the fully grown culture was then plated on selective mCGM-X agar plates to induce expression of *FnCas12a* and allow for the counterselection of the mutants. Agar plates were incubated for 24-48 h at 37°C and obtained colonies were screened through colony PCR. To perform colony PCR, colonies were resuspended in 50 µL PBS buffer (pH 7.4) and boiled for 10 min at 99°C. 1 µL of the boiled solution was then used as template for PCR using the Q5® High-Fidelity DNA Polymerase (NEB), following the manufacturer's instructions. The primers used for colony PCR are listed in Table S1. Lastly, to confirm the complete knock-out of the gene of interest, PCR amplicons were sequenced through Sanger sequencing (Macrogen). Each knock-out experiment was performed in triplicate and the average knock-out efficiency was calculated by defining the percentage of clean mutants (i.e. no mixed bands) versus wild type and mix bands.

### Plasmid curing

To cure the *C. beijerinckii* NCIMB 8052 knockout strains of the pCOMA plasmids, the cells were grown in 25 mL mCGM-G liquid medium without antibiotics for 24 h. 100  $\mu$ L of the grown culture was then plated on mCGM-G agar plates without the presence of antibiotics and grown for 24 h at 37°C. Obtained colonies were randomly selected and streaked out on one mCGM-G agar plate with antibiotics and on one mCGM-G agar plate without antibiotics and incubated for 24 h at 37°C. Colonies that did not grow on selective medium but grew on non-selective medium were selected and screened for the absence of plasmid through colony PCR. Mutant colonies which lost the respective pCOMA plasmid were grown in GAPES medium without antibiotics and glycerol stocks were made and stored at -80°C until further use.

### $\Delta Spo0A$ and WT *C. beijerinckii* NCIMB 8052 fermentation assays and morphology

The WT and  $\Delta Spo0A$  *C. beijerinckii* NCIMB 8052 strains were grown in GAPES medium without antibiotics for 48 h. At different time intervals, 1 mL of headspace was recovered and the solvent concentration was determined using gas chromatography (GC). 1 mL of liquid culture was also recovered, of which the pH, OD<sub>600</sub> and organic acid concentrations were determined using high-pressure liquid chromatography (HPLC).

A Shimadzu GC-2010 equipped with an Agilent technologies DB-WAX UI GC column (30 m x 0.53 mm) using a temperature gradient of 60-125°C over 10 min and a nitrogen flow rate of 115 mL min<sup>-1</sup> was used to separate metabolites. A split ratio of 20 and a carrier flow program with a constant pressure of 30 kPa was applied. References of GAPES medium containing 100, 50, 20 and 5 mM of acetone, ethanol and butanol were used to create a calibration curve. As internal standard, 5 mM of 1-propanol was used.

For HPLC, a Shimadzu LC-2030 with a Shimadzu RID-20A detector was used. To separate the metabolites, a Shodex SUGAR SH1821 column was operated at 45°C with a flow rate of 0.8 mL min<sup>-1</sup> and a flow time of 20 mins. 0.01N H<sub>2</sub>SO<sub>4</sub> was used as eluent. References of 100; 50; 25; 12.5; 6.25; 3.125, 1.5625 and 0.78125 mM of lactate, acetate and butyrate were used to create a calibration curve. As internal standard, 5 mM of crotonate was used.

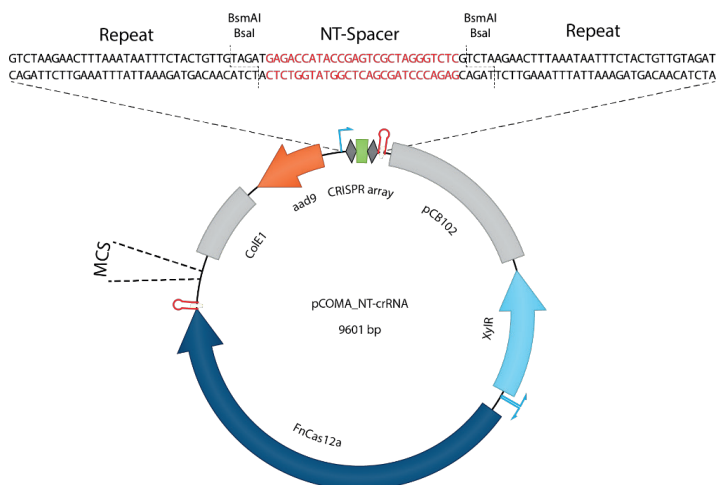
Pictures of WT and  $\Delta Spo0A$  *C. beijerinckii* NCIMB 8052 colonies were taken on mCGM agar plates using Carl Zeiss Axio Scope.A1, 100 x total magnification, phase 1.

## Results and discussion

### Markerless deletion of *Spo0A* through inducible expression of FnCas12a

Several CRISPR-Cas tools have previously been developed for genome engineering of *Clostridium beijerinckii* strains (65,67,406-410). However, multiplex gene editing in *Clostridium beijerinckii* has never been demonstrated. We sought to develop an easy-to-use and efficient genome editing tool for multiplex genome engineering in the strain *C. beijerinckii* NCIMB 8052 by using CRISPR-Cas12a. Cas12a was chosen as the appropriate Cas protein due to its ability to generate its own CRISPR-derived crRNA guides, which substantially facilitates multiplex gene targeting (416).

A previous report showed successful single-gene genome engineering of *C. beijerinckii* NCIMB 8052 using AsCas12a (65). The AsCas12a used in this study was derived from pDEST-hisMBP-AsCpf1-EC (Addgene plasmid #79007) which has a codon optimized nucleotide sequence for *E. coli*. Since *C. beijerinckii* NCIMB 8052 and *E. coli* differ in genomic GC-content (30% versus 51%, respectively), we reasoned that we should use a *Cas12a* gene that matches the codon usage of *C. beijerinckii* NCIMB 8052 and also follows the same translation speed and fidelity (Fig. S1) (419,420). To this end, we have chosen the wild type FnCas12a nuclease based on its low GC percentage (30%) and matching codon usage for *C. beijerinckii* NCIMB 8052 (Fig. S1).



**Figure 1. Backbone of the pCOMA plasmid series.** At the top of the plasmid the CRISPR array is shown, where a non-targeting (NT) spacer can be conveniently replaced through Golden Gate. Homologous arms for directed recombination can be inserted at the multiple cloning site (MCS) through Gibson Assembly.

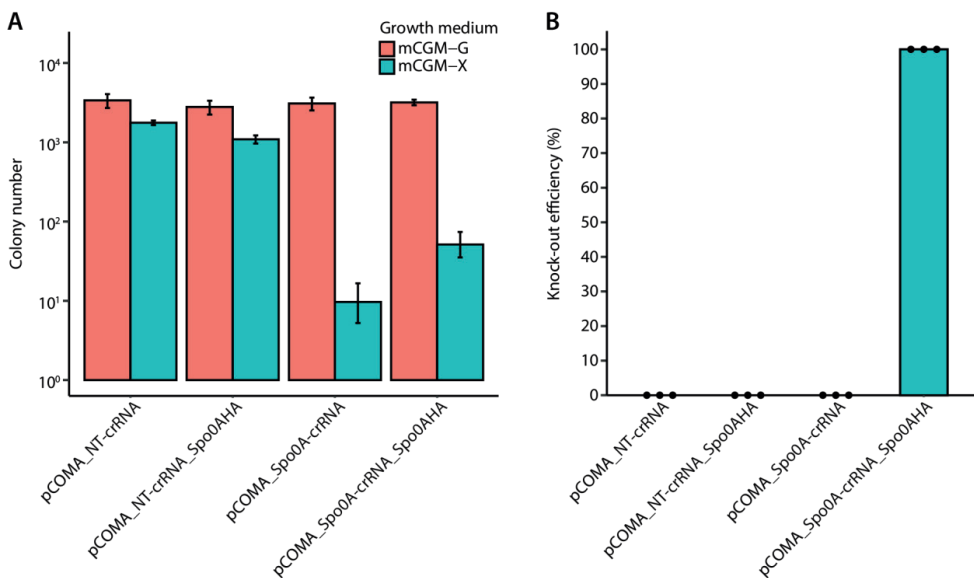
To develop a simple genome engineering tool for *C. beijerinckii* NCIMB 8052, a single plasmid approach was used containing the *FnCas12a* gene, the CRISPR array (repeat-spacer-repeat) with

an insertion site for easy exchange of the spacer through Golden Gate cloning, and a multiple cloning site (MCS) to insert the homology arms to facilitate directed recombination in general (Fig. 1), including gene knock-out as described below. To control the expression of FnCas12a and to avoid potential cell toxicity (66,406,421-423) we chose to use the xylose-inducible promoter derived from *C. difficile* based on its successful application in *C. beijerinckii* (67,424,425). Moreover, the use of the xylose-inducible system has a dual function in our setup as it can serve both as the inducer molecule for the expression of FnCas12a but also as carbon- and energy-source for growth. Therefore the use of glucose in the growth medium can be omitted and the effect of potential catabolite repression can be avoided (425). To express the crRNA, we chose to use the strong constitutive thiolase promoter (ThlP) of *C. beijerinckii*.

As a proof of principle, we selected the well-characterized *Spo0A* (Cbei\_1712) gene as a knock-out target. The  $\Delta Spo0A$  strain has a distinctive morphological and metabolite production phenotype, making identification of *Spo0A* knock-outs straightforward (30,426). To this end, plasmids pCOMA\_NT-crRNA (non-targeting control), pCOMA\_Spo0A-crRNA (targeting control), pCOMA\_NT-crRNA\_Spo0AHA (non-targeting with homology arms) and pCOMA\_Spo0A-crRNA\_Spo0AHA (targeting control with homology arms) were constructed and transformed to *C. beijerinckii* NCIMB 8052 cells. Selected transformants were grown for 48 h in mCGM-G liquid medium, after which 100  $\mu$ L of the culture were plated on mCGM-G solid medium (as a control) and 100  $\mu$ L of the culture on mCGM-X solid medium to induce the expression of FnCas12a and, hence, counter-select for mutants.

Transformants plated on mCGM-G solid medium showed comparable numbers of colonies (approximately  $10^3$ ), regardless of the transformed plasmid (Fig. 2A). Similarly, transformants carrying a non-targeting spacer, led to approximately  $10^3$  colonies when plated on mCGM-X. In contrast, transformants carrying a targeting spacer and plated on mCGM-X showed fewer colonies compared to the non-targeting controls. As expected, the pCOMA\_Spo0A-crRNA targeting control resulted in very few colonies ( $\sim 10$ ), which shows the functionality of FnCas12a to successfully target and cleave the genome of *C. beijerinckii* NCIMB 8052. The presence of a small number of colonies may be attributed to PAM, spacer, protospacer, crRNA or FnCas12a mutants that could escape the counter-selective properties of FnCas12a. A reduced number of colonies ( $\sim 50$ ) compared to the non-targeting controls, but higher than the pCOMA\_Spo0A-crRNA targeting control was observed when using the pCOMA\_Spo0A-crRNA\_Spo0AHA targeting control with homology arms.

To assess whether the colonies obtained on mCGM-X were successful *Spo0A* knock-outs, we screened colonies from each biological replicate (24 in total) by colony PCR. As expected, the non-targeting controls showed a 0% knock-out efficiency represented by a WT genotype of 1866 bp amplicons (Fig. 2B and S2). Similarly, the obtained pCOMA\_Spo0A-crRNA colonies showed a WT genotype (0% knock-out efficiency), further supporting that the obtained colonies indeed are escapees. Intriguingly, obtained colonies containing the pCOMA\_Spo0A-crRNA\_Spo0AHA had a 100% knock-out efficiency with a  $\Delta Spo0A$  genotype corresponding to 1044 bp amplicons (Fig. 2B, Fig. S2). Mutant colonies were further checked for the deletion of *Spo0A* by Sanger sequencing, confirming the scarless deletion of *Spo0A*.



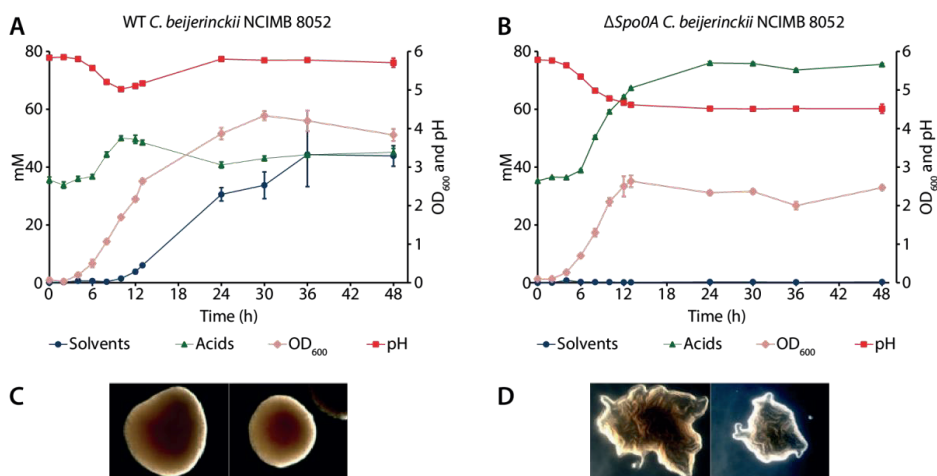
**Figure 2. Inducible counterselection and *Spo0A* knock-out in *C. beijerinckii* NCIMB 8052.** *C. beijerinckii* NCIMB 8052 cells were transformed either with the pCOMA\_NT-crRNA and pCOMA\_NT-crRNA\_Spo0AHA non-targeting plasmids (negative controls), the pCOMA\_Spo0A-crRNA targeting plasmid (positive control) or the pCOMA\_Spo0A-crRNA\_Spo0AHA plasmid. Transformants were plated on mCGM-G (no-induction) or on mCGM-X (induction of FnCas12a). This experiment was performed in biological triplicates. The error bars in (A) show the standard deviation. (A) Colony number obtained after plating the transformants on the appropriate medium. (B) *Spo0A* knock-out efficiency determined by screening eight colonies from each replicate. Dots represent the knock-out efficiency from each replicate.

### **$\Delta Spo0A$ *C. beijerinckii* NCIMB 8052 shows retarded growth, elimination of solvent production and increased production of acids**

As previously described (30,426),  $\Delta Spo0A$  *C. beijerinckii* strains show a distinctive phenotype which includes retarded growth, the elimination of solvent production, increased acid production



and altered colony morphology. To assess whether our  $\Delta Spo0A$  *C. beijerinckii* NCIMB 8052 mutants show the described phenotypic characteristics, we selected three  $\Delta Spo0A$  colonies and subjected them to plasmid curing. Three cured  $\Delta Spo0A$  *C. beijerinckii* NCIMB 8052 colonies and three WT *C. beijerinckii* NCIMB 8052 colonies were then grown in GAPES medium for 48 h and the fermentation products were analysed (Fig. 3A and B).



**Figure 3. Fermentation profile and morphology of WT and  $\Delta Spo0A$  *C. beijerinckii* NCIMB 8052 strains. (A)** Fermentation profile of WT *C. beijerinckii* NCIMB 8052. **(B)** Fermentation profile of  $\Delta Spo0A$  *C. beijerinckii* NCIMB 8052. The error bars in (A) and (B) indicate the standard deviation. Solvents represent acetone, butanol and ethanol. Acids represent acetate, butyrate and lactate. **(C)** Colony morphology of WT *C. beijerinckii* NCIMB 8052. **(D)** Colony morphology of  $\Delta Spo0A$  *C. beijerinckii* NCIMB 8052.

During the first 8 h of incubation, the  $\Delta Spo0A$  and WT strains performed similarly, as the acids were produced to equimolar amounts (approx. 50 mM), accompanied with the characteristic pH drop from pH 6.0 to around pH 5.0. However, after 8 h of incubation, the growth and product formation of the  $\Delta Spo0A$  and WT strains differed considerably. The WT strain ceased the production of acids and the production of solvents was initiated, raising the pH to around 5.8. The  $OD_{600}$  increased to 4.33 at 30 h of incubation and it decreased to 3.83 at the end of the fermentation. Solvents and acids reached a concentration of 43.85 mM and 45.11 mM, respectively, at the end of fermentation. In contrast, the  $\Delta Spo0A$  strain did not assimilate the produced acids after 8 h of growth, resulting in an increased production of acids and the absence of solvent production. Due to the lack of acid assimilation, the pH remained low, reaching a pH of 4.51 at the end of the fermentation. The  $OD_{600}$  did not increase after 12 h of fermentation, likely due to the acid crash caused by the increased acid production. The acids reached a concentration of 75.5 mM at the end of the fermentation and no solvents were produced. Lastly, the distinct, irregular colony

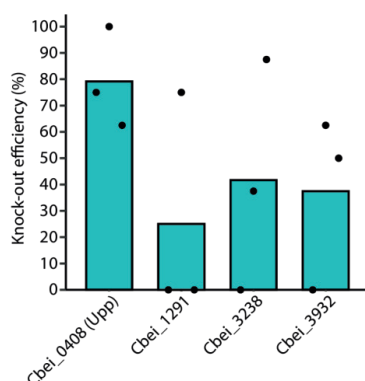
morphology was apparent for the  $\Delta Spo0A$  colonies whereas the WT colonies showed the typical round and smooth shape (Fig. 3C and D).

### Establishing single deletions of various *C. beijerinckii* NCIMB 8052 genes

To assess the applicability and knock-out efficiency of our tool to other genes (other than the *Spo0A*), we sought to delete four genes at different genomic loci: *Cbei\_0408* (*Upp*; 630 bp), *Cbei\_1291* (987 bp), *Cbei\_3238* (888 bp) and *Cbei\_3932* (813 bp). An identical protocol as the one described for *Spo0A* deletion was followed, and obtained colonies were screened for mutants through colony PCR (amplicon size) and Sanger sequencing.

Compared to the 100% *Spo0A* knock-out generation, the knock-out efficiency varied amongst the selected genes. The highest knock-out efficiency was observed for *Upp* (79%), followed by *Cbei\_3238* (42%), *Cbei\_3932* (38%) and *Cbei\_1291* (25%) (Fig. 4 and S3). More intriguingly, we clearly observed different knock-out efficiencies between the biological replicates of transformants carrying (essentially) the same plasmid variant. For example, for *Cbei\_1291*, two out of the three biological replicates showed 0% knock-out efficiency, whereas one of the replicates showed 75% knock-out efficiency (Fig. S3). The inconsistency in editing amongst the replicates was also observed for *Cbei\_3238* and *Cbei\_3932* where at least one of the replicates showed 0% knock-out efficiency. Nonetheless, one of the replicates for *Cbei\_3238* showed 88% knock-out efficiency and one of the replicates for *Cbei\_3932* showed 63% knock-out efficiency. Knocking-out the *Upp* gene was more consistent as biological replicates varied between 63 and 100% knock-out efficiency.

The variable knock-out efficiency amongst the different genes may be attributed to the selection of a good or bad spacer (427) which is often attributed to the secondary structure of the crRNA, the GC content of the target and the spacer, and the melting temperature of the spacer-protospacer pairing (428,429). However, the variable knock-out efficiency amongst the biological replicates targeting the same gene could be due to early escapees which dominated the culture during the 48 h growth before inducing the expression of FnCas12a for counterselection. Most likely, escapees survived counterselection because of mutations at the spacer or protospacer, at the PAM sequence or the *FnCas12* gene sequence.



**Figure 4. Single-gene knock-out of multiple genes in *C. beijerinckii* NCIMB 8052.** *Cbei\_0408 (Upp)*, *Cbei\_1291*, *Cbei\_3238* and *Cbei\_3932* were targeted for knock-out. The average knock-out efficiency for each gene is: *Cbei\_0408* (79.17%), *Cbei\_1291* (25%), *Cbei\_3238* (41.67%) and *Cbei\_3932* (37.5%). This experiment was performed in biological triplicates. The knock-out efficiency was determined by screening eight colonies from each replicate through colony PCR using the primers listed in Table S1. Dots represent the knock-out efficiency from each replicate.

### Multiplex gene knock-out in *C. beijerinckii* NCIMB 8052

Single-gene knock-out in *C. beijerinckii* NCIMB 8052 was previously achieved using CRISPR-AsCas12a (65). However, multiplex genome editing in *C. beijerinckii* has never been reported before. To establish a multiplex approach to simultaneously knock-out several genes in *C. beijerinckii* NCIMB 8052, we set out to disrupt the *Spo0A* and *Upp* genes in a single step (i.e., in one transformation event) using the established xylose-inducible system, as described above.

To target two genomic loci (*Spo0A* and *Upp*) with FnCas12a, two spacers (one for each target) were introduced into a synthetic CRISPR array (Table S1). Since the spacer sequence can affect the editing efficiency (430), we constructed two CRISPR arrays, one in which the *Spo0A* spacer preceded the *Upp1* spacer (*Spo0A-Upp1*) and one with a reversed order (*Upp1-Spo0A*). In addition, to further assess the effect of changing one of the targeting spacers with another spacer that targets the same gene but at a different site (protospacer), we replaced the *Upp1* spacer with the *Upp2* spacer. Similarly, two variant CRISPR arrays were constructed (*Spo0A-Upp2*, *Upp2-Spo0A*) (Table S1).

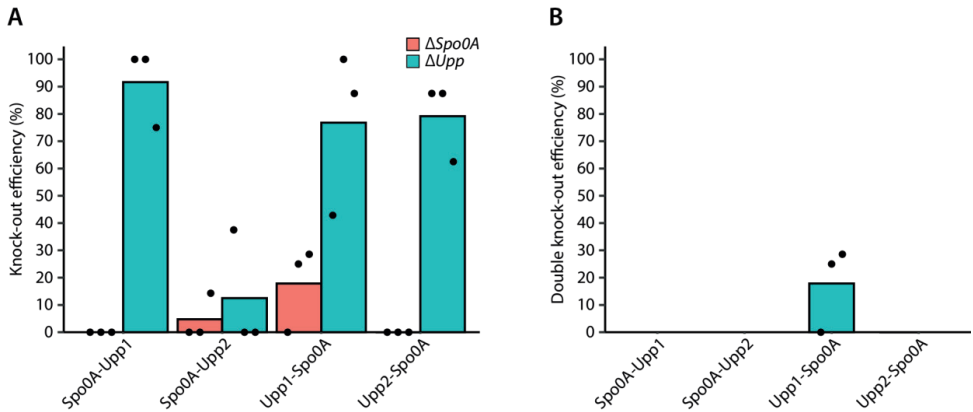
As expected, knock-out efficiencies varied between the different CRISPR array variants (Fig. 5). The highest knock-out efficiency for *Upp* (91%) was observed when the *Spo0A-Upp1* CRISPR array was used, whereas the lowest knock-out efficiency (13%) was observed when the *Spo0A-Upp2* CRISPR array was used. While the *Spo0A-Upp1* CRISPR array showed the highest knock-out efficiency for *Upp*, 0% knock-out efficiency was observed for *Spo0A*. In contrast, by switching

the position of the Spo0A and Upp1 spacers (i.e., from Spo0A-Upp1 to Upp1-Spo0A), 18% knock-out efficiency was observed for *Spo0A*. The knockout efficiency of *Upp* dropped to 77% when the Upp1-Spo0A CRISPR array was used, indicating that the position of the Upp1 spacer in the array does not affect largely the knock-out efficiency of *Upp*. Moreover, the knock-out efficiency of *Upp* was reduced from 92% to 13% when the Upp1 spacer was substituted with the Upp2 spacer and when the Spo0A spacer was the first spacer of the CRISPR array. However, the substitution of Upp1 to Upp2 yielded some (5/24) successful Spo0A knock-outs, although most (4/5) of them were mixed colonies, as indicated in Fig. S4. Important to note is that half (12/24) of the screened colonies had a mixed (WT and  $\Delta Upp$ ) genotype.

In summary, we achieved multiplex gene knock-out in *C. beijerinckii* NCIMB 8052, although with a relatively low editing efficiency. Clean (i.e. without the presence of mixed colonies) double *Spo0A* and *Upp* mutants were observed only when the Upp1-Spo0A CRISPR array was used. This observation is in agreement with a previous report by Liao *et al.* (2019) (430), in which it is clearly demonstrated that the abundance of crRNAs guides largely varies in different CRISPR array designs. The variation in the abundance of crRNAs (i.e. crRNA guides bound to the Cas nuclease) most likely correlates directly with the targeting efficiency of that guide. An important determinant of a crRNA guide's efficiency is its secondary structures, that in the worst case can inhibit the formation of the characteristic hairpin that is required for guide binding to the Cas12a effector, for guide maturation and eventually for specific targeting activity (416,430). To assess this possibility, we predicted the secondary structure of the transcribed pre-crRNAs using NUPACK (431).

As expected, complex secondary structures were formed in all the pre-crRNAs (Fig. S5). In the CRISPR arrays where the Spo0A spacer preceded the Upp spacer, an undisrupted hairpin was formed between nucleotide 22 and 35 of the pre-crRNA, required for the successful recognition and processing of the Spo0A crRNA by FnCas12a. However, a secondary structure was observed between the nucleotides of the Spo0A spacer, although with low equilibrium probability (Fig. S5). The secondary structure formed by the Spo0A spacer may reflect the low knock-out efficiency observed in all the multiplex editing assays, rendering the Spo0A crRNA as a poorly performing crRNA (427). Yet, 100% knock-out efficiency was observed when the Spo0A crRNA was used for single *Spo0A* knock-outs (Fig. 2B). In contrast to the Spo0A spacer, the necessary hairpin for processing the Upp1 or Upp2 crRNA was disturbed in the Spo0A-Upp1 and Spo0A-Upp2 arrays. Still, our results for both the Spo0A-Upp1 and Spo0A-Upp2 arrays show that the (hypothetically) well processed Spo0A crRNA yields low knock-out efficiency for the *Spo0A* gene, whereas the disturbed Upp1 or Upp2 crRNAs are not necessarily a limitation for knocking-out the *Upp* gene (Fig. 4A). When the Upp2 spacer preceded the Spo0A spacer, disturbed hairpins were observed for

both the Upp2 and Spo0A crRNAs (Fig. S5). In contrast, in the case where the Upp1 spacer preceded the Spo0A spacer, undisturbed hairpins were formed for both crRNAs. The Upp1-Spo0A array combination yielded the highest (18%) multiplex knock-out efficiency (Fig. 5B), which is very likely to be the result of undisturbed hairpins.



**Figure 5. Multiplex gene knock-out.** The *Spo0A* and *Upp* genes were targeted simultaneously for knock-out in a single step. Different CRISPR arrays were used with either the Spo0A spacer preceding the Upp spacer or the other way around. Two different Upp spacers were used, designated as Upp1 and Upp2. **(A)** Knock-out efficiency of either the *Spo0A* or the *Upp* gene using the different CRISPR arrays. **(B)** Double knock-out efficiency of the *Spo0A* and *Upp* genes using the different CRISPR arrays. The knock-out efficiency was determined by screening eight colonies from each replicate through colony PCR using the primers listed in Table S1. Dots represent the knock-out efficiency from each replicate.

Overall, our multiplex knock-out results cannot be fully explained by the pre-crRNA secondary structure. In most cases, the Spo0A crRNA hairpin is structured but yields very low knock-out efficiency, whereas the Upp crRNA hairpin is unstructured and often yields high knock-out efficiency. Based on our observations and the observations made by Liao *et al.* (2019) and Creutzburg *et al.* (2020), we can conclude that when a multiplex approach is considered, the test of multiple spacers targeting the same gene in combination with the change in spacer position in the CRISPR array should be applied for optimal results.

# Conclusion

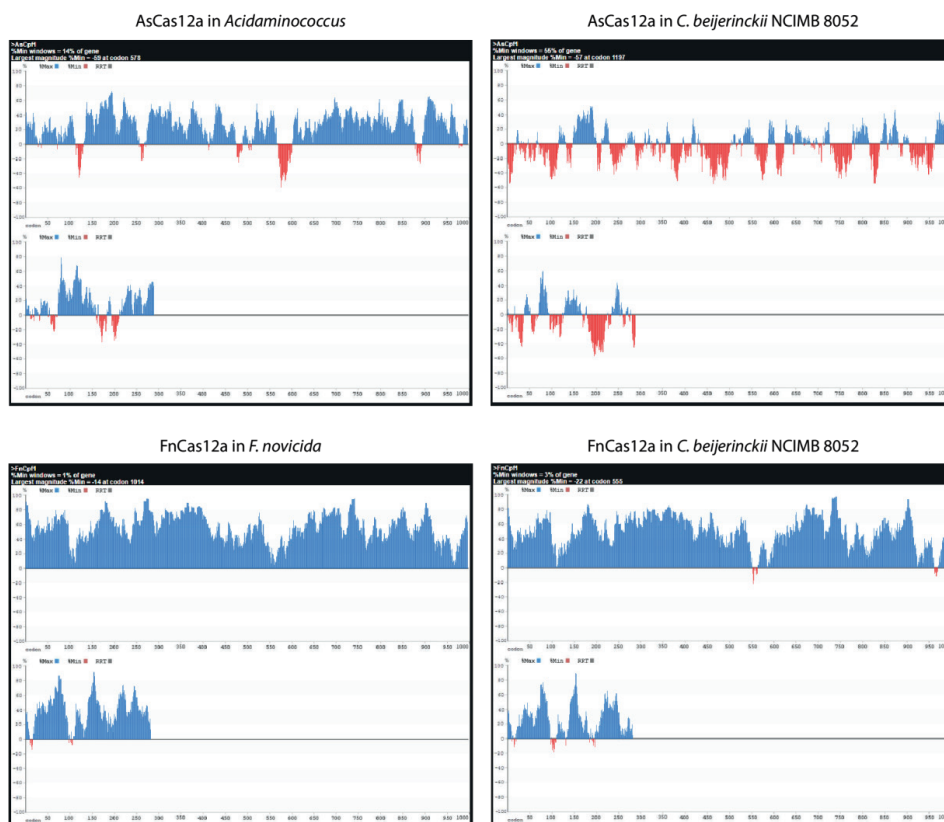
In this study, we successfully developed a CRISPR-FnCas12a genome engineering tool for *C. beijerinckii* NCIMB 8052 that can facilitate single- and multi-plex gene knock-out in a single step. The knock-out efficiency for single genes varied between 25 and 100%, indicating that different genomic loci are not targeted equally well. The knock-out efficiency for the simultaneous deletion

of two genes was 18% and dependent on the spacer sequence and position in the synthetic CRISPR array. In general, our tool expands the CRISPR-Cas toolbox in clostridia species and can contribute to the rapid and easy generation of mutants.

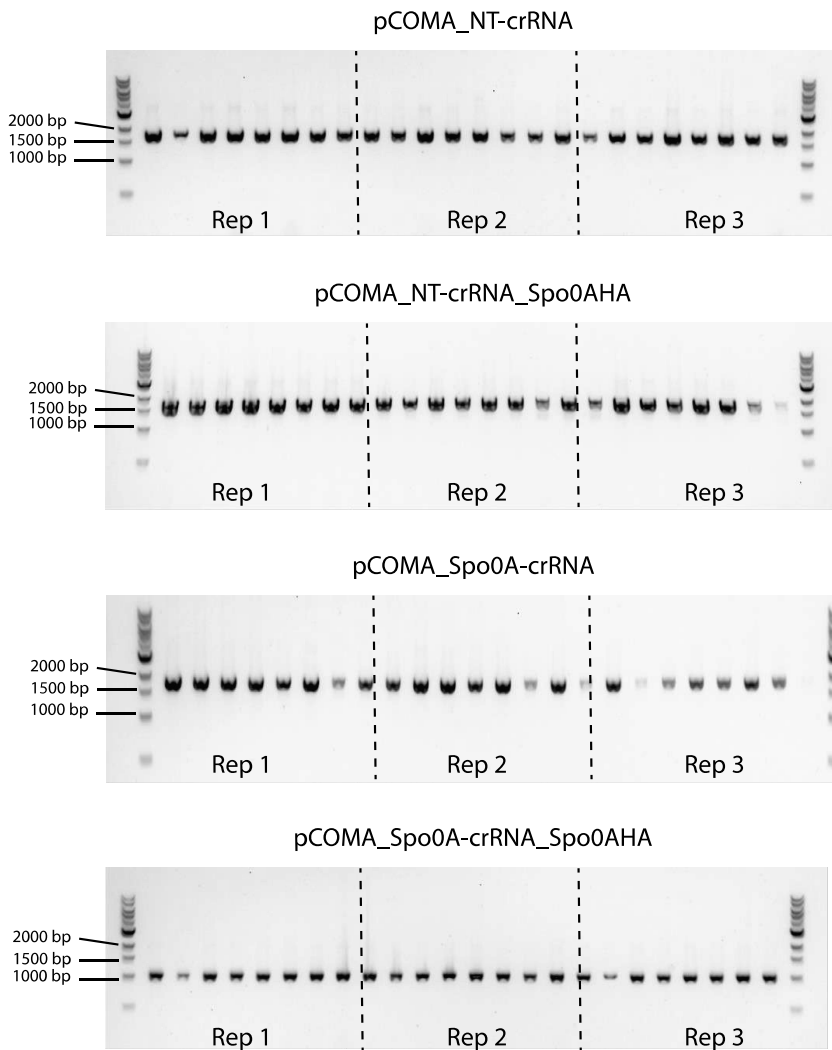
## **Acknowledgements**

We would like to thank Rob Joosten and Ton van Gelder for their technical support on this project. Also, we would like to thank Dr. Wen Wu and Dr. Prarthana Mohanraju for their advice and for sharing their knowledge on CRISPR-Cas12a.

## Supplementary information

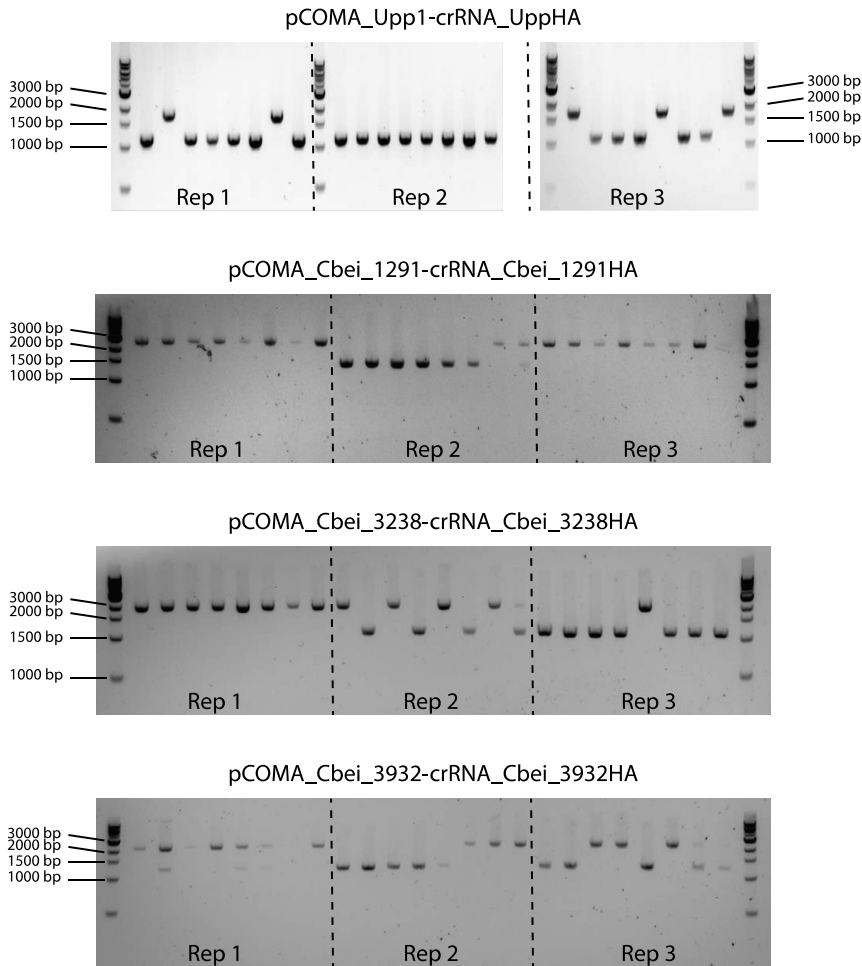


**Supplementary figure 1. Codon usage of AsCas12a and FnCas12a in the native (*Acidaminococcus* and *Francisella novicida*) and target (*C. beijerinckii* NCIMB 8052) organisms.** The codon usage for each organism was created using the codon harmonization tool developed by Claessens *et al.* (2017). The Cas12a codon usage in the native and target organisms was visualized using the codons.org website created by Clarke T.F. & Clark P.L. (2008).

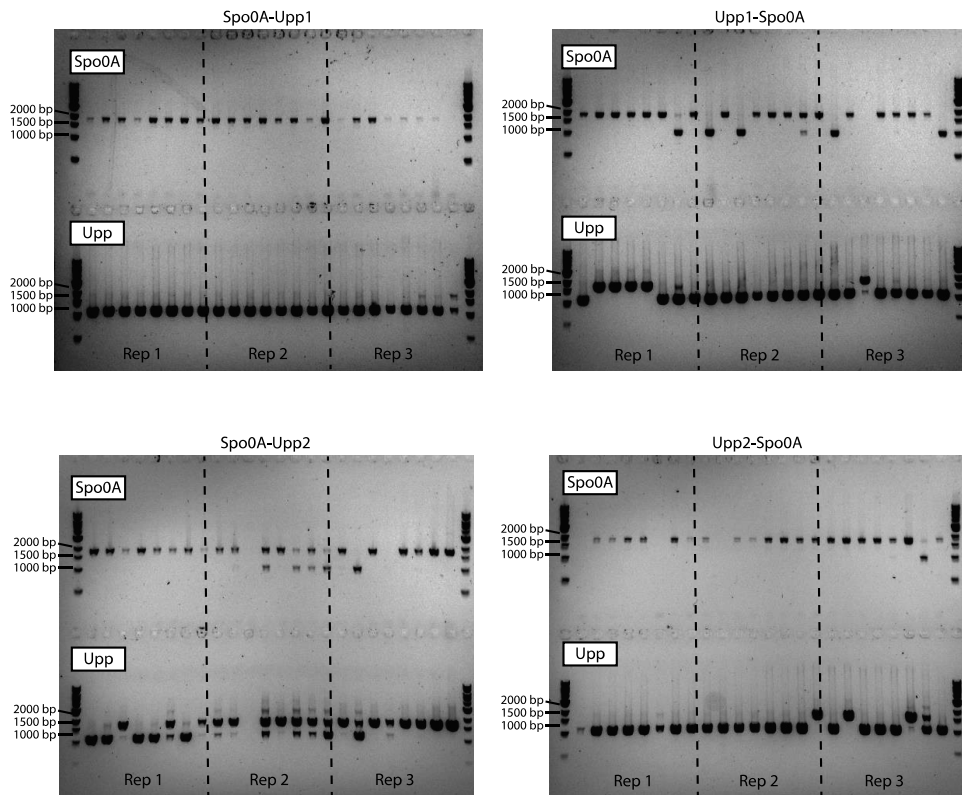


**Supplementary figure 2. Raw data for the single-gene knock-out of *Spo0A* using CRISPR-FnCas12a in *C. beijerinckii* NCIMB 8052.** *C. beijerinckii* NCIMB 8052 was transformed either with pCOMA\_NT-crRNA, pCOMA\_NT-crRNA\_Spo0AHA, pCOMA\_Spo0A-crRNA or pCOMA\_Spo0A-crRNA\_Spo0AHA and obtained colonies were screened through colony PCR using BG16483 and BG16484 oligos (Table S1). This experiment was performed in biological triplicates and the result of each triplicate is represented by Rep1, Rep2 and Rep3 at the bottom of each gel and separated by the dashed vertical lines. Mix amplicons (wild-type and knock-out bands) were not counted for the total knock-out efficiency percentage. Wild-type Spo0A: 1866 bp,  $\Delta$ Spo0A: 1044 bp.

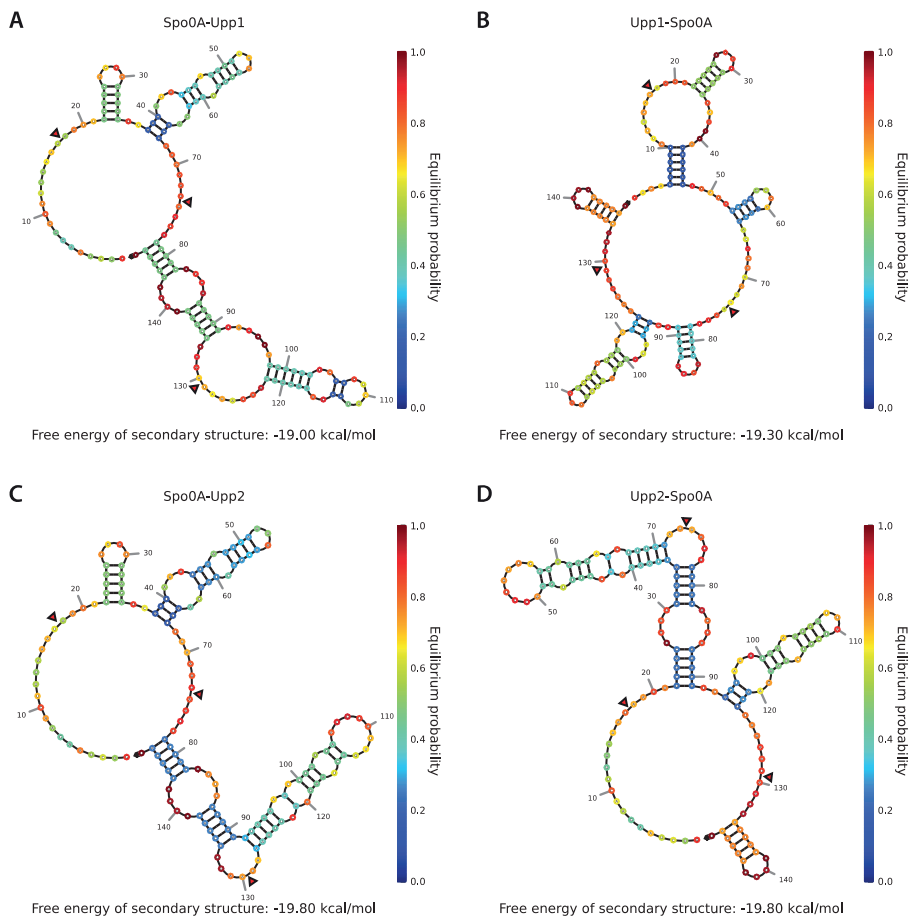




**Supplementary figure 3. Raw data for the single-gene knock-out of the *Upp*, *Cbei\_1291*, *Cbei\_3238* and *Cbei\_3932* genes using CRISPR-FnCas12a in *C. beijerinckii* NCIMB 8052.** *C. beijerinckii* NCIMB 8052 was transformed with either of pCOMA\_Upp1-crRNA\_UppHA, pCOMA\_Cbei\_1291-crRNA\_Cbei\_1291HA, pCOMA\_Cbei\_3238-crRNA\_Cbei\_3238HA or pCOMA\_Cbei\_3932-crRNA\_Cbei\_3932HA. Each knock-out experiment was performed in biological triplicates and the result of each triplicate is represented by Rep1, Rep2 and Rep3 at the bottom of each gel and separated by the dashed vertical lines. Mix amplicons (wild-type and knock-out bands) were not counted for the total knock-out efficiency percentage. The primers used for this experiment are listed in Table S1. Wild-type *Upp*: 1775 bp,  $\Delta$ *Upp*: 1145 bp. Wild-type *Cbei\_1291*: 2392 bp,  $\Delta$ *Cbei\_1291*: 1405 bp. Wild-type *Cbei\_3238*: 1979 bp,  $\Delta$ *Cbei\_3238*: 1091 bp. Wild-type *Cbei\_3932*: 2054 bp,  $\Delta$ *Cbei\_3932*: 1241 bp.



**Supplementary figure 4.** Raw data for the multiplex knock-out of *Spo0A* and *Upp* using CRISPR-FnCas12a in *C. beijerinckii* NCIMB 8052. pCOMA\_Spo0A-Upp1-crRNA\_Spo0AHA\_UppHA (top left), pCOMA\_Upp1-Spo0A-crRNA\_Spo0AHA\_UppHA (top right), pCOMA\_Spo0A-Upp2-crRNA\_Spo0AHA\_UppHA (bottom left) or pCOMA\_Upp2-Spo0A-crRNA\_Spo0AHA\_UppHA (bottom right) was used to transform *C. beijerinckii* NCIMB 8052. Each knock-out experiment was performed in biological triplicates and the result of each triplicate is represented by Rep1, Rep2 and Rep3 at the bottom of each gel and separated by the dashed vertical lines. Mix amplicons (wild-type and knock-out bands) were not counted for the total knock-out efficiency percentage. The primers used for this experiment are listed in Table S1. Wild-type *Spo0A*: 1866 bp,  $\Delta$ *Spo0A*: 1044 bp. Wild-type *Upp*: 1775 bp,  $\Delta$ *Upp*: 1145 bp.



**Supplementary figure 5. Predicted secondary structure of pre-crRNAs.** The pre-crRNAs include the three 36 nt repeats interspaced with the designated 20 nt spacers. Red triangles indicate the processing site by FnCas12a. A) pre-crRNA where the Spo0A spacer precedes the Upp1 spacer. B) pre-crRNA where the Upp1 spacer precedes the Spo0A spacer. C) pre-crRNA where the Spo0A spacer precedes the Upp2 spacer. D) pre-crRNA where the Upp2 spacer precedes the Spo0A spacer.

**Supplementary Table 1. Oligonucleotides used in this study.**

Oligo ID	Oligo sequence (5' to 3')	Description
<b>Oligonucleotides for the amplification of homologous arms</b>		
BG15796	GAGATCTCCATGGACGCGTG ACGTAATATAGTATTAATTTA TGGTGTTATATTATATAAAAG	500 bp homology arms upstream of Spo0A, forward
BG14151	ATTTTTTCTCTCCTTTTGTC	500 bp homology arms upstream of Spo0A, reverse
BG14152	GACAAAAGGAGAGAAAAAAT TTGGCTGAAGTAACATG	500 bp homology arms downstream of Spo0A, forward
BG14153	CCGGGGATCCTCTAGAGTCGT TTGAAATATTATGACCTTTTT G	500 bp homology arms downstream of Spo0A, reverse
BG16051	TCTCCATGGACGCGTGACGTA CATATTCAACAGGTTCTTG	500 bp homology arms upstream of Upp, forward
BG16052	TATTATTCCTCCAAGTTTGC	500 bp homology arms upstream of Upp, reverse
BG16053	GCAAACCTGGAGGAATAATA TTAGTAATTTATTAGAATTAA AAGCTATC	500 bp homology arms downstream of Upp, forward
BG16054	CCGGGGATCCTCTAGAGTCGT CCTGCTTTGAATCCCTTAT	500 bp homology arms downstream of Upp, reverse
BG16736	TCTCCATGGACGCGTGACGTC ATGAAACCTGCTGATAATG	500 bp homology arms upstream of Cbei_1291, forward
BG16737	GATTTCATAAATATTTACTG CCTAATAATCAATAAAATTTA TTCGGG	500 bp homology arms upstream of Cbei_1291, reverse
BG16738	AGTAAATATTTATAGAAATCG TAGGTAATGTG	500 bp homology arms downstream of Cbei_1291, forward
BG16739	CCGGGGATCCTCTAGAGTCGG TTAAATGAATTTTATGCTCA TATAAAAAAGTAATAG	500 bp homology arms downstream of Cbei_1291, reverse
BG18279	TCTCCATGGACGCGTGACGTG TGAAAATAAAATAGCTTCAA ATTATGAG	500 bp homology arms upstream of Cbei_3238, forward
BG18280	CGCATACATCTCCCTTTTG	500 bp homology arms upstream of Cbei_3238, reverse
BG18281	ACAAAAGGGAGATGTATGCG TGTTAATGAGAAAAGATACA AC	500 bp homology arms downstream of Cbei_3238, forward
BG18282	CCGGGGATCCTCTAGAGTCGA ATTATATCATACTTTCAGTAA TACT	500 bp homology arms downstream of Cbei_3238, reverse

## Chapter 5

BG16752	TCTCCATGGACGCGTGACGTA TATACTATTAAGTAAATAAA TAATACTAAAGGAAG	500 bp homology arms upstream of Cbei_3932, forward
BG16753	TATAATAACCTCCAAAAAACT ATCTATAATC	500 bp homology arms upstream of Cbei_3932, reverse
BG16754	GTTTTTGGAGGTTATTATAC CGCATGATTCATTAATTTGG	500 bp homology arms downstream of Cbei_3932, forward
BG16755	CCGGGGATCCTCTAGAGTCGG CATTAGGCCCGATTTC	500 bp homology arms downstream of Cbei_3932, reverse
<b>Oligonucleotides for knock-out confirmation</b>		
BG16483	GGAATATAAAATAAACATAG GG	$\Delta$ Spo0A, forward
BG16484	TTTCGCATACTATAATCCAC	$\Delta$ Spo0A, reverse
BG16862	AATTGGATGCGATCATGGTG	$\Delta$ Upp, forward
BG16863	ACTATTCTTAGAGAGTCATCG TCTTC	$\Delta$ Upp, reverse
BG18948	GTTTAACTTGCTTTTCCTCTC	$\Delta$ Cbei_1291, forward
BG18949	GATCCTCTTTAATTCAGGG	$\Delta$ Cbei_1291, reverse
BG18938	CGTGTATCTGCAATCAGTTTT G	$\Delta$ Cbei_3238, forward
BG18939	CTCCACCTATGTAATTTATT G	$\Delta$ Cbei_3238, reverse
BG18936	CAGTTAGGGATTGAAACATA CATG	$\Delta$ Cbei_3932, forward
BG18937	AAACCTAGCATCTAATGATT AG	$\Delta$ Cbei_3932, reverse
<b>Oligonucleotides for spacer insertion into pCOMA plasmid series; Spacers are underlined</b>		
BG15843	AGGTCTCATAGAT <u>CAGTATAT</u> <u>TAAATGATTATCGTCTAAGAG</u> ACCA	Spo0A spacer insertion through Golden Gate, forward
BG15844	TGGTCTCTTAGAC <u>GATAATCA</u> <u>TTTAATATACTGATCTATGAG</u> ACCT	Spo0A spacer insertion through Golden Gate, reverse
BG16041	AGGTCTCATAGAT <u>AACCTGTT</u> <u>TTTTCATCTCTTGCTAAGAG</u> ACCA	Upp1 spacer insertion through Golden Gate, forward
BG16042	TGGTCTCTTAGAC <u>AAGAGATG</u> <u>AAAAAACAGGTTATCTATGA</u> GACCT	Upp1 spacer insertion through Golden Gate, reverse

BG16756	AGGTCTCATAGAT <u>ATCCGGCG</u> <u>CGCATCTTTCT</u> CGTCTAAGAG ACCA	Cbei_1291 spacer insertion through Golden Gate, forward
BG16757	TGGTCTCTTAGAC <u>GAGAAA</u> <u>ATGCGCGCCGGAT</u> ATCTATGA GACCT	Cbei_1291 spacer insertion through Golden Gate, reverse
BG18319	AGGTCTCATAGAT <u>AAGAGGA</u> <u>ACGTGCAATACG</u> AGTCTAAG AGACCA	Cbei_3238 spacer insertion through Golden Gate, forward
BG18320	TGGTCTCTTAGACT <u>CGTATTG</u> <u>CACGTTCCTCTT</u> ATCTATGAG ACCT	Cbei_3238 spacer insertion through Golden Gate, reverse
BG16762	AGGTCTCATAGAT <u>TGACGCCC</u> <u>CATATATCTAAC</u> GTCTAAGAG ACCA	Cbei_3932 spacer insertion through Golden Gate, forward
BG16763	TGGTCTCTTAGAC <u>GTTAGATA</u> <u>TATGGGGCGTCA</u> ATCTATGAG ACCT	Cbei_3932 spacer insertion through Golden Gate, reverse
BG18724	AGGTCTCATAGAT <u>CAGTATAT</u> <u>TAAATGATTATCG</u> TCTAAGAA CTTTAAATAATTTCTACTGTT GTAGAT <u>AACCTGTTTTTCAT</u> <u>CTCTTGTCTAAG</u> AGACCA	Spo0A-Upp1 spacer insertion through Golden Gate, forward
BG18725	TGGTCTCTTAGACA <u>AAGAGATG</u> <u>AAAAAACAGGTT</u> ATCTACAA CAGTAGAAATTATTTAAAGTT CTTAGAC <u>GATAATCATTTAAT</u> <u>ATACTGATCTATG</u> AGACCT	Spo0A-Upp1 spacer insertion through Golden Gate, reverse
BG18726	AGGTCTCATAGAT <u>AACTGTT</u> <u>TTTTCATCTCTT</u> GTCTAAGAA CTTTAAATAATTTCTACTGTT GTAGAT <u>CAGTATATTAATGA</u> <u>TTATCGTCTAAG</u> AGACCA	Upp1-Spo0A spacer insertion through Golden Gate, forward
BG18727	TGGTCTCTTAGAC <u>GATAATCA</u> <u>TTTAATATACTGA</u> TCTACAAC AGTAGAAATTATTTAAAGTTC TTAGAC <u>AAGAGATGAAAAAA</u> <u>CAGGTTATCTATG</u> AGACCT	Upp1-Spo0A spacer insertion through Golden Gate, reverse
BG18728	AGGTCTCATAGAT <u>CAGTATAT</u> <u>TAAATGATTATCG</u> TCTAAGAA CTTTAAATAATTTCTACTGTT	Spo0A-Upp2 spacer insertion through Golden Gate, forward

## Chapter 5

	GTAGATT <u>AAAAATCCTTTGAAC</u> CTGTTGTCTAAGAGACCA	
BG18729	TGGTCTCTTAGACAACAGGTT <u>CAAAGGATTTTAATCTACAAC</u> AGTAGAAATTATTTAAAGTTC TTAGACGATAATCATTTAATA <u>TACTGATCTATGAGACCT</u>	Spo0A-Upp2 spacer insertion through Golden Gate, reverse
BG18730	AGGTCTCATAGATT <u>AAAAATCC</u> <u>TTGAACCTGTTGTCTAAGAA</u> CTTTAAATAATTTCTACTGTT GTAGATCAGTATATTAAATGA <u>TTATCGTCTAAGAGACCA</u>	Upp2-Spo0A spacer insertion through Golden Gate, forward
BG18731	TGGTCTCTTAGACGATAATCA <u>TTTAATATACTGATCTACAAC</u> AGTAGAAATTATTTAAAGTTC TTAGAC <u>ACAGGTTCAAAGG</u> <u>ATTTTAATCTATGAGACCT</u>	Upp2-Spo0A spacer insertion through Golden Gate, reverse

## Spacer insertion protocol

The protocol below has been adopted from Badianis *et al.* (2020) and it describes the easy and efficient insertion of CRISPR spacers in the Cas12a CRISPR array.

### Example of how to order spacers.

Red is an example of a 20 nt spacer. Black is the overhang to insert the spacer to the backbone containing BpiI or BbsI Type II S sites. Note: overhangs may vary in sequence according to which backbone is used.

5'-AGAAGACATAGATGACTGCTGGGGTCAGCTCCCGTCTATGTCTTCA-3'

5'-TGAAGACATAGACGGGAGCTGACCCCAGCAGTCATCTATGTCTTCT-3'

### Create the primer dimers for Golden Gate.

- Incubate at 95°C for 5 mins. Cool at room temperature for 2 h.

Oligo annealing	
Solution	Volume (μL)
Oligo 1 (stock solution: 100 μM)	1
Oligo 2 (stock solution: 100 μM)	1
NaCl (1 M)	1
MQ	47

- Dilute the oligo dimers 10 times (i.e. 1 μL in 9 μL MQ). Final concentration is 0.2 pM μL<sup>-1</sup>.

Prepare the MetaMix stock (or if you have stock use it directly) for Golden Gate. Store MetaMix at -20°C until use.

MetaMix	
Solution	Volume (μL)
*Type II S enzyme (NEB)	10
T4 ligation buffer (NEB)	15
T4 ligase (NEB)	10
MQ	15

\*Make sure your enzyme can work in the T4 ligation buffer or whether it need additional elements.



## Chapter 5

Mix the below in one PCR tube.

Golden Gate	
Solution	Volume ( $\mu\text{L}$ )
Diluted oligo dimers	2
Plasmid (30 ng $\mu\text{L}^{-1}$ stock)	2
MetaMix	2

Incubate following the below PCR program for Golden Gate.

Temperature ( $^{\circ}\text{C}$ )	Time (secs-mins)	
37	300-5	
16	300-5	
37	300-5	x 15*
37	300-5	
80	1200-20	
12	$\infty$	

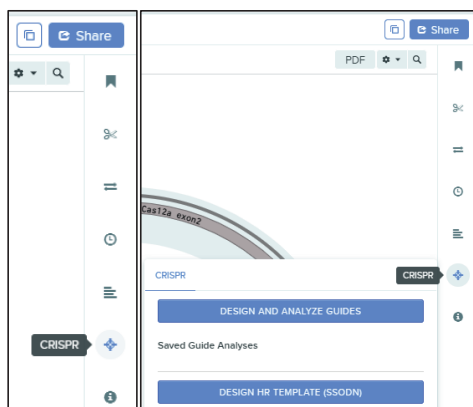
\*Increase to 30 cycles for increased efficiency

- Use 1  $\mu\text{L}$  of the reaction to transform heat competent *E. coli* DH5a (NEB)

## Spacer selection protocol

The selection of Cas12a CRISPR spacers is based on the knowledge obtained from Zetsche *et al.* (2015), Hui *et al.* (2017), Creutzburg *et al.* (2020) and Liao *et al.* (2019) (416,427,428,430). The CRISPR Guide RNA design tool developed by Benchling is used to find spacers in the target genomic locus.

- The target genomic locus is selected and the CRISPR tool in Benchling is used.



- The settings single guide, 20 bp for the guide length, custom PAM (TTV) and 5' to reassurance that the PAM is 5' to 3' are selected. If present, select the genome of the organism of interest. If the organism of interest is not present in the drop down list, then a closely related species can be used. In the case where your organism of interest or a closely related species are not present in the list, choose a random genome. The outcome in finding spacer sequences will not change but the off-target score will not be reliable (we tackle this issue below).
- Note: the PAM sequence for Cas12a can be 5'-TTTV-3' but in organisms with a high in GC content genome, 5'-TTTV-3' PAM sequences may be rare. After completing the selection, click the finish button.

Design CRISPR Guides: Guide parameters

Design Type

☒ Single guide

Wild-type Cpf1, single gRNA

☐ Paired guides

Cpf1 nickase is not available

☐ Guides for "base editing" (Komor et al., 2016)

C → T (or G → A) substitution, no dsDNA breaks

Guide Length

20

Genome

ASM584v2 (Escherichia coli str. K-12 substr. MG1655)

Don't see the genome you're looking for? We may be able to import it, just ask.

PAM

Custom PAM

TTV

☐ 3' ☒ 5'

On-target scores are only defined for Cas9 PAMs. Resulting analysis will omit the on-target score.

Advanced Settings

☐ Save these as my default CRISPR settings

Finish

- If your target region is already selected, proceed by clicking the create button. Otherwise, you need to select your target region by selecting it on the sequence map and then press the create button.

To get started, create a target region by selecting it on the sequence map and pressing

Target Region

5532

2206

Create

- A new tab will appear showing a list with potential spacers as indicated below. Sort the spacers by Off-Target Score (the higher the better). Sorting should be done only in the case where the genome of interest was present in the list of genomes.

LINEAR MAP DESCRIPTION METADATA PLASMID DESIGN CRISPR X

Untitled Save Settings

Target Region 7176 7195

☒ Start End Annotations Genome Region

☒ 5532 2206 Cas12a exon2 No region set

You don't have a genome region set above, so scores may not match other sites such as [crispr.mit.edu](https://crispr.mit.edu). Benchmarking uses the genome region to locate your target region and to ignore potential off-target sites in that part of the genome.

Export Save Assemble 0 selected 1-100 of 575 Prev Next

<input type="checkbox"/>	Position	Strand	Guide Sequence	PAM	On-Target Score	Off-Target Score
<input type="checkbox"/>	7213	+	GACAAGGATGAGCATTTT	TTA	100.0	100.0
<input type="checkbox"/>	7184	+	GTCAGTCAGAAGATAAGGCA	TTA	100.0	100.0
<input type="checkbox"/>	7178	+	ATATTAGTCAGTCAGAAGAT	TTC	100.0	100.0
<input type="checkbox"/>	7177	-	TCTAAAATATTTGCCTTATC	TTG	100.0	100.0
<input type="checkbox"/>	7164	-	CCTTATCTTCTGACTGACTA	TTG	100.0	100.0
<input type="checkbox"/>	7159	-	TCTTCTGACTGACTAATATG	TTA	100.0	100.0
<input type="checkbox"/>	7154	-	TGACTGACTAATATGAAATA	TTC	100.0	100.0
<input type="checkbox"/>	7128	-	GTTTATGTAAGAGATTATTA	TTA	100.0	100.0
<input type="checkbox"/>	7123	-	TGTAAGAGATTATTAGTTTG	TTA	100.0	100.0
<input type="checkbox"/>	7111	-	TTAGTTTGATCTAAAAGATC	TTA	100.0	100.0
<input type="checkbox"/>	7108	-	GTTTGATCTAAAAGATCCTT	TTA	100.0	100.0
<input type="checkbox"/>	7103	-	ATCTAAAAGATCCTTGATAG	TTG	100.0	100.0
<input type="checkbox"/>	7078	-	ACATCATCTTCGCACTAGC	TTA	100.0	100.0

ASSEMBLY WIZARD SPLIT WORKSPACE

- Continue by exporting the list of spacers. You can either export all spacers or select the ones you prefer and export the selection that you made.

Export Save Assemble 100 selected 1-100 of 575 Prev Next

Export Selected (.tsv)

Export All (.tsv)

<input type="checkbox"/>	Position	Strand	Guide Sequence	PAM	On-Target Score	Off-Target Score
<input checked="" type="checkbox"/>	7184	+	GTCAGTCAGAAGATAAGGCA	TTA	100.0	100.0
<input checked="" type="checkbox"/>	7178	+	ATATTAGTCAGTCAGAAGAT	TTC	100.0	100.0

- Paste the list in excel.

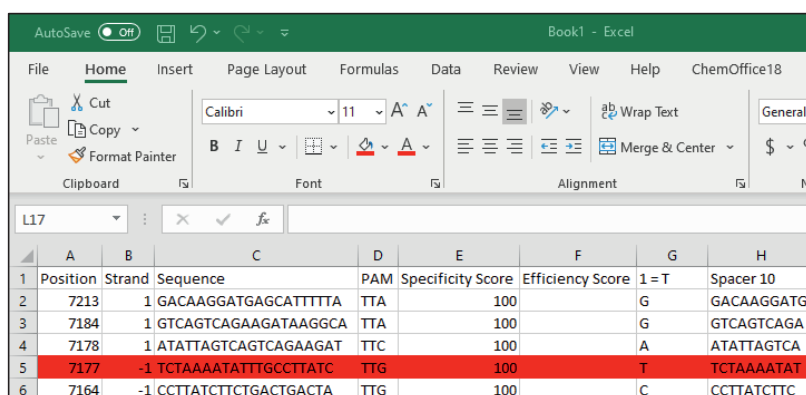
Chapter 5

	A	B	C	D	E	F
1	Position	Strand	Sequence	PAM	Specificity Score	Efficiency Score
2	7213	1	GACAAGGATGAGCATTTTAA	TTA	100	
3	7184	1	GTCAGTCAGAAGATAAGGCA	TTA	100	
4	7178	1	ATATTAGTCAGTCAGAAGAT	TTC	100	
5	7177	-1	TCTAAAATATTGCTTATC	TTG	100	
6	7164	-1	CCTTATCTCTGACTGACTA	TTG	100	

- Create a new column and name it “1 = T”. Then use the function “=LEFT(text,[num\_chars])” (e.g. “=LEFT(C2, 1)”). This function extracts the first character (nt) from the selected cell. We do this to spot and remove the spacers that start with a T, since such spacers are very inefficient according to Hui *et al.* (2017). Label the spacers with red color to avoid from choosing them in further steps (or delete them from your list).

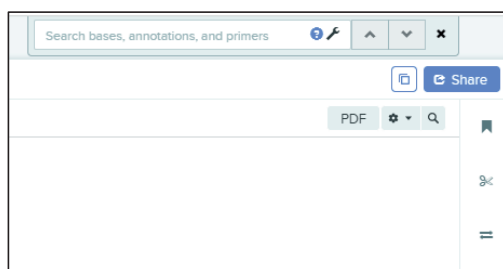
	A	B	C	D	E	F	G
1	Position	Strand	Sequence	PAM	Specificity Score	Efficiency Score	1 = T
2	7213	1	GACAAGGATGAGCATTTTAA	TTA	100		G
3	7184	1	GTCAGTCAGAAGATAAGGCA	TTA	100		G
4	7178	1	ATATTAGTCAGTCAGAAGAT	TTC	100		A
5	7177	-1	TCTAAAATATTGCTTATC	TTG	100		T
6	7164	-1	CCTTATCTCTGACTGACTA	TTG	100		C

- Optional step: Create a new column and name it Spacer 10. Then use the function “=LEFT(C2, 10)”. This function extracts the first 10 characters (nts) from the selected cell. We will use the 10 first nts of the spacers to manually check for off targets against the genome of the organism of interest. We use the first 10 bases as this region contains the seed region (5 first nts) of the spacer and additional 5 nucleotides that we consider important for off-targeting. Note: This step is optional since Benchling may not have the genome of the organism of interest and therefore off-target scores may be inaccurate.



	A	B	C	D	E	F	G	H
	Position	Strand	Sequence	PAM	Specificity Score	Efficiency Score	1 = T	Spacer 10
2	7213	1	GACAAGGATGAGCATTTT	TTA	100		G	GACAAGGATG
3	7184	1	GTCAGTCAGAAGATAAGGCA	TTA	100		G	GTCAGTCAGA
4	7178	1	ATATTAGTCAGTCAGAAGAT	TTC	100		A	ATATTAGTCA
5	7177	-1	TCTAAAATATTTGCCCTTATC	TTG	100		T	TCTAAAATAT
6	7164	-1	CCTTATCTTCTGACTGACTA	TTG	100		C	CCTTATCTTC

- **Optional step:** To manually check for off-targets, download the complete genome of your organism of interest and use it to create a linear DNA sequence in benchling. Then, press ctrl+F to open the “Search bases, annotations and primers” tab. Then type TTV followed by the 10 nts retrieved from the Spacer 10 column (see above). There should be at least one exact match for the searched sequence which is the target sequence retrieved from the Benchling CRISPR tool. In the case where more than one matches are found, off-targeting could be possible and such spacers should be avoided.



- After selecting the appropriate spacer, the secondary RNA structure of the CRISPR array should be checked to avoid secondary RNA structures that inhibit the formation of the characteristic hairpin required for Cas12a processing (427,430). To do this, new columns are created in Excel containing the CRISPR repeat sequence and the spacer sequence as indicated below. In the example below, a repeat-spacer-repeat array is used as the CRISPR-array but a repeat-spacer can also be used.

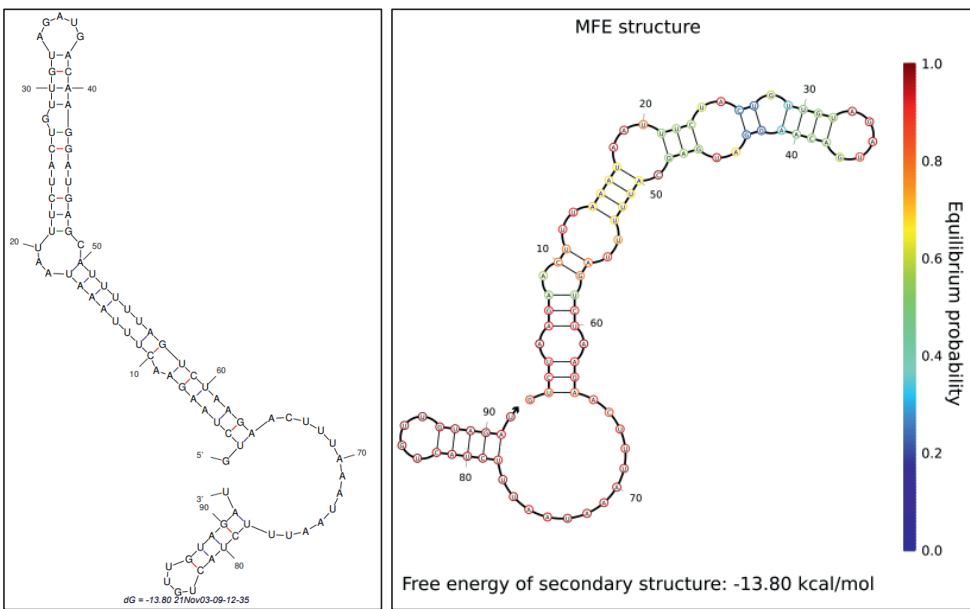
Chapter 5

Position	Strand	Sequence	PAM	Specificity Score	Efficiency Score	1 = T	Spacer 10	Repeat	Spacer	Repeat
7213	1	GACAAGGATGAGCATTTT	TTA	100	G	GACAAGGATG	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GACAAGGATGAGCATTTT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
7184	1	GTCAGTCAGAGAAGTAAGCA	TTA	100	G	GTCAGTCAG	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCAGTCAGAGAAGTAAGCA	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
7178	1	ATATTAGTCAGTCAGAAGAT	TTC	100	A	ATATTAGTC	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	ATATTAGTCAGTCAGAAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
7177	-1	TCTAAAATATTGCGCTTATC	TTG	100	T	TCTAAAAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	TCTAAAATATTGCGCTTATC	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
7164	-1	CCTATCTCTGACTGACTA	TTG	100	C	CCTATCTCT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	CCTATCTCTGACTGACTA	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT

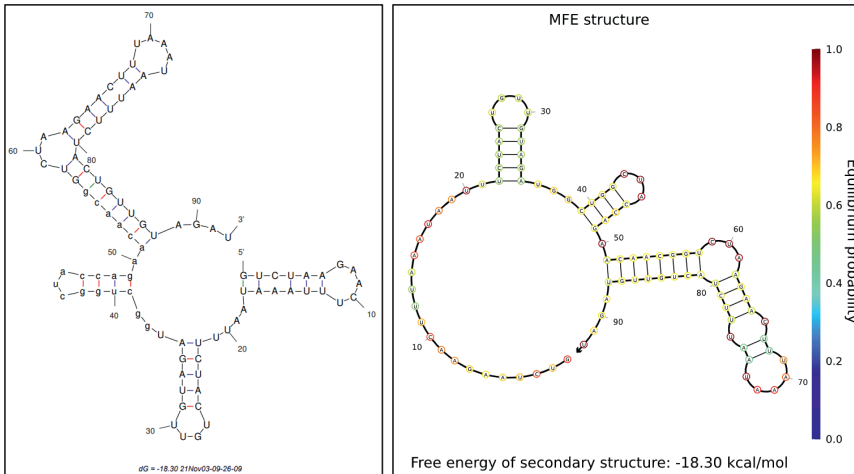
- Copy the repeat-spacer-repeat sequence and use it to predict the RNA secondary structure in online tools such as mFold (UNAFold Web Server) or NUPACK.

Position	Strand	Sequence	PAM	Specificity Score	Efficiency Score	1 = T	Spacer 10	Repeat	Spacer	Repeat
7213	1	GACAAGGATGAGCATTTT	TTA	100	G	GACAAGGATG	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GACAAGGATGAGCATTTT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
7184	1	GTCAGTCAGAGAAGTAAGCA	TTA	100	G	GTCAGTCAG	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCAGTCAGAGAAGTAAGCA	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
7178	1	ATATTAGTCAGTCAGAAGAT	TTC	100	A	ATATTAGTC	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	ATATTAGTCAGTCAGAAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
7177	-1	TCTAAAATATTGCGCTTATC	TTG	100	T	TCTAAAAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	TCTAAAATATTGCGCTTATC	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT
7164	-1	CCTATCTCTGACTGACTA	TTG	100	C	CCTATCTCT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	CCTATCTCTGACTGACTA	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT	GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT

- The secondary structure outcome of the selected repeat-spacer-repeat RNA is shown below (Left: mFold, Right: NUPACK). We consider the secondary structure as “not optimal” since the hairpin (nts 22 to 35) required for Cas12a processing is inhibited. However, the secondary structure of the RNA is an *in silico* prediction and may not represent the actual secondary structure in an *in vivo* situation.

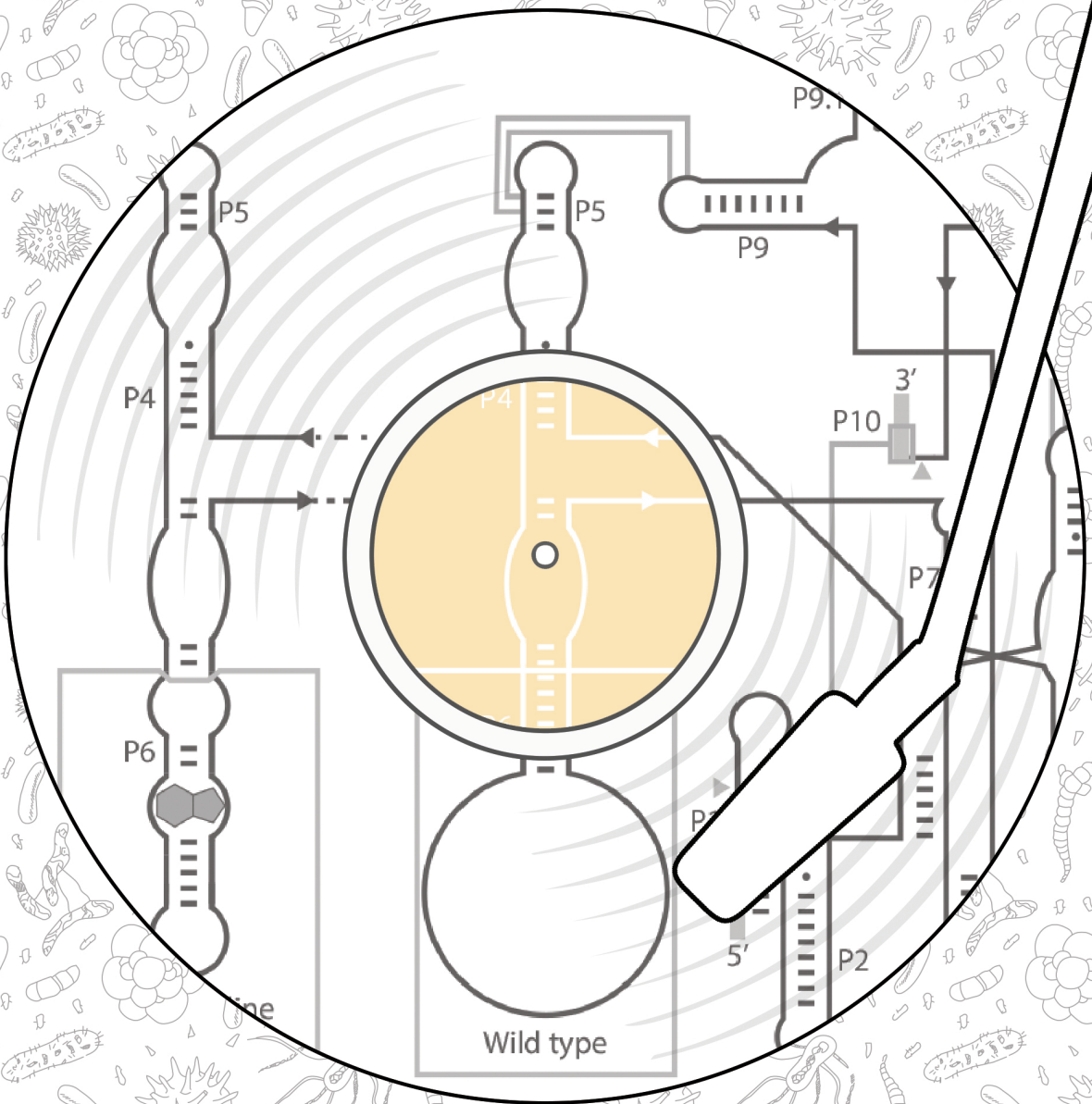


- Below we show an example of what we consider a “good” spacer. From the two predictions (Left: mFold, Right: NUPACK), we can see that the hairpin (nts 22 to 35) required for Cas12a processing is not inhibited.





# CHAPTER 6



## Chapter 6

# Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas

Constantinos Patinios<sup>1a</sup>, Sjoerd C.A. Creutzburg<sup>1a</sup>, Adini Q. Arifah<sup>1</sup>, Belén Adiego-Pérez<sup>1</sup>, Evans A. Gyimah<sup>1</sup>, Colin J. Ingham<sup>2</sup>, Servé W.M. Kengen<sup>1</sup>, John van der Oost<sup>1</sup>, Raymond H.J. Staals<sup>1\*</sup>

<sup>1</sup>Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands.

<sup>2</sup>Hoekmine Besloten Vennootschap, Kenniscentrum Technologie en Innovatie, Hogeschool Utrecht, 3584 CS, Utrecht, The Netherlands.

<sup>a</sup>These authors have contributed equally to the work

\*Corresponding author: Dr. Raymond H.J. Staals (raymond.staals@wur.nl)

Chapter adapted from publication:

Nucleic Acid Research; <https://doi.org/10.1093/nar/gkab893>

## Abstract

CRISPR-Cas is a powerful tool for genome editing in bacteria. However, its efficacy is dependent on host factors (such as DNA repair pathways) and/or exogenous expression of recombinases. In this study, we mitigated these constraints by developing a simple and widely applicable genome engineering tool for bacteria which we termed SIBR-Cas (Self-splicing Intron-Based Riboswitch-Cas). SIBR-Cas was generated from a mutant library of the theophylline-dependent self-splicing T4 *td* intron that allows for tight and inducible control over CRISPR-Cas counter-selection. This control delays CRISPR-Cas counter-selection, granting more time for the editing event (e.g. by homologous recombination) to occur. Without the use of exogenous recombinases, SIBR-Cas was successfully applied to knock-out several genes in three wild-type bacteria species (*Escherichia coli* MG1655, *Pseudomonas putida* KT2440 and *Flavobacterium* IR1) with poor homologous recombination systems. Compared to other genome engineering tools, SIBR-Cas is simple, tightly regulated and widely applicable for most (non-model) bacteria. Furthermore, we propose that SIBR can have a wider application as a simple gene expression and gene regulation control mechanism for any gene or RNA of interest in bacteria.

## Introduction

Homologous recombination (HR) combined with CRISPR-Cas counter-selection is a powerful approach for genome editing in a wide range of bacterial species (48,61). Double stranded DNA (dsDNA) breaks generated by CRISPR-Cas can initiate recombinational repair through RecBCD (432), but its efficiency might be outcompeted by the strong CRISPR-Cas counter-selective pressure (433). We therefore anticipate that gaining control over CRISPR-Cas counter-selection, would be beneficial for genome editing purposes. In other words, to achieve high editing efficiencies, generation of the desired edit through HR should precede CRISPR-Cas counter-selection. Typically, the efficiency of HR in prokaryotes is low, making genome editing through CRISPR-Cas counter-selection cumbersome.

To enhance HR frequencies, the heterologous expression of recombinases has been used. However, this method is often laborious (maintenance of multiple plasmids) and success is not guaranteed as the recombinases may be incompatible with the target organism (434). As an alternative, several regulation systems have been developed to control the expression and activity of the CRISPR-Cas module. Examples include the use of inducible promoters (65-67), inducible intein splicing (68,69), split Cas proteins (70,71), inducible conformation change (72), inducible inhibition through aptamers (73) and inducible translation through riboswitches (74). Other approaches focused on the inducible guide RNA functionality using ribozymes (75), riboswitches (76,77) and photocaging (78-81). The control of the CRISPR-Cas modules gives enough time for HR to occur before CRISPR-Cas counter-selection is induced. Whilst the existing approaches are tailored for the organism, Cas protein or guide RNA (gRNA) of interest, these solutions are typically not widely applicable. Therefore, CRISPR-Cas engineering tools would benefit from widely applicable and tight regulation of their counter-selective properties to provide enough time for HR to take place.

Ribozymes and riboswitches are gene regulation systems found in a wide range of bacterial species. The catalytic and/or regulatory functionality of these RNA molecules relies on their primary, secondary and tertiary structures, making them great candidates for developing universal tools for regulating gene expression, without the use of proteins (435-439). To this end, several studies used ribozymes and riboswitches to control the expression of a gene of interest (GOI), but also for regulating the activity and function of CRISPR-Cas (73-77,436). Nevertheless, these approaches leave room for improvement. For example, the technology developed by Tang *et al.* (2017) (75), requires base pairing of the CRISPR spacer sequence with the 5' end of the hammerhead ribozyme; something that requires modification in case the CRISPR spacer needs to be changed. Moreover, the studies by Kundert *et al.* (2019) (77), Siu *et al.* (2019) (76) and Zhao *et al.* (2020) (73) rely on

the secondary structure of the Cas9 single guide RNA (sgRNA), which rules out the use of other CRISPR-Cas systems. Lastly, the RiboCas technology developed by Cañadas *et al.* (2019) (74), regulates the expression of Cas9 by masking the RBS with a theophylline-dependent riboswitch. Whereas this is a smart alternative to previous approaches, it can be cumbersome to use in organisms that either do not use the canonical RBS sequence, or in cases that the secondary structure of the 5'-UTR sequence interferes with the theophylline aptamer (440-442).

A unique type of ribozymes includes the self-splicing Group I introns. Group I introns have been described to control gene expression and RNA processing in bacteria and phages but also in some eukaryotes (protozoa and plants) (443-445). Due to their prevalence and simplistic nature, Group I introns have the potential to be used as universal, synthetic ribozymes to control gene expression. Especially when ribozymes are associated with a specific ligand-binding sequence (RNA aptamer), the presence/absence of such a ligand acts as an ON/OFF switch for splicing (riboswitch), thereby controlling the expression of an associated gene. An example of a natural Group I intron-based riboswitch has been discovered in the bacterium *Clostridium difficile*, where its sequence resides between the RBS and the ATG start codon of an adjacent gene (446,447). After transcription, this results in a secondary structure in the 5'-UTR that prevents recruitment of the ribosome, hence hampering translation initiation. After induction by intracellular GTP or c-di-GMP, this ribozyme induces its splicing from the precursor transcript, resulting in appropriate repositioning of the RBS upstream the start codon, thereby allowing for the ribosome to start the translation process (446,447). Although this natural mechanism is a beautiful case of gene expression control, its requirement for specific endogenous inducers (GTP and c-di-GMP) as well as its dependency on specific secondary structures (including both the ribozyme and the coding sequence) complicates its general applicability. A synthetic alternative was provided by Thompson *et al.* (2002), when they combined the self-splicing Group I intron of the T4 bacteriophage with a theophylline aptamer towards a functional inducible gene expression system (448). Although this system was restricted to controlling the original *thymidylate synthase* (*td*) gene, we here describe its repurposing as a generic system to tune gene expression.

In this study, we developed the Self-splicing Intron-Based Riboswitch (SIBR) system. SIBR is based on the bacteriophage T4 *td* Group I self-splicing intron and has been engineered and repurposed as a modular, tightly regulated system that can control the expression of any GOI in a wide range of bacterial species. To illustrate this, we used SIBR to control the Cas12a nuclease from *Francisella novicida* (named SIBR-Cas) and demonstrated efficient genome editing in three wild-type (WT) bacterial species (*Escherichia coli* MG1655, *Pseudomonas putida* KT2440 and *Flavobacterium IR1*) without the use of exogenous recombinases nor the use of inducible

promoters. SIBR-Cas is an elegant solution for the widespread problem of engineering prokaryotic organisms with poor recombination efficiencies. We also suggest that SIBR can be used as a universal OFF/ON or ON/OFF switch for individual genes or multiple genes in a polycistronic operon.

## Materials and methods

### Bacterial strains, handling and growth conditions

*E. coli* DH5a (NEB) was used for general plasmid propagation and standard molecular techniques. *E. coli* DH10B T1<sup>R</sup> (Invitrogen) was used for the LacZ assays. *E. coli* MG1655 (ATCC) was used for targeting and knock-out assays. Unless specified otherwise, *E. coli* strains were grown at 37°C in LB liquid medium (10 g L<sup>-1</sup> tryptone, 5 g L<sup>-1</sup> yeast extract, 10 g L<sup>-1</sup> NaCl) or on LB agar plates (LB liquid medium, 15 g L<sup>-1</sup> bacteriological agarose) containing the appropriate antibiotics: spectinomycin (100 mg L<sup>-1</sup>), kanamycin (50 mg L<sup>-1</sup>), ampicillin (100 mg L<sup>-1</sup>) or chloramphenicol (35 mg L<sup>-1</sup>). Transformation of electro-competent *E. coli* cells was performed in 2 mm electroporation cuvettes with an ECM 63 electroporator (BTX) at 2500 V, 200 Ω and 25 μF.

*P. putida* strain KT2440 was obtained from DSMZ. Cells were grown at 30°C in LB liquid medium or on LB agar plates containing kanamycin (50 mg L<sup>-1</sup>). Electro-competent *P. putida* cells were transformed in 2 mm electroporation cuvettes using 2500 V, 200 Ω and 25 μF.

*Flavobacterium* species Iridescence 1 (sp. IR1) was kindly provided by Hoekmine BV. WT *Flavobacterium* IR1 was grown at 25°C in ASW medium (5 g L<sup>-1</sup> peptone, 1 g L<sup>-1</sup> yeast extract, 10 g L<sup>-1</sup> sea salt) or plated on ASW agar (ASW medium, 15 g L<sup>-1</sup> agar) containing erythromycin (200 mg L<sup>-1</sup>) where appropriate. Electro-competent *Flavobacterium* IR1 cells were transformed in 1 mm electroporation cuvettes using 1500 V, 200 Ω and 25 μF.

### Electro-competent cell preparation

Pre-cultures of *E. coli* or *P. putida* were grown overnight at 37°C in fresh 10 mL LB broth. 5 mL of the overnight culture was inoculated in 500 mL of pre-warmed 2×YP medium (16 g L<sup>-1</sup> peptone, 10 g L<sup>-1</sup> yeast extract) and incubated at 37°C shaking at 200 rpm until an OD<sub>600</sub> of 0.4 was reached. The culture was then cooled down to 4°C. Next, the culture was aliquoted into two sterile 450 mL centrifuge tubes and centrifuged at 3000 g for 10 minutes at 4°C. The supernatant was decanted and the pellet was washed with 250 mL ice-cold sterile miliQ water followed by centrifugation at 3000 g for 10 minutes. The supernatant was decanted and the pellet was resuspended using 5 mL of ice cold 10% glycerol. The two resuspensions were combined in one tube and ice-cold 10% glycerol was added to reach a final volume of 250 mL, followed by centrifugation at 3000 g for 10 minutes. The supernatant was decanted and the pellet was washed with 250 mL of ice-cold 10% glycerol and centrifuged at 3000 g for 10 minutes. The supernatant was decanted and the pellet was resuspended with 2 mL of ice-cold 10% glycerol and aliquoted into tubes of 40 μL. All the electrocompetent cells were stored at -80°C prior to transformation. To prepare electro-competent

cells, *Flavobacterium* IR1 was grown overnight in 10 mL ASW at 25°C, shaking at 200 rpm. The overnight culture was used to inoculate 500 mL ASW in 2 L baffled flask to a starting OD<sub>600</sub> of 0.05 and incubated at 25°C and shaking at 200 rpm until the cell density reached an OD<sub>600</sub> equal to 0.3-0.4. The cells were cooled down at 4°C and kept cold on ice for the rest of the procedure. The culture was divided into two sterile 450 mL centrifuge tubes and centrifuged at 3000 g for 10 minutes at 4°C. The supernatant was decanted, and the cell pellet was washed twice with 250 mL ice cold washing buffer (10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>) followed by centrifugation at 3000 g for 10 min at 4°C. The supernatant was removed, and the pellet was resuspended using 5 mL of the washing buffer. All the resuspensions were combined in one tube and washed by adding 250 mL of the washing buffer. It was then washed once with 10% glycerol followed by centrifugation at 3000 g for 10 minutes. The supernatant was decanted, and the resulting cell pellet was resuspended with 5 mL of ice cold 10% (v/v) glycerol and 100 µL aliquots were stored at -80°C until use.

### Plasmid construction

The LacZ reporter plasmid series were constructed from pEA001 [PWW]. The LacZ reporter plasmid series contain the *E. coli* LacZ gene under the control of the constitutive lacUV5 promoter. Ten amino acids flanking the T4 *td* intron (five from each side) were introduced between D6 and S7 of LacZ, omitting the intron itself. For cloning purposes, the ten amino acids were in turn flanked by a BspTI and PstI restriction sites. Generating the complete mutant series was performed by PCR, digestion with BspTI and PstI (Thermo Fisher Scientific) and ligation into pEA001 [PWW].

The main components (origin of replication, antibiotic resistance gene and promoters) of the SIBR-Cas plasmids for *E. coli* and *P. putida* were designed to be functional in both organisms. The constitutive lacUV5 promoter was used to drive the expression of the *FnCas12a* variants (WT and Int1-4) and an additional lacUV5 promoter was used to drive the expression of the crRNA. The empty vectors pSIBR001-005 were designed to allow convenient insertion of new spacers through Golden Gate Assembly using the BbsI-HF® enzyme and the T4 DNA ligase (NEB), following the protocol as previously described by Batianis *et al.* (2019) (377). Homology arms (500 bp) were introduced to the SIBR-Cas plasmids at a multiple cloning site (MCS). Briefly, homology arms were amplified from genomic DNA and introduced to the MCS of the linearized SIBR-Cas plasmid using the NEBuilder® HiFi DNA Assembly Master Mix (NEB). The plasmid was linearized using Esp3I (NEB). The DNA sequence of all plasmids was verified through Sanger sequencing (Macrogen Europe B.V.). To construct the SIBR-Cas plasmids for *Flavobacterium* IR1, the backbone of pSpyCas9Fb\_NT (449) was used but the Cas9 and the sgRNA were replaced with the *FnCas12a* variants (WT and Int1-4) and the crRNA, respectively. Spacers and homology arms were



introduced through Golden Gate using BsaI-HF® enzyme and NEBuilder® HiFi DNA Assembly, respectively, as described above. Other plasmids and oligonucleotides used in this study are listed in Supplementary Table 1 and 2 respectively. The complete plasmid maps including the intron sequences, the Cas12a coding sequences, homology arms and the crRNAs sequences used for each plasmid can be accessed through the Benchling links in Supplementary Table 1. Also, the spacer moieties of the crRNA used in this study are highlighted in Supplementary Table 2.

### Chemicals and reagents

Unless otherwise specified, all chemical reagents were purchased from Sigma-Aldrich. Sea salt was purchased from Sel Marine. A 40 mM theophylline (Sigma-Aldrich) stock was prepared by dissolving theophylline in dH<sub>2</sub>O followed by filter sterilization using a 0.2 µm Whatman® puradisc syringe filter. When necessary, 0-10 mM theophylline was added to the liquid or solid medium. 20 mg mL<sup>-1</sup> X-Gal (Sigma-Aldrich) stock was prepared by dissolving X-Gal in N,N-Dimethylmethanamide. The final X-Gal concentration for blue/white colony screening was 0.2 mg mL<sup>-1</sup>.

### β-galactosidase activity assay

LacZ activity was assayed in *E. coli* DH10B T1<sup>R</sup> in triplicate. Transformed *E. coli* cells carrying a single variant of the pEA001 plasmid series (Table S1) were grown overnight at 37°C, after which 20 µL of culture was mixed with 80 µL of permeabilization solution (100 mM Na<sub>2</sub>HPO<sub>4</sub>, 20 mM KCl, 2 mM MgSO<sub>4</sub>, 0.8 g L<sup>-1</sup> CTAB, 0.4 g L<sup>-1</sup> sodium deoxycholate and 5.4 mL L<sup>-1</sup> β-mercaptoethanol) and incubated at 30°C for 30 min. 600 µL of pre-warmed substrate solution (60 mM Na<sub>2</sub>HPO<sub>4</sub>, 40 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 g L<sup>-1</sup> o-nitrophenyl-β-D-galactopyranoside and 2.7 mL L<sup>-1</sup> β-mercaptoethanol) was added and incubated at 30°C until sufficient colour had developed. 700 µL of stop solution (1 M Na<sub>2</sub>CO<sub>3</sub>) was added to quench the reaction. The reaction was filtered through a 0.2 µm filter and measured in a spectrophotometer at 420 nm in a 1 cm cuvette. LacZ activity was calculated according to the following equation:

$$LacZ = \frac{A_{420}}{t} \cdot \frac{V_{total}}{V_{culture} \cdot OD_{600}}$$

The LacZ activities of all clones were divided by the LacZ activity exhibited by the WT intron.

### SIBR-Cas targeting and editing assays in *E. coli*

For both the targeting and the editing assays, *E. coli* MG1655 electrocompetent cells were transformed with a single variant of the pSIBR-Cas series (Table S1). 10 ng of plasmid DNA was

used to transform 20  $\mu\text{L}$  electro-competent cells as described above. Transformed cells were recovered in 1 mL LB liquid medium for 1 h at 30°C and shaking at 200 rpm. For the targeting assay, the transformants were ten-fold serially diluted in LB liquid medium and 3  $\mu\text{L}$  were used for spot dilution assays on LB agar plates containing kanamycin (50  $\text{mg L}^{-1}$ ) in the presence or absence of 2 mM theophylline and incubated for 24 h at 30°C. For the editing assay, transformants were plated on LB agar plates containing kanamycin (50  $\text{mg L}^{-1}$ ) and 7 mM theophylline and incubated at 30°C for 24 h. Colony PCR was performed on 16 colonies from each transformation to define the editing efficiencies. Triplicate transformations were used for each SIBR-Cas plasmid. Mutant colonies were sequenced through Sanger sequencing (Macrogen BV) to confirm complete deletion of the target gene.

### **SIBR-Cas targeting and editing assays in *P. putida***

40  $\mu\text{L}$  electro-competent *P. putida* cells were transformed with 200 ng of a single variant of the pSIBR-Cas series (Table S1) and recovered in 1 mL LB liquid medium for 2 h at 30°C, shaking at 200 rpm. Targeting was assayed by spot dilution assays on LB agar plates containing kanamycin (50  $\text{mg L}^{-1}$ ) in the presence or absence of 2 mM theophylline, followed by overnight incubation at 30°C. *P. putida* cells bearing the editing plasmid were plated on LB agar plates containing kanamycin (50  $\text{mg L}^{-1}$ ) and 2 mM theophylline and incubated at 30°C for 24 h. Grown colonies were screened through colony PCR to define the editing efficiency. For each SIBR-Cas plasmid, transformations were performed in triplicate. Mutant colonies were sequenced through Sanger sequencing (Macrogen BV) to confirm complete deletion of the target gene.

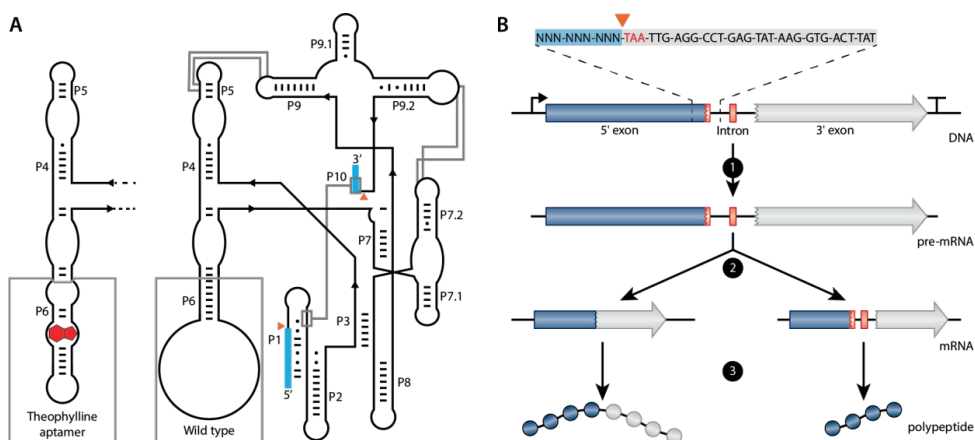
### **SIBR-Cas editing assays in *Flavobacterium IR1***

Electro-competent *Flavobacterium IR1* cells (100  $\mu\text{L}$ ) were transformed with 2  $\mu\text{g}$  plasmid of a single variant of the pSIBR-Cas series (Table S1). Transformed cells were recovered in 1 mL ASW and incubated for 4 h at 25°C, shaking at 200 rpm. Due to very low transformation efficiency, the recovered cells were transferred in 10 mL ASW liquid medium containing erythromycin (200  $\text{mg L}^{-1}$ ) and incubated for 96 h at 25°C, shaking at 200 rpm.  $10^{-6}$  or  $10^{-7}$  cells were then plated on ASW agar containing erythromycin (200  $\text{mg L}^{-1}$ ) and 2 mM theophylline. Plates were incubated at 25°C for 2-3 d and grown colonies were screened for editing through colony PCR. Each editing assay was performed in triplicate. Mutant colonies were sequenced through Sanger sequencing (Macrogen BV) to confirm complete deletion of the target gene.

## Results and discussion

### The flanking regions of the T4 *td* intron are amenable to modifications

To create a versatile gene control system, we focused on promoter-, sequence- and organism-independent mechanisms. We chose the T4 bacteriophage Group I self-splicing intron that resides in the *thymidylate synthase* (*td*) gene as the appropriate mechanism to control the expression of the GOI (Fig. 1A). The self-splicing ribozyme activity of the T4 *td* intron requires only ubiquitous cofactors such as GTP and  $Mg^{2+}$ , making its use widely applicable in bacterial species (450,451). Moreover, similar to other introns, the T4 *td* intron terminates the translation of the unspliced precursor mRNA due to the presence of in-frame stop codons (Fig. 1B). Therefore, the presence of the intron in the precursor mRNA will result in a truncated non-functional protein, whereas the spliced mRNA allows for the full translation of the protein of interest.

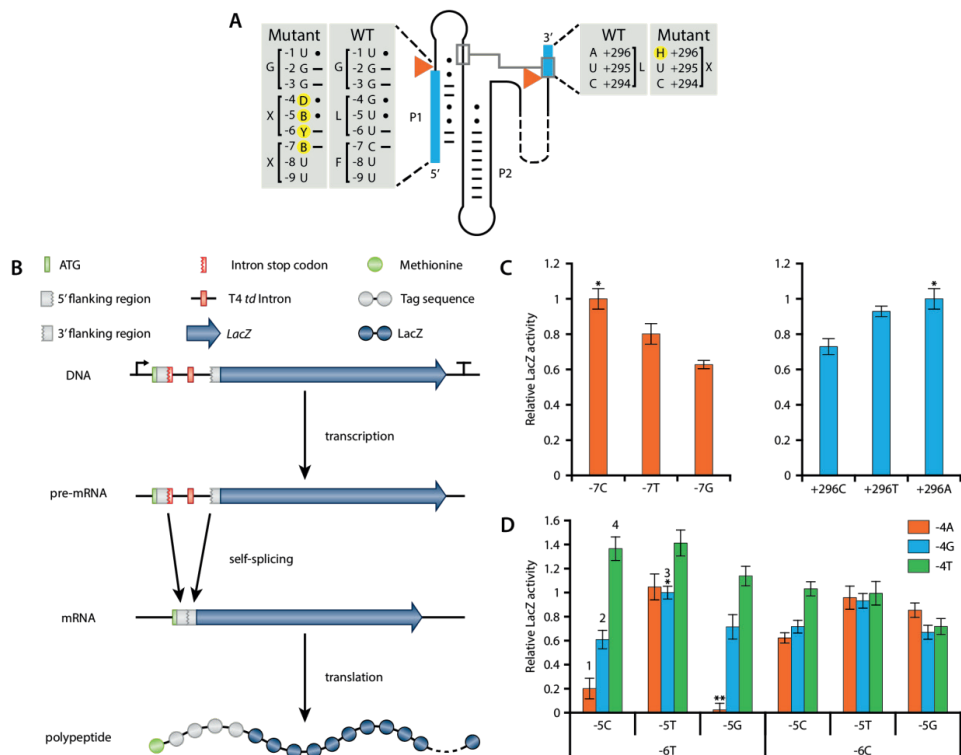


**Figure 1. Schematics and function of the T4 *td* intron.** (A) Schematic representation of the predicted secondary and tertiary structure of the wild-type (WT; right) and the theophylline-dependent T4 *td* intron (left). The structure follows the format of Cech *et al.* (1994)(452). P1 to P10 represent the pairing domains of the intron. Exon sequences are indicated as blue boxes. Orange triangles indicate the splicing sites. Base pairs are indicated by “-” and wobble pairs by “•”. The grey boxes at P6 highlight the difference between the WT and the theophylline-dependent ribozyme. Grey lines show interactions within the intron. (B) Schematic representation of the transcription and translation of a gene containing the T4 *td* intron in its open reading frame. At the top, the intron in-frame stop codon (TAA) is depicted in red, the 5' flanking region is highlighted with a blue box and a part of the intron sequence is highlighted with a light grey box. 1 depicts transcription, 2 depicts self-splicing (left path) or no self-splicing of the intron (right path) and 3 depicts translation of the full protein (left path) or the translation of a truncated protein (right path) when the intron is spliced or retained, respectively.

The naturally occurring T4 *td* intron is specific for the *td* gene because the exonic flanking regions are necessary to preserve the secondary structure of the P1 and P10 stems of the T4 *td* intron (453)

(Fig. 1A and 2A). Hence, transferring the intron along with the flanking exonic regions to another gene will disrupt the coding sequence of the target gene. On the other hand, changing the exonic flanking regions of the intron to preserve the coding sequence of the target gene may affect (or completely inhibit) the splicing activity of the intron. Since moving the T4 *td* intron without changing the flanking exons is not a viable option, we opted to create several variants of the T4 *td* intron by mutating the flanking exons. Some of the boundaries that determine the effect of the flanking exons on the splicing efficiency have been elucidated previously (453). However, to predict the effect of the altered flanking regions on splicing, we studied them in more detail.

To assess whether alterations at the flanking exons can affect the splicing efficiency, the T4 *td* intron variants were placed in between the coding sequence encoding amino acids D6 and S7 of the *LacZ* gene (Fig. 2A and B). We hypothesize that by placing the intron at the beginning of the *LacZ* gene, potential interference by exon-intron interactions or by other disturbances related to translation will be avoided (454). Splicing was assessed through the well-established  $\beta$ -galactosidase assay in *E. coli* DH10B.



**Figure 2. T4 *td* intron mutant library generation and LacZ assays.** (A) Detailed illustration of the WT and mutant 5' and 3' flanking regions of the T4 *td* intron. Exons are indicated with blue boxes. Orange triangles show the splicing sites. Base pairs are indicated by “-” and wobble pairs by “•”. Mutant nucleotides are highlighted with yellow circles and follow the IUPAC nucleotide nomenclature (B = C/G/T, D = A/G/T, H = A/C/T and Y = C/T). (B) LacZ transcription-translation cascade controlled by the T4 *td* intron library. (C) LacZ activity of position -7 and +296 mutants. “\*” indicates the WT intron and its relative LacZ activity is set to 1. All other LacZ activities are a fraction of the WT intron LacZ activity. (D) LacZ activity of all possible combinations for pair, wobble pair and mismatch at positions -6 to -4. “\*” indicates the WT intron and its relative LacZ activity is set to 1. All other LacZ activities are a fraction of the WT intron LacZ activity. “\*\*\*” indicates the stop codon UGA. The numbers above the bars refer to the intron variants that were selected for the subsequent experiments (1:Int1, 2:Int2, 3:Int3 and 4:Int4). Bars represent the means and error bars represent the standard deviation of three independent experiments.

Single base modifications at the -7 (C to T or G) or +296 (A to T or C) positions decreased the splicing activity of the intron compared to the WT T4 *td* intron sequence (Fig. 2C). Position -7 (C) preferably pairs with position +15 (G) in the WT intron since a weaker interaction in the form of a wobble base pair (T) or no interaction in the form of a mismatch (G), impeded the splicing of the intron by 20% and 35%, respectively. The opposite was observed for position +296 where a mismatch (A) allows for the highest intron splicing activity. The weak wobble base pair (T) impeded splicing by 8%, while the stronger pair (C) decreased the splicing by 29%.

Regardless of the impeded self-splicing of the mutant introns at position -7 and +296, self-splicing was still observed indicating that the flanking regions of the T4 *td* intron are amenable to modifications. To this end, we created pair, wobble or mismatch base substitutions at the -4, -5 and -6 positions and characterized the self-splicing activity of the resulting T4 *td* intron variants (Fig. 2D). For simplicity reasons, the rest of the flanking regions of the T4 *td* intron (including the -7 and +296 positions) were kept the same as the WT intron sequence.

Surprisingly, several intron variants showed better LacZ activity compared to the WT intron (-4G, -5T, -6T; Fig. 2D). A mismatch at position -4 (T) is preferred in almost all variants, except for those in which both -5 (G) and -6 (C) positions are mismatched too. Compared to the WT intron, a 40% increase in LacZ activity was observed when position -4 was mismatched (T) accompanied by a paired (C) or a wobble paired (T) -5 position and a paired (T) -6 position. A wobble base pair at position -5 (T) negates to a large extent the effect that -4 and -6 have on the splicing. In contrast, a pair (C) or a mismatch (G) at position -5 and depending on -4 and -6 positions, can alter the splicing efficiency from very high (136%) to very low (20%). Position -6 in general appears in favour of being paired (T). However, the complete stabilisation of the secondary structure of P1 (-4A, -5C, -6T) is inhibiting splicing almost completely as it shows 20% relative LacZ activity. Completely mismatching positions -4 to -6 (-4T, -5G, -6C) impedes splicing to around 71%. Complete stabilization or complete destabilization of the -4, -5 and -6 positions of the P1 stem was previously observed by Pichler *et al.* (2002) (453). However, the authors report contradicting results to our study as both their stabilized and destabilized variants show increased splicing efficiency. The observed differences in splicing efficiency by stabilizing or destabilizing the P1 stem may be attributed to the different experimental setup as we investigated splicing efficiency based on enzymatic activities whereas Pichler *et al.* (2002) performed *cis* splicing assays by isolating total RNA from *E. coli* cells carrying the different intron variants. Lastly, the -6T, -5G, -4A acts as a negative control as this combination forms a stop codon (UGA), as reflected by the absence of relative LacZ activity using this combination.

Taken together, these results demonstrate that T4 *td* intron variants were generated with a range of splicing efficiencies, allowing for tuneable control over the LacZ protein expression based solely on the intron variant. In addition, the transfer of the T4 *td* intron variants from the *td* gene to the start of the *LacZ* gene demonstrates the flexibility of the intron and its potential use as a tunable gene expression control mechanism.

## SIBR-Cas targeting efficiency is tuneable and inducible

To translate our setup to a CRISPR-Cas engineering context, we tested the ability of the T4 *td* intron variants to control the expression of a Cas nuclease. We selected Cas12a from *Francisella novicida* (FnCas12a), due to our prior expertise on this particular Cas nuclease (416). We selected four intron variants with distinct splicing efficiency (Int1: -4A, -5C, -6T; Int2: -4G, -5C, -6T; Int3: -4G, -5T, -6T; Int4: -4T, -5C, -6T; Fig. 2D) and inserted them directly after the start codon of the *FnCas12a* gene (Fig. 3A). The intron variants are numbered in order of increasing splicing efficiency with Int1 having the worst and Int4 the highest splicing efficiency. Moreover, to develop a tool compatible for most non-model organisms, where inducible promoters are either not known or not characterised, we used the constitutive lacUV5 promoter as a representative constitutive promoter to drive the expression of the Intron-Cas12a variants in *E. coli*.

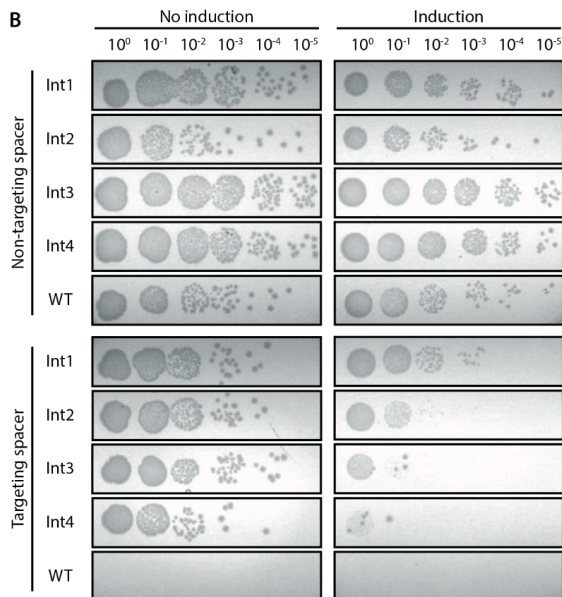
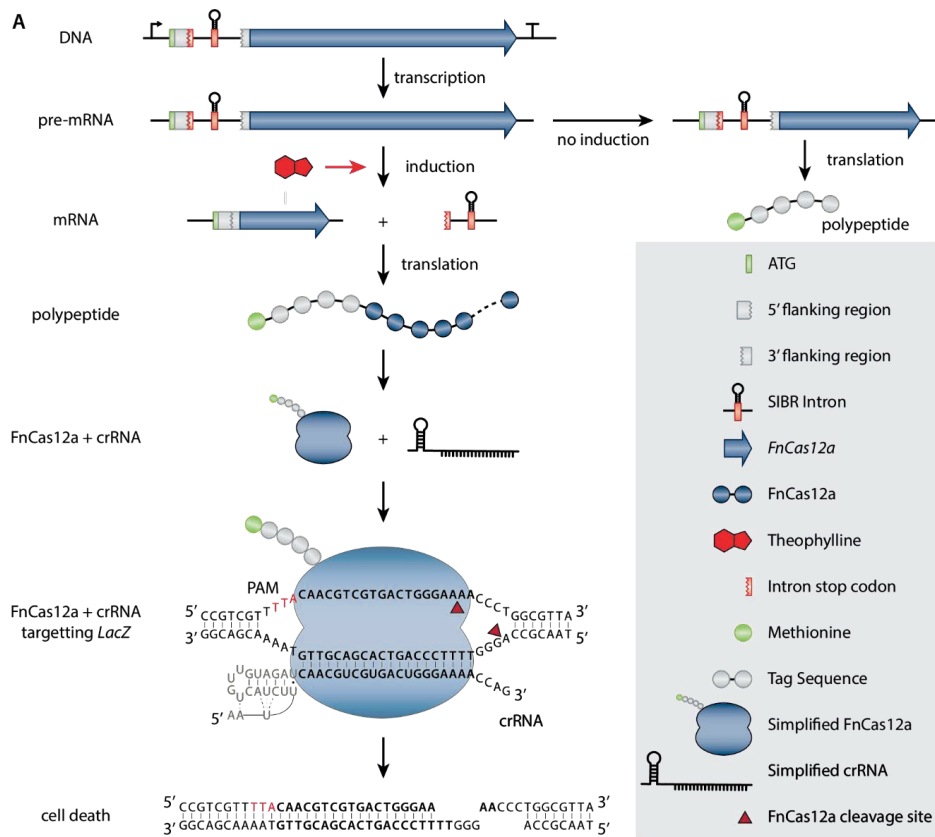
According to our design, unspliced precursor mRNAs will result short (5 amino acid) peptides due to the TAA stop codon present at the start of the intron (position +1, +2, +3). In contrast, excision of the intron will result in the full FnCas12a protein fused to a short 4 amino acid tag (SSGL for Int1,2 and 4 or SLGL for Int3) at its N terminus. Furthermore, to make splicing inducible, we added a theophylline aptamer at the P6 stem loop of the T4 *td* intron as previously described (448), resulting in a new tightly-controlled CRISPR-Cas system, which we named SIBR-Cas (Fig. 3A). Wild-type FnCas12a (WT-*FnCas12a*, without intron) was used as a reference for comparison to the SIBR-Cas variants. The efficiency of targeting for the SIBR-Cas and WT-*FnCas12a* variants was assessed by transforming *E. coli* MG1655 cells with plasmids expressing either of the different Cas12a variants with either a *LacZ* targeting (T) or a non-targeting (NT) crRNA. After transformation, the cells were serially diluted and plated on media with or without the presence of the theophylline inducer (Fig. 3B and S1).

The NT crRNA controls showed colonies up to the  $10^{-5}$  dilution, both in the presence or absence of theophylline for all the SIBR-Cas variants and WT-*FnCas12a* (Fig. 3B). No colonies were observed when the T crRNA and the WT-*FnCas12a* combination was used, regardless of induction with theophylline, demonstrating the strong Cas12a-mediated counter-selection. In contrast, transformants targeting *LacZ* and expressing either of the four SIBR-Cas variants (Int1-4), showed a notable reduction in colony number formation only when theophylline was present in the medium. Intriguingly, the targeting efficiency directly reflected the splicing efficiency of the intron variants tested for *LacZ* (Fig 2D), with Int1 (-4A, -5C, -6T) showing the least targeting efficiency (10 fold reduction) and Int4 (-4T, -5C, -6T) showing the highest targeting efficiency ( $10^3$  fold reduction) upon induction.

Our results demonstrate consistency in the splicing activity of the intron variants regardless of its genomic context (*LacZ* or *FnCas12a*). We therefore hypothesized that the minimalistic nature of the ‘tag’ sequence that SIBR leaves behind after splicing (i.e. twelve nucleotides at the 5’ of the mRNA, four amino acid residues at the N-terminus of the resulting protein) will be of minor influence on the stability, translation and folding efficiency, although this might be different with other genes. The difference between the intron variants Int1, Int2 and Int4 is only a single nucleotide (Int1: UCCUCAGGU; Int2: UCCUCGGGU; Int4: UCCUCUGGU). Therefore, it is unlikely to be of major importance for the stability and folding of the mRNA, strongly suggesting that the difference in enzyme activity can mostly be attributed to the splicing efficiency of the intron. Moreover, the varied protein activity (*LacZ* and *FnCas12a*) is unlikely to be caused by the small N-terminal tag as Int1, Int2 and Int4 have the exact (SSGL for Int1, 2 and 4) or very similar (SLGL for Int3) N-terminal amino acid sequence.



Chapter 6

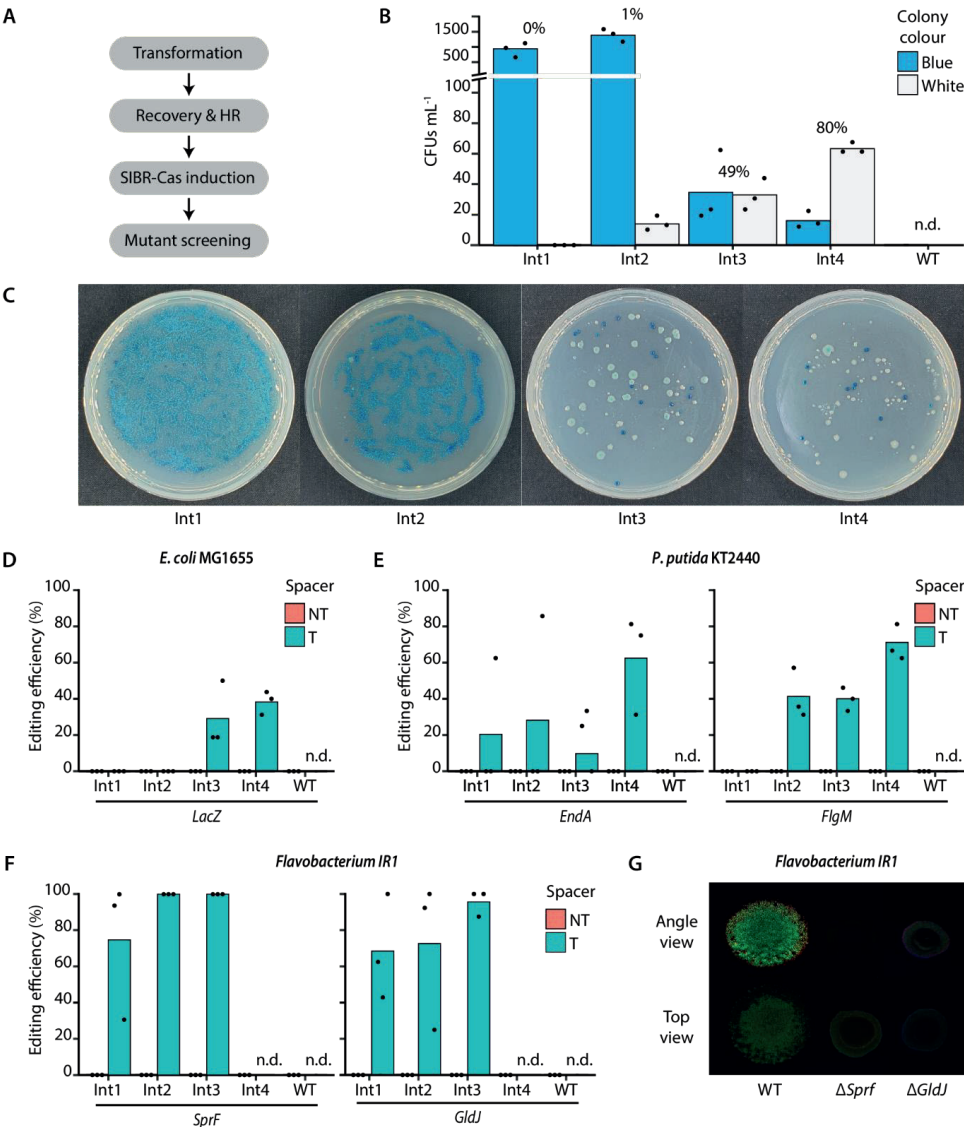


**Figure 3. SIBR-Cas targeting assays in *E. coli* MG1655.** (A) Schematic diagram of SIBR-Cas in the presence or absence of the theophylline inducer. In the presence of theophylline, the intron is self-spliced leading to the translation of FnCas12a with an additional 4 amino acids at its N terminus (left path). In the presence of a targeting CRISPR-RNA (crRNA), FnCas12a targets the genome and cleaves it, causing cell death. The *LacZ* target site is shown as an example. The bold 5'-CAACGTCGTGACTGGGAAAA-3' sequence indicates the target *LacZ* protospacer. The red 5'-TTA-3' nucleotides indicate the PAM sequence. The repeat sequence of the crRNA is indicated with grey nucleotides. In the absence of theophylline, the intron cannot splice itself out of the pre-mRNA, leading to the translation of a short, 5 amino acid long, peptide (right path). (B) Targeting and induction efficiency of SIBR-Cas in *E. coli* MG1655. The genome of *E. coli* MG1655 was targeted at the *LacZ* locus with a targeting or a non-targeting spacer. Four intron variants (Int1-4; bad splicer to good splicer) were used to control the translation of *FnCas12a* and a WT-*FnCas12a* was used as a control. Transformed *E. coli* cells were serially plated on LB solid medium with the appropriate antibiotic in the absence (left panel) or the presence (right panel) of the theophylline inducer.

### SIBR-Cas is an efficient genome engineering tool for bacteria

For efficient genome editing in bacteria, HR should precede CRISPR-Cas counter-selection (Fig. 4A). To assess whether tight control over CRISPR-Cas targeting could bolster the efficiency of CRISPR-Cas mediated genome editing by allowing more time for HR to occur, we used SIBR-Cas and targeted the *LacZ* gene of *E. coli* MG1655 for knock-out through HR and CRISPR-Cas counter-selection using a blue/white screening colony assay. To facilitate HR, we added 500 bp up- and down-stream homology arms to the plasmids expressing the four SIBR-Cas (Int1-4) and WT-*FnCas12a* variants that target the *LacZ* gene. After 1 hour recovery, we induced the expression of the SIBR-Cas variants to counter-select the WT from the mutant colonies.

The WT-*FnCas12a* variant targeting the *LacZ* gene produced no colonies, demonstrating the targeting efficiency of WT-*FnCas12a* but also the inefficient HR system of the WT *E. coli* MG1655 strain (Fig. 4B). In contrast, SIBR-Cas variants produced multiple colonies of which 80% of the total CFUs mL<sup>-1</sup> were white when Int4 was used, followed by Int3 (49%), Int2 (1%) and Int1 (0%) variants (Fig. 4B and C). Similar to the previous results, the high editing efficiencies obtained with Int4 suggest that its high splicing efficiency translates into a stronger counter-selective pressure. No white colonies were observed for the non-targeting controls, demonstrating that the efficiency of editing without CRISPR-Cas counter-selection is negligible (Fig. S2).



**Figure 4. SIBR-Cas genome editing assays in *E. coli* MG1655, *P. putida* KT2440 and *Flavobacterium* IR1. (A)** Schematic for SIBR-Cas editing procedure. The time required for recovery and HR differs amongst the different bacterial species as described in the materials and methods section. **(B)** Editing efficiency of the *LacZ* gene in *E. coli* MG1655. Blue/white screening was performed to distinguish the edited (white) from the unedited (blue) colonies when using either of the four different SIBR-Cas variants Int1 (0%), Int2 (1%  $\pm$  0.35%), Int3 (49%  $\pm$  8.75%) and Int4 (80%  $\pm$  5.57%) or the WT-*FnCas12a* (0%, n.d.). The percentage on top of each variant indicates the percentage of white colonies from the total number of colony forming units mL<sup>-1</sup> (CFUs mL<sup>-1</sup>). **(C)** Representative plates of edited *E. coli* MG1655 cells at the *LacZ* locus using the four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer). **(D)** Unbiased (omitting the presence of X-Gal in the medium) editing efficiency of the *LacZ* gene using the four different SIBR-Cas variants Int1 (0%), Int2 (0%), Int3 (29%  $\pm$  18.04%), Int4 (38%  $\pm$  6.41%) or the WT-*FnCas12a*

(0%, n.d.). All of the NT controls showed 0% targeting efficiency. (E) Editing efficiency of the *EndA* [Int1 (21%  $\pm$  36.08), Int2 (29%  $\pm$  49.49%), Int3 (19%  $\pm$  17.35%), Int4 (63%  $\pm$  27.24%) or the WT-*FnCas12a* (0%, n.d.)] and *FlgM* [Int1 (0%), Int2 (41%  $\pm$  13.84%), Int3 (40%  $\pm$  6.41%), Int4 (70%  $\pm$  9.85%) or the WT-*FnCas12a* (0%, n.d.)] genes in *P. putida* KT2440. All of the NT controls showed 0% targeting efficiency. (F) Editing efficiency of the *SprF* [Int1 (75%  $\pm$  38.29), Int2 (100%  $\pm$  0%), Int3 (100%  $\pm$  0%), Int4 (0%, n.d.) or the WT-*FnCas12a* (0%, n.d.)] and *GldJ* [Int1 (68%  $\pm$  29.03%), Int2 (72%  $\pm$  41.26%), Int3 (96%  $\pm$  7.22%), Int4 (0%, n.d.) or the WT-*FnCas12a* (0%, n.d.)] genes in *Flavobacterium* IR1. Individual bars represent the mean of triplicate experiments and “•” represents the value of each replicate. N.d., not determined. All of the NT controls showed 0% targeting efficiency. (G) Comparison between WT *Flavobacterium* IR1 and  $\Delta SprF$  and  $\Delta GldJ$  strains generated with SIBR-Cas. Images were taken after incubation at room temperature for 2 d by inoculating 3  $\mu$ L spot on ASWBC solid medium (ASW supplemented with black ink and carrageenan). The WT strain (left) is 18 mm across.

Since disruption of *LacZ* can also be achieved through non-HR mediated approaches (spontaneous mutations or occasional error-prone DNA repair following DNA cleavage by Cas12a), not all gene deletions can be screened phenotypically. Therefore, we repeated our experiment, but X-gal was omitted from the medium to eliminate the possibility of false-positives. Randomly selected colonies that were obtained were screened by PCR for *LacZ* deletion showing a 0%, 0%, 29% and 38% editing efficiency for Int1, Int2, Int3 and Int4 SIBR-Cas variants, respectively (Fig. 4D and S3). The WT-*FnCas12a* variant targeting *LacZ* did not yield any colonies and all the colonies obtained from the NT controls had the intact, wild-type *LacZ* locus. The observed decrease in editing efficiency (compared to the blue/white screening) might be attributed to spontaneous *LacZ* mutations that escape CRISPR-Cas counter-selection. Nevertheless, an editing efficiency of 38% was observed when SIBR-Cas Int4 was used without the use of recombinases or any other complex systems.

Following the successful use of SIBR-Cas in *E. coli*, we continued to demonstrate the efficiency of SIBR-Cas by testing it in other bacteria. For this purpose, we selected *Pseudomonas putida* KT2440, an organism with rather complex engineering tools and low HR efficiencies (64). After establishing the successful induction and targeting of SIBR-Cas in *P. putida* (Fig. S4), genome editing experiments were conducted to knock-out the *EndA* and *FlgM* genes. High knock-out efficiencies were obtained when Int4 was used, with editing efficiencies of 63% and 70% for *EndA* and *FlgM*, respectively (Fig. 4E, S5 and S6). Lower editing efficiencies (<40%) were observed for the other introns, whereas no transformants were obtained with the WT-*FnCas12a* variant when used with a targeting spacer. Control transformants with the NT crRNA had a WT genotype (Fig. S6 and S7).

Lastly, we focused on the non-model organism *Flavobacterium* IR1, which is a recent isolate best known for its iridescent, structural colour (455-457). The lack of genomic tools, low transformation

efficiency and the low HR efficiency of IR1 are currently the main bottlenecks holding back the fundamental characterization and commercial exploitation of this phenomenon (i.e. development of photonic paints). In addition, *Flavobacterium* species do not have a canonical RBS (TAAAA rather than GGAGG) (440-442) which render other widely applicable gene control systems, such as Ribo-Cas (74), inadequate for this type of bacterial species. To this end, we transformed *Flavobacterium* IR1 with a series of pSIBR plasmids (Table S1) and assessed the editing efficiency after 96 h of growth by plating the transformants on plates that contained the theophylline inducer. By using SIBR-Cas, 100% editing efficiency was achieved for both Int2 and Int3 variants when *SprF* was targeted for knock-out (Fig. 4F and S8). In accordance, the phenotype of the *SprF* mutants displayed similar characteristics when compared to previous studies (455,456) (Fig. 4G). 100% editing efficiency was also achieved for some of the replicates of Int2 (1 out of 3) and Int3 (2 out of 3) when the *GldJ* gene was targeted for knock-out (Fig. 4F and S9). Furthermore, SIBR-Cas was successful in creating a clean *GldJ* mutant, that could not be achieved by previous endeavours using transposon mutagenesis (455,456). Similar to the *SprF* mutant, the *GldJ* mutant could not develop the iridescence phenotype displayed by WT *Flavobacterium* IR1 (Fig. 4G). The WT-*FnCas12a* variants did not yield any colonies when the *SprF* or *GldJ* genes were targeted for knock-out and the non-targeting controls were all confirmed to be unedited (Fig. 4F, S8 and S9). To achieve high editing efficiency in *Flavobacterium* IR1, a 96 h incubation (without Cas induction) was required as shorter incubation times (24 and 48 hours) did not yield any viable colonies (Fig. S10). This is in agreement with our assumption that more time is required for HR to occur before the induction of CRISPR-Cas counter-selection and that dsDNA-break stimulated repair by HR was of minor importance. These results also demonstrate that, in this context, the prolonged incubation time did not enrich for escapees (crRNA, protospacer, PAM or Cas12a mutations). Surprisingly and in contrast to *E. coli* and *P. putida*, Int4 failed to sustain growth in the recovery stage (data not shown) and hence was not plated. We expect this to be caused by leakiness of the Int4 variant during the recovery phase.

Highly efficient CRISPR-Cas tools have previously been developed for *E. coli* MG1655 (61,62,458-460). The study by Jiang *et al.* (2015) (62) for example, demonstrated editing efficiencies close to 100%. This study uses a double vector system with inducible promoters to induce Cas counter-selection and bolstered homologous recombination with the use of the  $\lambda$  red recombinase. Another study used a more simplistic one-plasmid system (63), where the control of Cas9 was under the inducible L-arabinose promoter, allowing for inducible counter-selection. Without the use of any recombinases, 20.8% knock-out efficiency was observed when the *PoxB* gene was targeted for editing. In contrast, when the  $\lambda$  red recombinase was included, 100% editing

efficiency of *PoxB* was observed, indicating again the importance of exogenously expressed recombinases for high editing efficiency. In general, even for the model organism *E. coli*, the use of recombinases and inducible promoters is required for efficient genome engineering. SIBR-Cas was specifically developed to alleviate the need for these requirements, as both recombinases and inducible promoter might not be available for (or incompatible with) the host, especially in non-model organisms.

Similar to *E. coli*, various CRISPR-Cas tools have been developed for *P. putida* KT2440 (64,461-464). In the study of Sun *et al.* (2018), for example, a double vector system was used to express the  $\lambda$  red recombinase genes, the Cas nuclease (either Cas9 or Cas12) and the gRNA scaffold. Here, the requirement for exogenously expressed recombinases was even more evident, showing 93% editing with and 0% without the use of the  $\lambda$  red recombination system respectively (64). Nevertheless, successful genome editing in *P. putida* KT2440 has been achieved without the use of exogenous recombinases (464). Here, a plasmid was constructed expressing a thermostable variant of Cas9 (ThermoCas9) under the control of a 3-methylbenzoate-inducible Pm-promoter, a sgRNA targeting the *PyrF* gene and contained a homologous recombination template. Transformants were first screened for plasmid integration by PCR, after which a single colony was used for overnight growth in selective media, followed by inoculation in fresh medium. After another 6 h of growth in inducing media, dilutions were plated, ultimately resulting in a 50% editing efficiency rate (464). Lastly, ssDNA recombineering and CRISPR-Cas counter-selection approaches for editing *P. putida* KT2440 have been developed by Aparicio *et al.* (2018) (461) and Wu *et al.* (2019) (462). Both studies used inducible recombinases (Ssr or Red $\beta$ ) expressed from an additional plasmid. Of note is the editing efficiency of *EndA* (54.2%) and *FlgM* (93.2%) as reported by Aparicio *et al.* (2018) (461) compared to our results (63% for *EndA* and 70% for *FlgM*). Whilst high editing efficiencies are reported by other CRISPR tools for *P. putida* KT2440, SIBR-Cas is a more widely applicable and simplistic approach (single plasmid, no need for integrant verification, no recombinase or inducible promoters needed) achieving similar or, in some instances, higher editing efficiencies.

We further showed that SIBR-Cas is compatible for non-model organisms as exemplified by our successful engineering attempt in *Flavobacterium* IR1. Since *Flavobacterium* IR1 is a relatively new isolate (455-457), inducible promoters or other gene regulation systems are not described yet. Therefore, SIBR presented a simple and efficient solution to control the expression of CRISPR-Cas and to achieve editing in case the more canonical strategy of using homologous recombination combined with CRISPR-Cas counter-selection fails, as demonstrated here (Fig. 4F and S10). This is the first report for genome engineering of *Flavobacterium* IR1 using CRISPR-Cas and can serve as the basis for the easy and efficient engineering of other *Flavobacteria* species.

Collectively, our results show that SIBR-Cas is a tight and inducible genome engineering tool that can successfully be applied to a wide variety of bacterial species. By delaying CRISPR-Cas counter-selection and thus allowing enough time for HR to occur, we achieved high editing efficiencies in WT model and non-model bacterial species that naturally have very low HR efficiencies. We propose that this tool could be the solution for the difficulties of using CRISPR-Cas for prokaryotic genome engineering, especially in organisms where HR efficiencies are low, the use of recombinases is not possible or inducible promoters are not known or not characterized. We also foresee that SIBR-Cas will significantly decrease the time required for and complexity of CRISPR-Cas mediated genome engineering in prokaryotes.

Lastly, despite the success of SIBR-Cas in the three WT bacterial species chosen for our study, replicative plasmids and transformation protocols existed for all the three species. Therefore, for the application of SIBR-Cas in different (non-model) bacterial species, the existence of replicative plasmids and transformation protocols is a prerequisite. Also, different counterselection schemes may need to be developed for each different bacterium (due to differences in doubling time and/or efficiencies of recombination). For example, 96 h of growth was required for *Flavobacterium* IR1 to obtain knock-out mutants, whereas only 1 h was required to obtain knock-out mutants in *E. coli* MG1655. Prolonged incubation time, however, may increase the chances of having escapee mutants. This was likely the case in our study as we observed variable editing efficiencies even amongst biological replicates (Fig. 4D-F). This observation may imply that escapees developed early in the recovery phase, replicated further during the recovery phase and then avoided counterselection upon induction. The unpredictable nature of these events demonstrates the need to include replicates for these experiments, as also illustrated by the variable editing obtained in this study. Moreover, we shall not ignore several factors that may affect the splicing efficiency of the T4 *td* intron and hence the editing efficiency of SIBR-Cas. For example, the splicing efficiency of the T4 *td* intron may be affected by temperature, the salt and mineral concentration present in the growth medium, the pH of the growth medium and the presence of splicing inhibitors (e.g. certain antibiotics and co-factors) (465-475). Also, the uptake of the theophylline inducer by the bacterial cell is a requirement for splicing of SIBR; although theophylline uptake is common among diverse bacterial species (74,476). To turn SIBR-Cas into a truly universal genome engineering tool for bacteria, future studies should focus on the generation of SIBR systems with a variety of aptamers and a variety of self-splicing introns that overcome the aforementioned limitations.

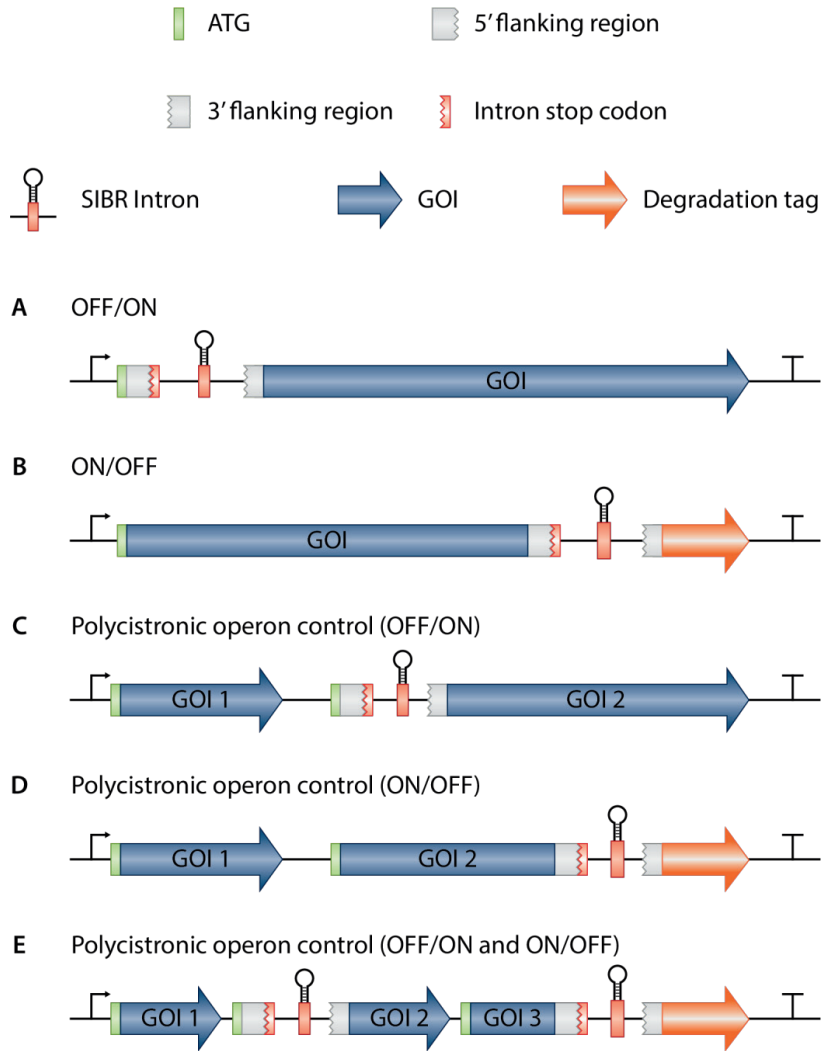
**SIBR-X as a modular, tight and inducible protein expression tool**

SIBR was successfully applied to control the expression of the *FnCas12a* gene in an OFF to ON manner. We suggest that SIBR-X (where X can be any gene/RNA of interest) can be a broader gene regulation tool for virtually any GOI (Fig. 5A). This is mainly attributed to the host factor-independent splicing mechanism of the intron variants created during this study. Furthermore, our design of placing the intron directly after the ATG start codon means that it should be compatible with most GOI, leaving only a short four amino acid tag at the N-terminus of the POI, diminishing the risk of interfering with the protein's functionality. Therefore, with the combination of the intron variants and the theophylline inducer concentration, a temporal and tuneable gene expression can be achieved.

SIBR can also be used as an ON to OFF switch (Fig. 5B). For example, SIBR can be inserted at the 3' of the coding sequence with a downstream degradation tag (e.g. SsrA degradation tag). This design will allow for constitutive translation of the POI in the absence of the inducer (terminated at the stop codon of the intron), but will trigger rapid protein degradation after splicing of the intron due to the attached degradation tag. Akin to other down-regulation approaches (such as RNAi and CRISPRi), SIBR-mediated knockdown efficiencies are highly dependent on the turnover speed (and stability) of the protein in question. Other fusions can be envisioned as well, such as a (nuclear) localization tags, signal peptides, etc. Lastly, SIBR can be used as a polycistronic operon control mechanism in different configurations (Fig. 5C, D and E). This approach will be especially useful in organisms where temporal and inducible expression is difficult to achieve by other means (e.g. operons with multiple and/or uncharacterized promoters and terminators).

Conclusively, we foresee various applications within industry and fundamental research, where SIBR-X can be a valuable tool in both model and non-model organisms.





**Figure 5. Potential applications of SIBR.** (A) OFF to ON switch by interrupting the translation of the GOI. (B) ON to OFF switch by degrading the POI after inducible attachment of a degradation tag. The degradation tag can be replaced by a localization tag as well. (C) Polycistronic operon control allowing constitutive expression of the 1<sup>st</sup> gene (Gene of interest 1; GOI 1) and inducible OFF to ON expression of the 2<sup>nd</sup> gene (GOI 2). (D) Polycistronic operon control by allowing constitutive expression of both GOI 1 and GOI 2 and inducible protein degradation of GOI 2 upon induction. (E) A combination of all the potential SIBR applications in one single polycistronic operon.

## Funding

This work was supported by the Graduate School VLAG, Wageningen University and Research, Netherlands; R.H.J.S was supported by a VENI grant [016.Veni.171.047], awarded to R.H.J.S, from ‘The Netherlands Organization for Scientific Research’ (NWO); J.v.d.O was supported by the ‘European Research Council’ (ERC) [ERC-AdG-834279], awarded to J.v.d.O.

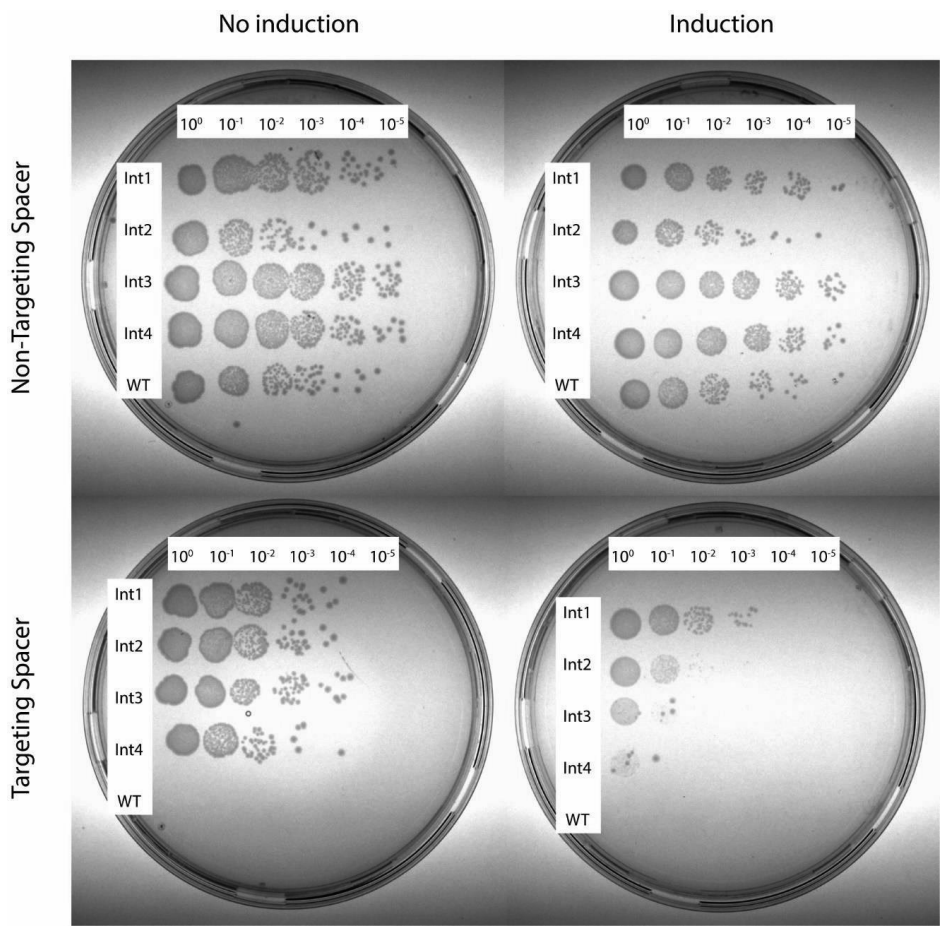
*Conflict of interest statement.* The authors (C.P., S.C.A.C, J.v.d.O. and R.H.J.S) have filed a patent application based on the results reported in this study.

## Acknowledgements

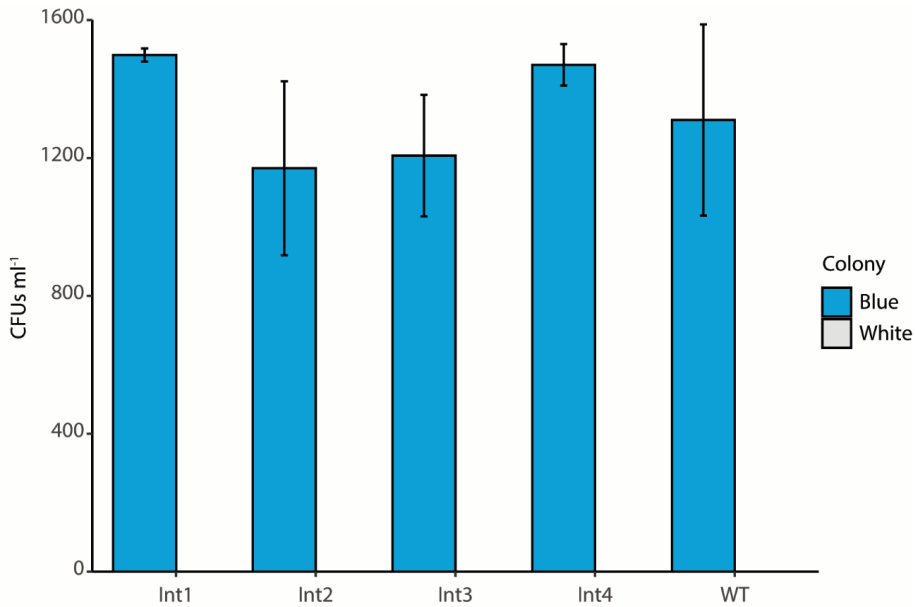
We would like to thank Hoekmine B.V. for sharing their IR1 strain and expertise. We thank Rob Joosten and Steven Aalvink for their technical support.

*Author contribution:* C.P., S.C.A.C., A.Q.A., S.W.M., J.v.d.O. and R.H.J.S. designed the study. C.P., S.C.A.C., A.Q.A., B.A.P. and E.A.G. created all the plasmids. S.C.A.C. and E.A.G. generated and characterized the intron variants. C.P. and A.Q.A. generated and characterized SIBR-Cas. C.J.I. provided lab facilities and expertise for handling *Flavobacterium* IR1 through Hoekmine B.V.. C.P., S.C.A.C. and R.H.J.S. wrote the manuscript and all authors contributed in reviewing, editing and approving the final manuscript.

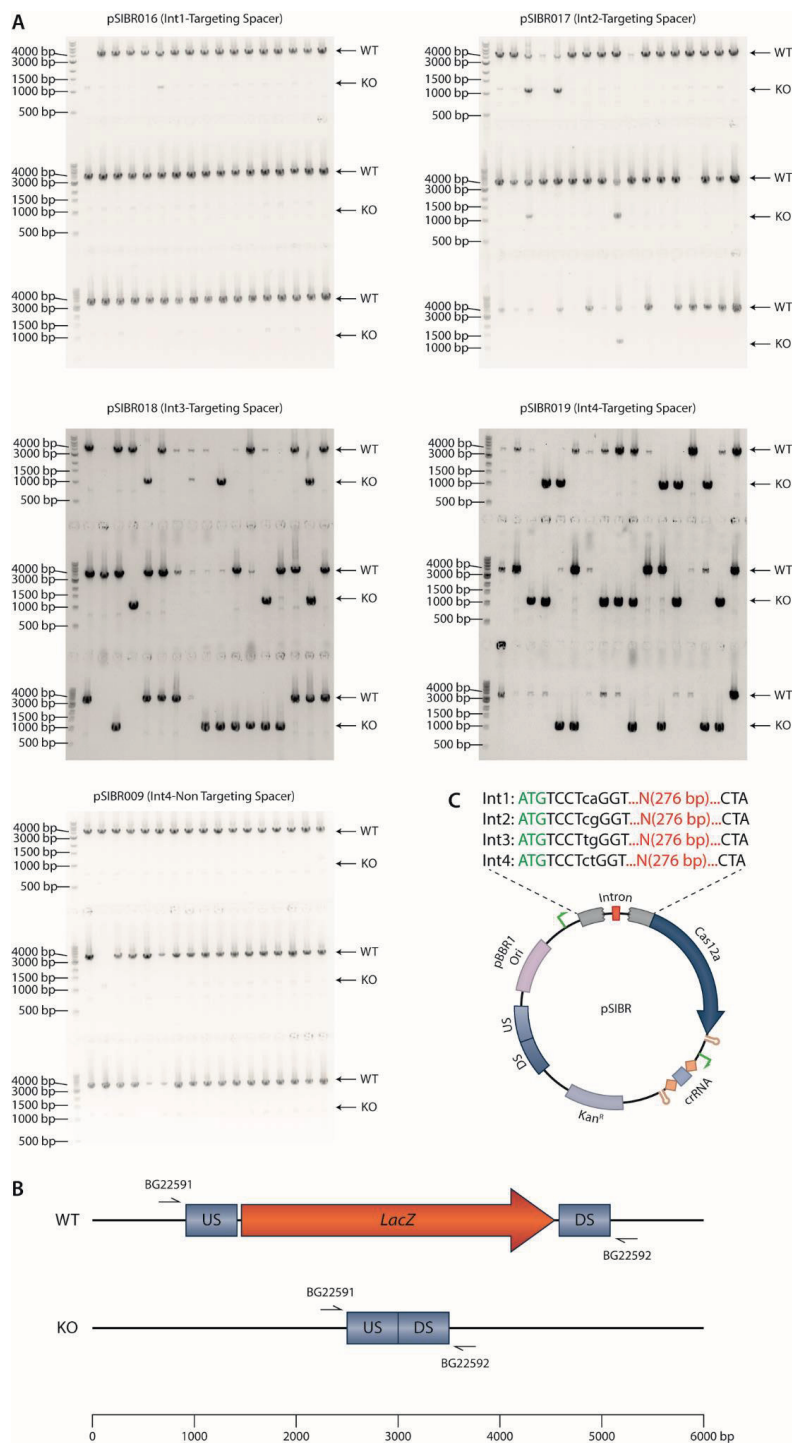
Supplementary information



**Figure S1. Raw data of SIBR-Cas targeting assay in *E. coli* MG1655.** Four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer) and the WT-*FnCas12a* were used for counter-selection by targeting *LacZ* with a targeting or a non-targeting spacer (Table S1). Transformants were serially diluted 5 times and plated on LB solid medium containing kanamycin (50 mg L<sup>-1</sup>) in the presence or absence of the theophylline inducer (2 mM).

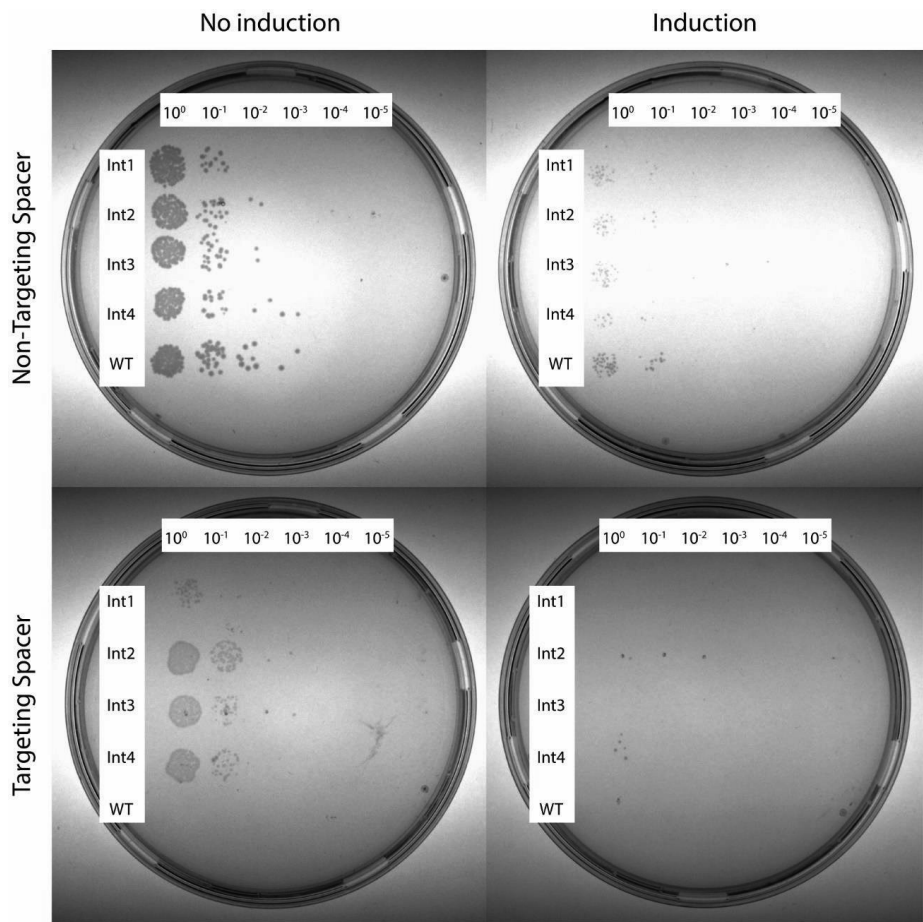


**Figure S2. Blue/white colony screening of *E. coli* MG1655 transformed with plasmids containing non-targeting spacers.** *E. coli* MG1655 was transformed with plasmids containing either of the four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer) or the WT-*FnCasI2a*, a non-targeting spacer and 500 bp homology arms to facilitate the knock-out of *LacZ* (Table S1). Values and error bars represent the means and s.d. of triplicate experiments.

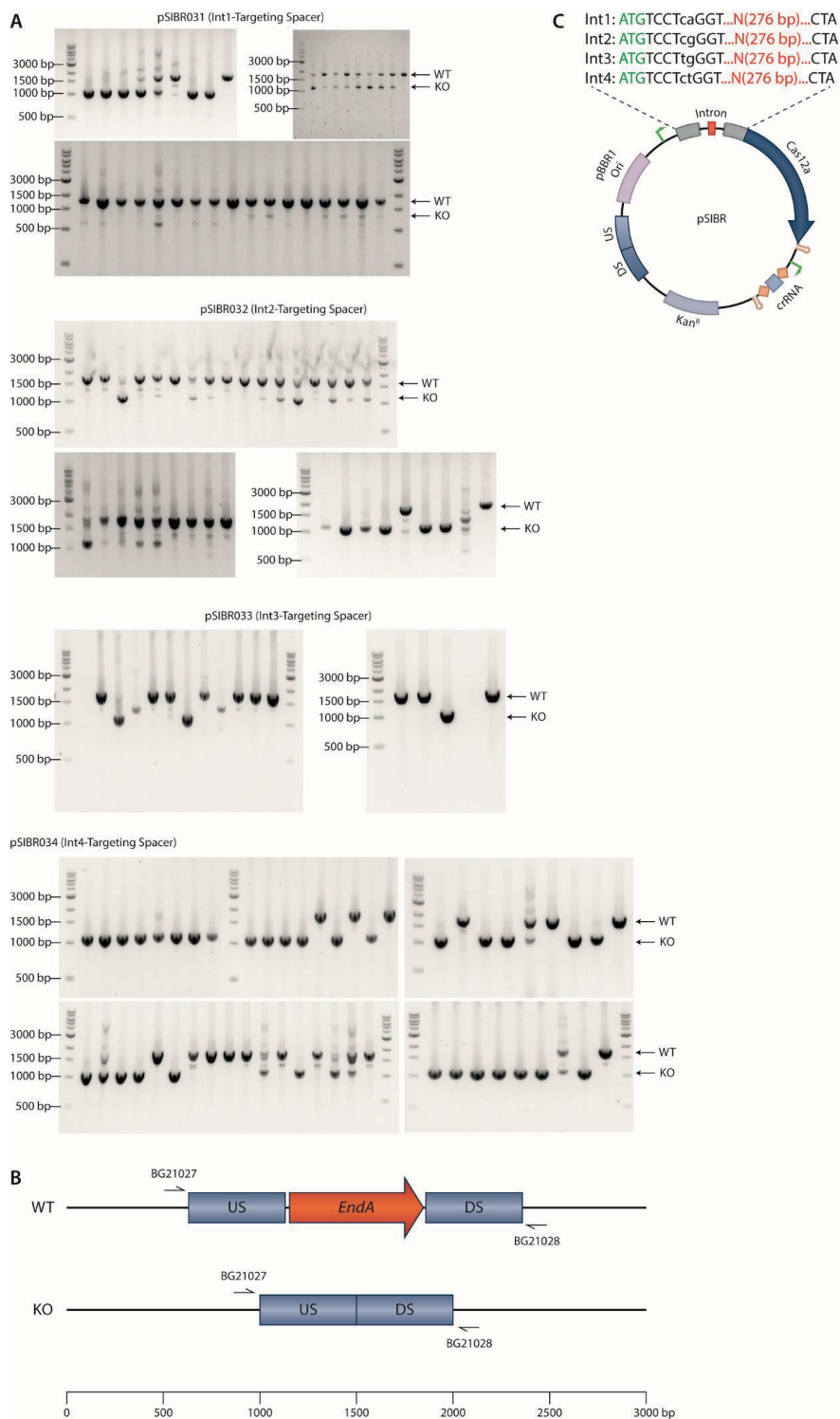


**Figure S3. *LacZ* knock-out using SIBR-Cas.** (A) *E. coli* MG1655 was transformed with plasmids containing either of the four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer) or the WT-*FnCas12a*, a

*LacZ* targeting or a non-targeting spacer and 500 bp homology arms to facilitate the knock-out of *LacZ* (Table S1). Each variant was performed in triplicate and 16 colonies (if present) were randomly selected from each replicate for colony PCR. The WT-*FnCasI2a* variant targeting *LacZ* did not yield any colonies. The non-targeting variants were all WT and for this reason only Int4 is shown as a representative. Mix amplicons (WT and knock-out bands) were not counted for the total knock-out efficiency percentage. WT: 4229 bp, Knock-out (KO): 1154 bp. **(B)** Schematic representation of WT or edited (KO) *E. coli* MG1655 genome at the *LacZ* locus. BG22591 and BG22592 represent the primers used for PCR (Table S2). **(C)** Schematic representation of the pSIBR plasmids used to knock-out *LacZ*. Int1-4 represent the four different SIBR-Cas variants and their sequence difference is depicted with lowercase nucleotide letters.



**Figure S4. Raw data of SIBR-Cas targeting assay in *P. putida* KT2440.** Either of the four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer) or the WT-*FnCas12a* were used for counter-selection by targeting *EndA* with a targeting or a non-targeting spacer (Table S1). Transformants were serially diluted 5 times and plated on LB medium containing kanamycin ( $50 \text{ mg L}^{-1}$ ) in the presence or absence of the theophylline inducer ( $2 \text{ mM}$ ).





## Chapter 6

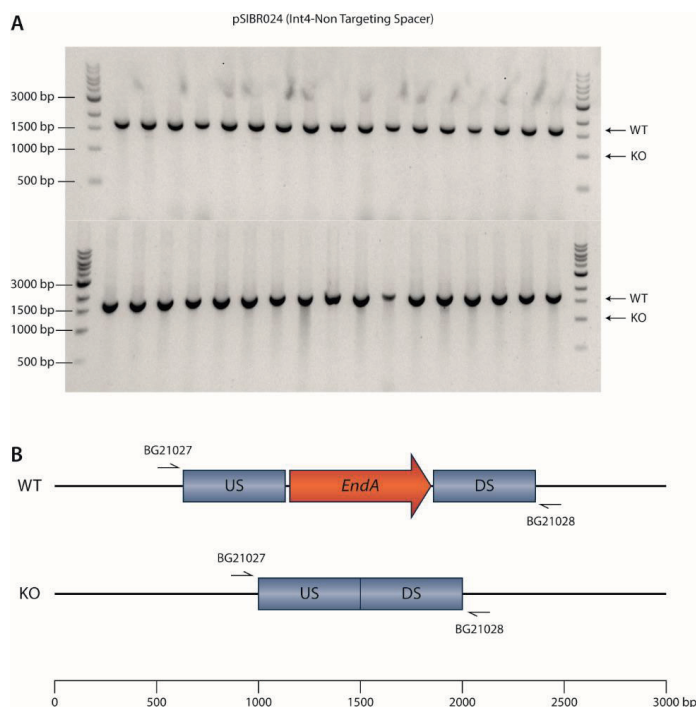
**Figure S5. *EndA* knock-out using SIBR-Cas.** (A) *P. putida* KT2440 was transformed with plasmids containing either of the four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer) or the WT-*FnCas12a*, an *EndA* targeting spacer and 500 bp homology arms to facilitate the knock-out of *EndA* (Table S1). Each variant was performed in triplicate and 16 colonies (if present) were randomly selected from each replicate for colony PCR. The WT-*FnCas12a* variant targeting *EndA* did not yield any colonies. Mix amplicons (WT and knock-out bands) were not counted for the total knock-out efficiency percentage. WT: 1814 bp, Knock-out (KO): 1121 bp. (B) Schematic representation of WT or edited (KO) *P. putida* KT2440 genome at the *EndA* locus. BG21027 and BG21028 represent the primers used for PCR (Table S2). (C) Schematic representation of the pSIBR plasmids used to knock-out *EndA*. Int1-4 represent the four different SIBR-Cas variants and their sequence difference is depicted with lowercase nucleotide letters.

## Chapter 6

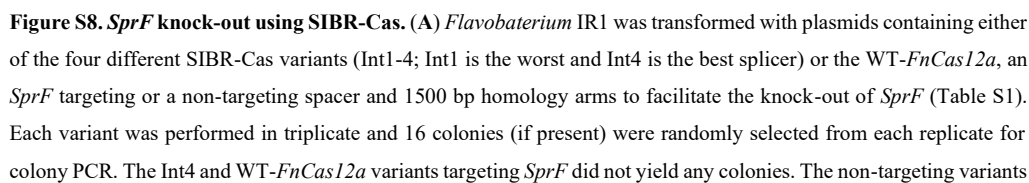


## Chapter 6

**Figure S6. *FlgM* knock-out using SIBR-Cas.** (A) *P. putida* KT2440 was transformed with plasmids containing either of the four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer) or the WT-*FnCas12a*, an *FlgM* targeting or a non-targeting spacer and 500 bp homology arms to facilitate the knock-out of *FlgM* (Table S1). Each variant was performed in triplicate and 16 colonies (if present) were randomly selected from each replicate for colony PCR. The WT-*FnCas12a* variant targeting *FlgM* did not yield any colonies. The non-targeting variants were all WT and for this reason only Int4 is shown as a representative. Mix amplicons (WT and knock-out bands) were not counted for the total knock-out efficiency percentage. WT: 1523 bp, Knock-out (KO): 1208 bp. (B) Schematic representation of WT or edited (KO) *P. putida* KT2440 genome at the *FlgM* locus. BG21663 and BG21664 represent the primers used for PCR (Table S2). (C) Schematic representation of the pSIBR plasmids used to knock-out *FlgM*. Int1-4 represent the four different SIBR-Cas variants and their sequence difference is depicted with lowercase nucleotide letters.



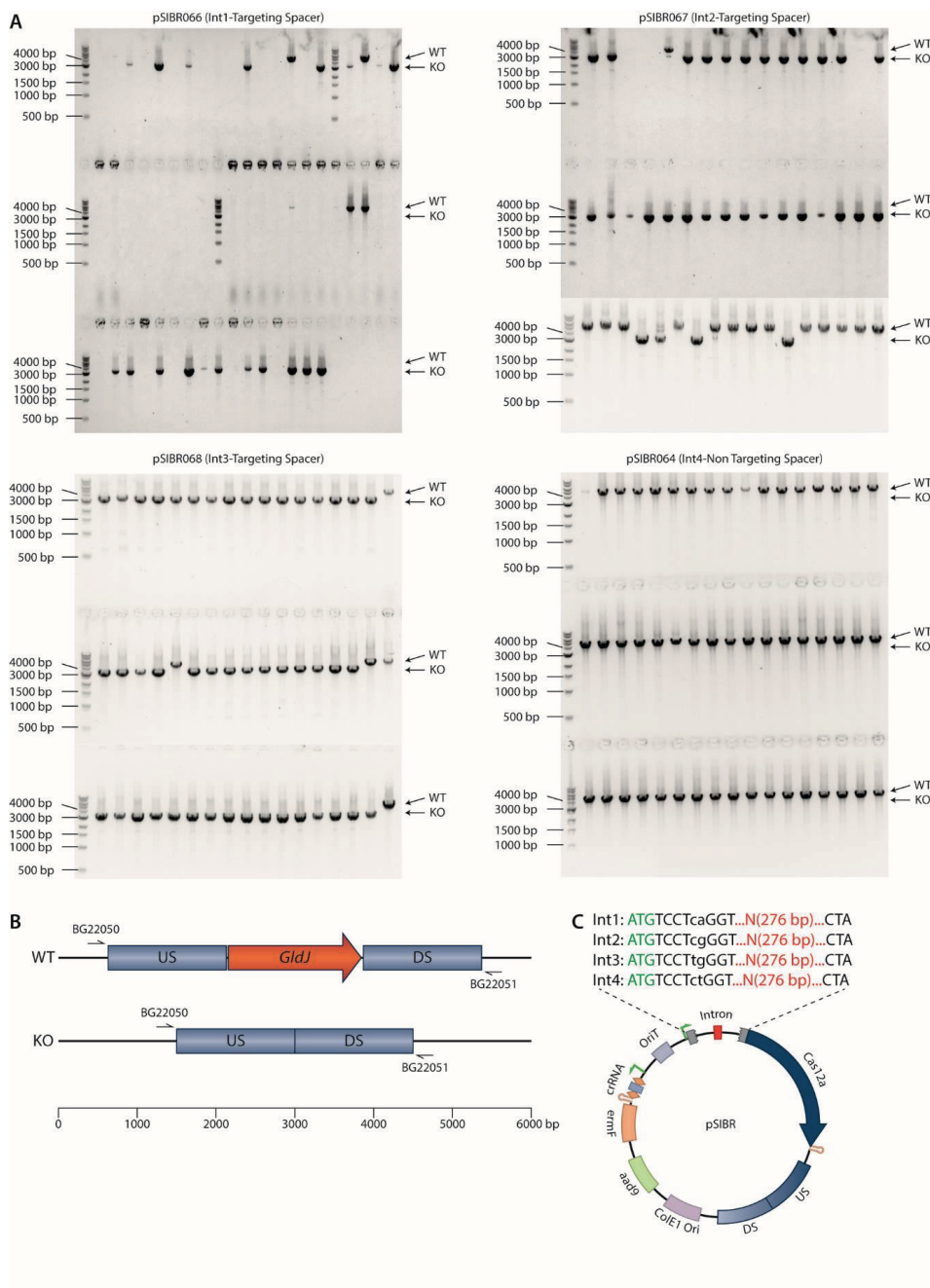
**Figure S7. *EndA* knock-out using non-targeting SIBR-Cas.** (A) *P. putida* KT2440 was transformed with plasmids containing either of the four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer) or the WT-*FnCas12a*, an *EndA* non-targeting spacer and 500 bp homology arms to facilitate the knock-out of *EndA* (Table S1). Each variant was performed in triplicate and 16 colonies (if present) were randomly selected from each replicate for colony PCR. All variants were WT and for this reason only the Int4 is shown as a representative. WT: 1814 bp, Knock-out (KO): 1121 bp. (B) Schematic representation of WT or edited (KO) *P. putida* KT2440 genome at the *EndA* locus. BG21027 and BG21028 represent the primers used for PCR (Table S2).



### Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas

were all WT and for this reason only Int4 is shown as a representative. Mix amplicons (WT and knock-out bands) were not counted for the total knock-out efficiency percentage. WT: 4081 bp, Knock-out (KO): 3082 bp. **(B)** Schematic representation of WT or edited (KO) *Flavobacterium* IR1 genome at the *SprF* locus. BG18323 and BG18324 represent the primers used for PCR (Table S2). **(C)** Schematic representation of the pSIBR plasmids used to knock-out *SprF*. Int1-4 represent the four different SIBR-Cas variants and their sequence difference is depicted with lowercase nucleotide letters.

Chapter 6

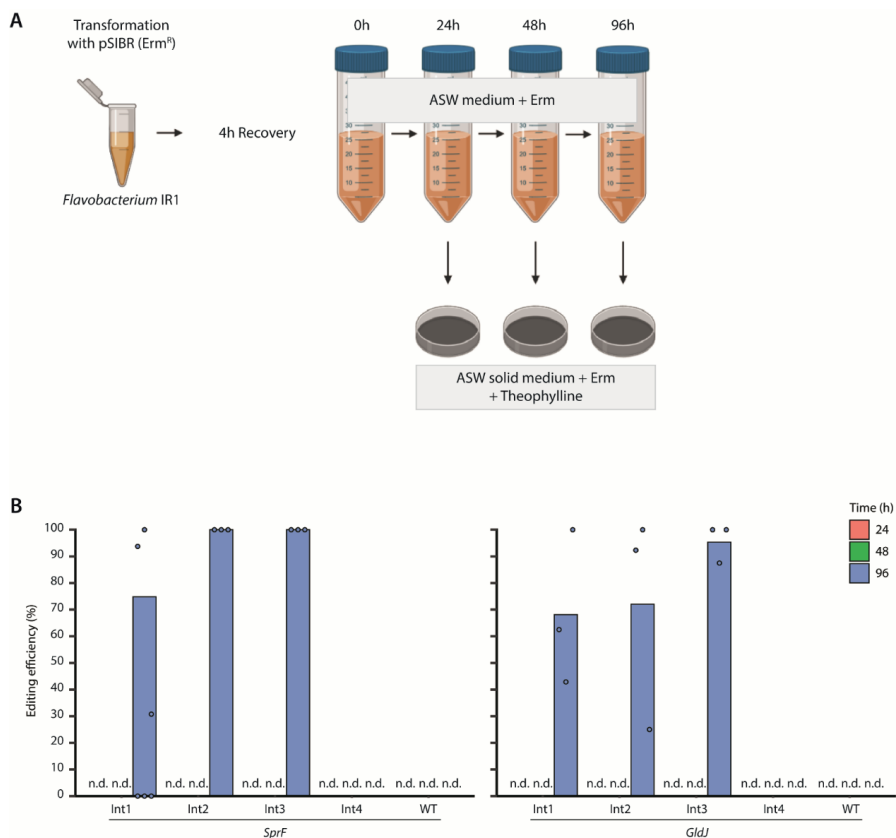


**Figure S9. *GldJ* knock-out using SIBR-Cas.** (A) *Flavobacterium* IR1 was transformed with plasmids containing either of the four different SIBR-Cas variants (Int1-4; Int1 is the worst and Int4 is the best splicer) or the WT-*FnCas12a*, a *GldJ* targeting or a non-targeting spacer and 1500 bp homology arms to facilitate the knock-out of *GldJ* (Table S1). Each variant was performed in triplicate and 16 colonies (if present) were randomly selected from each replicate for colony PCR. The Int4 and WT-*FnCas12a* variants targeting *GldJ* did not yield any colonies. The non-targeting variants

### Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas

were all WT and for this reason only the Int4 is shown as a representative. Mix amplicons (WT and knock-out bands) were not counted for the total knock-out efficiency percentage. WT: 4944 bp, Knock-out (KO): 3255 bp. **(B)** Schematic representation of WT or edited (KO) *Flavobacterium* IR1 genome at the *GldJ* locus. BG22050 and BG22051 represent the primers used for PCR (Table S2). **(C)** Schematic representation of the pSIBR plasmids used to knock-out *GldJ*. Int1-4 represent the four different SIBR-Cas variants and their sequence difference is depicted with lowercase nucleotide letters.





**Figure S10. *SprF* and *GldJ* knock-out in *Flavobacterium* IR1 using SIBR-Cas and following a 96 h incubation approach.** (A) Schematic representation of the 96 h incubation approach. After transformation of *Flavobacterium* IR1 with plasmids containing either of the four different SIBR-Cas variants (Int1-4; Int1 is the worst splicer) or the WT-*FnCas12a*, an *SprF* or a *GldJ* targeting spacer and 1500 bp homology arms to facilitate either the knock-out of *SprF* or *GldJ* (Table S1), cells were recovered for 4 hours at 30°C in 1 mL ASW medium. After recovery, the cells were transferred in 10 mL ASW medium containing erythromycin (Erm; 200 mg L<sup>-1</sup>) and incubated for 96 hours. 1 mL of culture was recovered at 24, 48 and 96 hours of incubation, centrifuged to remove excess medium and plated on ASW agar containing erythromycin (200 mg L<sup>-1</sup>) and the theophylline inducer (2 mM). This image was created using BioRender.com. (B) Editing efficiency of *SprF* and *GldJ* genes in *Flavobacterium* IR1 following the 96 h incubation approach as indicated in (A). Colonies were obtained only at 96 hours (blue bars) after transformation and the editing efficiency of the biological triplicates is represented with a blue dot. 24 hours (red bars) and 48 hours (green bars) of incubation did not yield any colonies. Int1 to Int4 represent the four Intron-*FnCas12a* variants and WT represents the WT *FnCas12a* gene. N.d., not determined.

Table S1. Plasmids used in this study.

Plasmid name	Description and relevant characteristics	Reference	Plasmid map (Benchling link)
<b>LacZ assays in <i>E. coli</i></b>			
pEA001 [-7W]	KanR, lacUV5p-[-7W] Intron-LacZ	This study	<a href="https://benchling.com/s/seq-b29hHw7mAwS9wCC9XjSh">https://benchling.com/s/seq-b29hHw7mAwS9wCC9XjSh</a>
pEA001 [-7M]	KanR, lacUV5p-[-7M] Intron-LacZ	This study	<a href="https://benchling.com/s/seq-b29hHw7mAwS9wCC9XjSh">https://benchling.com/s/seq-b29hHw7mAwS9wCC9XjSh</a>
pEA001 [PPP]	KanR, lacUV5p-PPP Intron-LacZ	This study	<a href="https://benchling.com/s/seq-TaewzYsGuJPgQQIUMGWK">https://benchling.com/s/seq-TaewzYsGuJPgQQIUMGWK</a>
pEA001 [PPW]	KanR, lacUV5p-PPW Intron-LacZ	This study	<a href="https://benchling.com/s/seq-EOKis573cMPYC19kMe1K">https://benchling.com/s/seq-EOKis573cMPYC19kMe1K</a>
pEA001 [PPM]	KanR, lacUV5p-PPM Intron-LacZ	This study	<a href="https://benchling.com/s/seq-WoEO8dYbzWWEfgP4YHZ3">https://benchling.com/s/seq-WoEO8dYbzWWEfgP4YHZ3</a>
pEA001 [PWP]	KanR, lacUV5p-PWP Intron-LacZ	This study	<a href="https://benchling.com/s/seq-v5xeWzkey7xqyDVz049D">https://benchling.com/s/seq-v5xeWzkey7xqyDVz049D</a>
pEA001 [PWW]	KanR, lacUV5p-PWW Intron-LacZ	This study	<a href="https://benchling.com/s/seq-fyuHmtagQ1frgrXSHLTI">https://benchling.com/s/seq-fyuHmtagQ1frgrXSHLTI</a>
pEA001 [PWM]	KanR, lacUV5p-PWM Intron-LacZ	This study	<a href="https://benchling.com/s/seq-FCyiOnKpcH9rjLCKPBuQ">https://benchling.com/s/seq-FCyiOnKpcH9rjLCKPBuQ</a>
pEA001 [PMP]	KanR, lacUV5p-PMP Intron-LacZ	This study	<a href="https://benchling.com/s/seq-ITxfhMjjvaKJqg160zNT">https://benchling.com/s/seq-ITxfhMjjvaKJqg160zNT</a>
pEA001 [PMW]	KanR, lacUV5p-PMW Intron-LacZ	This study	<a href="https://benchling.com/s/seq-JPwy5FGBvEskE2GVbv7P">https://benchling.com/s/seq-JPwy5FGBvEskE2GVbv7P</a>
pEA001 [PMM]	KanR, lacUV5p-PMM Intron-LacZ	This study	<a href="https://benchling.com/s/seq-x3TLolbNngfV2YQilNRY">https://benchling.com/s/seq-x3TLolbNngfV2YQilNRY</a>
pEA001 [MPP]	KanR, lacUV5p-MPP Intron-LacZ	This study	<a href="https://benchling.com/s/seq-0Xylkf8SqsFB5t8M1R17">https://benchling.com/s/seq-0Xylkf8SqsFB5t8M1R17</a>
pEA001 [MPW]	KanR, lacUV5p-MPW Intron-LacZ	This study	<a href="https://benchling.com/s/seq-6WHDIEHCNkgJwg9QMgko">https://benchling.com/s/seq-6WHDIEHCNkgJwg9QMgko</a>
pEA001 [MPM]	KanR, lacUV5p-MPM Intron-LacZ	This study	<a href="https://benchling.com/s/seq-8Pv9l5ldfQDM4uE9mEsk">https://benchling.com/s/seq-8Pv9l5ldfQDM4uE9mEsk</a>
pEA001 [MPW]	KanR, lacUV5p-MPW Intron-LacZ	This study	<a href="https://benchling.com/s/seq-6WHDIEHCNkgJwg9QMgko">https://benchling.com/s/seq-6WHDIEHCNkgJwg9QMgko</a>
pEA001 [MWW]	KanR, lacUV5p-MWW Intron-LacZ	This study	<a href="https://benchling.com/s/seq-o5RmlTaeke7KfC26ucN2">https://benchling.com/s/seq-o5RmlTaeke7KfC26ucN2</a>
pEA001 [MWM]	KanR, lacUV5p-MWM Intron-LacZ	This study	<a href="https://benchling.com/s/seq-LXTAIKbOkh0LkUtwS6jN">https://benchling.com/s/seq-LXTAIKbOkh0LkUtwS6jN</a>
pEA001 [MMP]	KanR, lacUV5p-MMP Intron-LacZ	This study	<a href="https://benchling.com/s/seq-5tcNZAVJGT4y9edfAQ45">https://benchling.com/s/seq-5tcNZAVJGT4y9edfAQ45</a>
pEA001 [MMW]	KanR, lacUV5p-MMW Intron-LacZ	This study	<a href="https://benchling.com/s/seq-ZA7JveYYuPSFZYHL6R14">https://benchling.com/s/seq-ZA7JveYYuPSFZYHL6R14</a>
pEA001 [MMM]	KanR, lacUV5p-MMM Intron-LacZ	This study	<a href="https://benchling.com/s/seq-eTGayMzfAzP8rWzB6bGF">https://benchling.com/s/seq-eTGayMzfAzP8rWzB6bGF</a>
pEA001 [296P]	KanR, lacUV5p-[+296P] Intron-LacZ	This study	<a href="https://benchling.com/s/seq-r1rCLMW5YYzhYzQMnFu0">https://benchling.com/s/seq-r1rCLMW5YYzhYzQMnFu0</a>
pEA001 [296W]	KanR, lacUV5p-[+296W] Intron-LacZ	This study	<a href="https://benchling.com/s/seq-7WTrwg0RE4wBvLycHil7">https://benchling.com/s/seq-7WTrwg0RE4wBvLycHil7</a>
<b>SIBR-Cas targeting and editing in <i>E. coli</i> MG1655</b>			
pSIBR001	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t	This study	<a href="https://benchling.com/s/seq-6nB5lmpoZV4I4MGsFa08">https://benchling.com/s/seq-6nB5lmpoZV4I4MGsFa08</a>
pSIBR002	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t	This study	<a href="https://benchling.com/s/seq-HqcmCgFky18So0ILu4H">https://benchling.com/s/seq-HqcmCgFky18So0ILu4H</a>
pSIBR003	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t	This study	<a href="https://benchling.com/s/seq-ITK2KqCHUltu7RRdTM7z">https://benchling.com/s/seq-ITK2KqCHUltu7RRdTM7z</a>

## Chapter 6

pSIBR004	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t	This study	<a href="https://benchling.com/s/seq-ynlTp7uQfMXeA3ozptCF">https://benchling.com/s/seq-ynlTp7uQfMXeA3ozptCF</a>
pSIBR005	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t	This study	<a href="https://benchling.com/s/seq-oMUyGhlUeozuKBzAw4jc">https://benchling.com/s/seq-oMUyGhlUeozuKBzAw4jc</a>
pSIBR006	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-KfOvLIHpPQ3Nc3Letvi2">https://benchling.com/s/seq-KfOvLIHpPQ3Nc3Letvi2</a>
pSIBR007	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-A9yxFE6tM147TxLnuBp5">https://benchling.com/s/seq-A9yxFE6tM147TxLnuBp5</a>
pSIBR008	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-GsV7JDjZCptXs74ru8KP">https://benchling.com/s/seq-GsV7JDjZCptXs74ru8KP</a>
pSIBR009	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-bOnvk6U3KpdodrtOdwqa">https://benchling.com/s/seq-bOnvk6U3KpdodrtOdwqa</a>
pSIBR010	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-UfbGPfRlByNQMPfIRQgf">https://benchling.com/s/seq-UfbGPfRlByNQMPfIRQgf</a>
pSIBR011	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t	This study	<a href="https://benchling.com/s/seq-uiQtuVeqvz5zjF7z0ny3">https://benchling.com/s/seq-uiQtuVeqvz5zjF7z0ny3</a>
pSIBR012	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t	This study	<a href="https://benchling.com/s/seq-JQKbacCnlyWpi8euHvrn">https://benchling.com/s/seq-JQKbacCnlyWpi8euHvrn</a>
pSIBR013	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t	This study	<a href="https://benchling.com/s/seq-yIKKtgjKF6DIFbcDwHCQ">https://benchling.com/s/seq-yIKKtgjKF6DIFbcDwHCQ</a>
pSIBR014	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t	This study	<a href="https://benchling.com/s/seq-fkln1V3COjxwQkzR99Tn">https://benchling.com/s/seq-fkln1V3COjxwQkzR99Tn</a>
pSIBR015	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t	This study	<a href="https://benchling.com/s/seq-uT1YvCJDtZK0T3ZV9TSZ">https://benchling.com/s/seq-uT1YvCJDtZK0T3ZV9TSZ</a>
pSIBR016	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-2zEd9FIQA0IliAMMtng9">https://benchling.com/s/seq-2zEd9FIQA0IliAMMtng9</a>
pSIBR017	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-7KxCoxBOjFf1l5JKPY0V">https://benchling.com/s/seq-7KxCoxBOjFf1l5JKPY0V</a>
pSIBR018	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-uyyUS1G1D8LGq0CXwdIO">https://benchling.com/s/seq-uyyUS1G1D8LGq0CXwdIO</a>
pSIBR019	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-fAPkx41fWWhXfMxP33Bp">https://benchling.com/s/seq-fAPkx41fWWhXfMxP33Bp</a>
pSIBR020	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-LacZ spacer-B1002t, 500 bp homologous arms for <i>LacZ</i>	This study	<a href="https://benchling.com/s/seq-HL3kNsN2thXzkaSbBCuY">https://benchling.com/s/seq-HL3kNsN2thXzkaSbBCuY</a>

SIBR-Cas targeting and editing in <i>P. putida</i> KT2440			
pSIBR021	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-SxuPs17CAIx8DACfcRHC">https://benchling.com/s/seq-SxuPs17CAIx8DACfcRHC</a>
pSIBR022	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-3q2GgUwX70pVGd86YoZz">https://benchling.com/s/seq-3q2GgUwX70pVGd86YoZz</a>
pSIBR023	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-BWRXhXnYy2XmfKUuFNnW">https://benchling.com/s/seq-BWRXhXnYy2XmfKUuFNnW</a>
pSIBR024	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-XZnDqXp8dQWu5CSMU41k">https://benchling.com/s/seq-XZnDqXp8dQWu5CSMU41k</a>
pSIBR025	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-NT spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-IYm1KhIuepDFo3EWjPIR">https://benchling.com/s/seq-IYm1KhIuepDFo3EWjPIR</a>
pSIBR026	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t	This study	<a href="https://benchling.com/s/seq-dgVIJzUOcRjip42yGgb1">https://benchling.com/s/seq-dgVIJzUOcRjip42yGgb1</a>
pSIBR027	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t	This study	<a href="https://benchling.com/s/seq-WQU6RFH5WcTmIlG3b5hF">https://benchling.com/s/seq-WQU6RFH5WcTmIlG3b5hF</a>
pSIBR028	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t	This study	<a href="https://benchling.com/s/seq-ESDD9SEYInV9PEXrds52">https://benchling.com/s/seq-ESDD9SEYInV9PEXrds52</a>
pSIBR029	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t	This study	<a href="https://benchling.com/s/seq-NaPP64FzauzJVp63HOJz">https://benchling.com/s/seq-NaPP64FzauzJVp63HOJz</a>
pSIBR030	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t	This study	<a href="https://benchling.com/s/seq-u2ZsyKxbD4tPqMorvQnJ">https://benchling.com/s/seq-u2ZsyKxbD4tPqMorvQnJ</a>
pSIBR031	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-bx2nsbDh3fepPVsTN2iY">https://benchling.com/s/seq-bx2nsbDh3fepPVsTN2iY</a>
pSIBR032	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-CdpX42lFfRmfEdZJ9oUG">https://benchling.com/s/seq-CdpX42lFfRmfEdZJ9oUG</a>
pSIBR033	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-9R6r5e6hNU6dBaj3EsiT">https://benchling.com/s/seq-9R6r5e6hNU6dBaj3EsiT</a>
pSIBR034	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-vfPXGovDiFmGAsHEBT4n">https://benchling.com/s/seq-vfPXGovDiFmGAsHEBT4n</a>
pSIBR035	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-EndA spacer-B1002t, 500 bp homologous arms for EndA	This study	<a href="https://benchling.com/s/seq-YRxyMOOAVIX5leF9nxBX">https://benchling.com/s/seq-YRxyMOOAVIX5leF9nxBX</a>
pSIBR036	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-NT-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-gDxNAZRAqfBfgvEKvtlS">https://benchling.com/s/seq-gDxNAZRAqfBfgvEKvtlS</a>

pSIBR037	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-NT-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-x4ZSBS8StDdzT1yEifmA">https://benchling.com/s/seq-x4ZSBS8StDdzT1yEifmA</a>
pSIBR038	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-NT-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-JsWtfIN7RWaeK1EqyCv9">https://benchling.com/s/seq-JsWtfIN7RWaeK1EqyCv9</a>
pSIBR039	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-NT-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-xelvILM9aFbrFN5JGAGs">https://benchling.com/s/seq-xelvILM9aFbrFN5JGAGs</a>
pSIBR040	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-NT-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-6knJB0dffuipQURdFqxqW">https://benchling.com/s/seq-6knJB0dffuipQURdFqxqW</a>
pSIBR041	KanR, lacUV5p-Int1-FnCas12a-B1002t, lacUV5p-FlgM spacer-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-Wb4iXW1NIk2iKe2vPhTL">https://benchling.com/s/seq-Wb4iXW1NIk2iKe2vPhTL</a>
pSIBR042	KanR, lacUV5p-Int2-FnCas12a-B1002t, lacUV5p-FlgM spacer-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-yKGbB2QsknL1JcFwahSF">https://benchling.com/s/seq-yKGbB2QsknL1JcFwahSF</a>
pSIBR043	KanR, lacUV5p-Int3-FnCas12a-B1002t, lacUV5p-FlgM spacer-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-50Urz54c5xTwjw316Nes">https://benchling.com/s/seq-50Urz54c5xTwjw316Nes</a>
pSIBR044	KanR, lacUV5p-Int4-FnCas12a-B1002t, lacUV5p-FlgM spacer-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-qByXhASTnuawws86f6Ou">https://benchling.com/s/seq-qByXhASTnuawws86f6Ou</a>
pSIBR045	KanR, lacUV5p-WT-FnCas12a-B1002t, lacUV5p-FlgM spacer-B1002t, 500 bp homologous arms for FlgM	This study	<a href="https://benchling.com/s/seq-P8mK0IPbLhXKnuyQsToC">https://benchling.com/s/seq-P8mK0IPbLhXKnuyQsToC</a>
<b>SIBR-Cas editing in <i>Flavobacterium</i> IR1</b>			
pSpyCas9Fb_NT	AmpR ( <i>E. coli</i> ), ErmR ( <i>Flavobacterium</i> ), ompAp-SpyCas9, Hup-ransom spacer sgRNA-ompAt	Carrión <i>et al.</i> (2019)	<a href="https://benchling.com/s/seq-iBmrky1QUKQRxSTF4alx">https://benchling.com/s/seq-iBmrky1QUKQRxSTF4alx</a>
pSIBR046	SpecR ( <i>E. coli</i> ), ErmR ( <i>Flavobacterium</i> ), ompAp-Int1-FnCas12a-mapt, Hup-NT spacer-ompAt	This study	<a href="https://benchling.com/s/seq-tryMFuIMBcQowzpUhUox">https://benchling.com/s/seq-tryMFuIMBcQowzpUhUox</a>
pSIBR047	SpecR ( <i>E. coli</i> ), ErmR ( <i>Flavobacterium</i> ), ompAp-Int2-FnCas12a-mapt, Hup-NT spacer-ompAt	This study	<a href="https://benchling.com/s/seq-fuYCpx2KmNQbSW3ruCvt">https://benchling.com/s/seq-fuYCpx2KmNQbSW3ruCvt</a>
pSIBR048	SpecR ( <i>E. coli</i> ), ErmR ( <i>Flavobacterium</i> ), ompAp-Int3-FnCas12a-mapt, Hup-NT spacer-ompAt	This study	<a href="https://benchling.com/s/seq-34dQ4N74oKkn6jYkSxoh">https://benchling.com/s/seq-34dQ4N74oKkn6jYkSxoh</a>
pSIBR049	SpecR ( <i>E. coli</i> ), ErmR ( <i>Flavobacterium</i> ), ompAp-Int4-FnCas12a-mapt, Hup-NT spacer-ompAt	This study	<a href="https://benchling.com/s/seq-SnpUmWz3Tb6iEiB1sKcZ">https://benchling.com/s/seq-SnpUmWz3Tb6iEiB1sKcZ</a>
pSIBR050	SpecR ( <i>E. coli</i> ), ErmR ( <i>Flavobacterium</i> ), ompAp-WT-FnCas12a-mapt, Hup-NT spacer-ompAt	This study	<a href="https://benchling.com/s/seq-w5aqQtyTT8YnsNFvnLBf">https://benchling.com/s/seq-w5aqQtyTT8YnsNFvnLBf</a>

pSIBR051	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int1 FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-dZilphxSOvgSfNPC0xVI">https://benchling.com/s/seq-dZilphxSOvgSfNPC0xVI</a>
pSIBR052	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int2 FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-IWXgaNozjrxheoaH3GsO">https://benchling.com/s/seq-IWXgaNozjrxheoaH3GsO</a>
pSIBR053	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int3 FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-iOJXCZaLpoLePAGEWyRj">https://benchling.com/s/seq-iOJXCZaLpoLePAGEWyRj</a>
pSIBR054	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int4 FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-e67DTZnqKUWgHReK5w9i">https://benchling.com/s/seq-e67DTZnqKUWgHReK5w9i</a>
pSIBR055	SpecR (E. coli), ErmR (Flavobacterium), ompAp-WT FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-Ef4po4YgCN3MoTmjNk5Z">https://benchling.com/s/seq-Ef4po4YgCN3MoTmjNk5Z</a>
pSIBR056	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int1 FnCas12a-mapt, Hup-SprF spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-WkoYYG87CQOlaLzXNI4o">https://benchling.com/s/seq-WkoYYG87CQOlaLzXNI4o</a>
pSIBR057	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int2 FnCas12a-mapt, Hup-SprF spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-Fu0ho5T6eUO2GZOmiu0Z">https://benchling.com/s/seq-Fu0ho5T6eUO2GZOmiu0Z</a>
pSIBR058	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int3 FnCas12a-mapt, Hup-SprF spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-RljMtyXgvb1x2BflEyaX">https://benchling.com/s/seq-RljMtyXgvb1x2BflEyaX</a>
pSIBR059	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int4 FnCas12a-mapt, Hup-SprF spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-6Ma0owMHNGMER9Goxp39">https://benchling.com/s/seq-6Ma0owMHNGMER9Goxp39</a>
pSIBR060	SpecR (E. coli), ErmR (Flavobacterium), ompAp-WT FnCas12a-mapt, Hup-SprF spacer-ompAt, 1500 bp homologous arms for SprF	This study	<a href="https://benchling.com/s/seq-xCL5zkoGDjmQql4zMpHm">https://benchling.com/s/seq-xCL5zkoGDjmQql4zMpHm</a>
pSIBR061	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int1 FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-chB8po62TOoKZWxUDUba">https://benchling.com/s/seq-chB8po62TOoKZWxUDUba</a>
pSIBR062	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int2 FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-3YTUB840ckExcZGvba39">https://benchling.com/s/seq-3YTUB840ckExcZGvba39</a>
pSIBR063	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int3	This study	<a href="https://benchling.com/s/seq-xchTm7PHiocBjwZ9SFMp">https://benchling.com/s/seq-xchTm7PHiocBjwZ9SFMp</a>

## Chapter 6

	FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for GldJ		
pSIBR064	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int4 FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-mBvNACYHsDYyDCxoRYiH">https://benchling.com/s/seq-mBvNACYHsDYyDCxoRYiH</a>
pSIBR065	SpecR (E. coli), ErmR (Flavobacterium), ompAp-WT FnCas12a-mapt, Hup-NT spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-xUQNzF0bQqrysCOUPDRE">https://benchling.com/s/seq-xUQNzF0bQqrysCOUPDRE</a>
pSIBR066	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int1 FnCas12a-mapt, Hup-GldJ spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-NzQNSVzLjYTyhEV2ZyR2">https://benchling.com/s/seq-NzQNSVzLjYTyhEV2ZyR2</a>
pSIBR067	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int2 FnCas12a-mapt, Hup-GldJ spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-U85vC5BwueDjpfqOL2oE">https://benchling.com/s/seq-U85vC5BwueDjpfqOL2oE</a>
pSIBR068	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int3 FnCas12a-mapt, Hup-GldJ spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-HNAiHbFECKxxe3t9PdWk">https://benchling.com/s/seq-HNAiHbFECKxxe3t9PdWk</a>
pSIBR069	SpecR (E. coli), ErmR (Flavobacterium), ompAp-Int4 FnCas12a-mapt, Hup-GldJ spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-WZi2dhVTrALGUiDMmljs">https://benchling.com/s/seq-WZi2dhVTrALGUiDMmljs</a>
pSIBR070	SpecR (E. coli), ErmR (Flavobacterium), ompAp-WT FnCas12a-mapt, Hup-GldJ spacer-ompAt, 1500 bp homologous arms for GldJ	This study	<a href="https://benchling.com/s/seq-TQHsQtPuiTAqZp0P2XwE">https://benchling.com/s/seq-TQHsQtPuiTAqZp0P2XwE</a>

**Table S2. Oligonucleotides used in this study. The spacer moieties of the crRNA used in this study are highlighted with red.**

Oligo ID	Sequence (5' to 3')	Description
BG22591	CGGCGAGGATGAGTGCACAG	cPCR for LacZ KO in <i>E. coli</i> , forward
BG22592	GGGAAGGCGACTGGAGTGCC	cPCR for LacZ KO in <i>E. coli</i> , reverse
BG21027	CGAAGTGATGGCCAAGCTGGG	cPCR for EndA KO in <i>P. putida</i> , forward
BG21028	CTGGCGATGGTAGCGATGACC	cPCR for EndA KO in <i>P. putida</i> , reverse
BG21663	CGTGGACGGTATTCGTGCCG	cPCR for FlgM KO in <i>P. putida</i> , forward
BG21664	GGTTGCGGGGCATCGGATTC	cPCR for FlgM KO in <i>P. putida</i> , reverse
BG18323	AATAGACGCTTTAGAGCTAC	cPCR for SprF KO in <i>Flavobacterium IR1</i> , forward
BG18324	CTTAGGGCAATAATTAGTGC	cPCR for SprF KO in <i>Flavobacterium IR1</i> , reverse
BG22050	ACAATTCCTGTGTTCGAGGC	cPCR for GldJ KO in <i>Flavobacterium IR1</i> , forward
BG22051	CACAGACAAAAGCTGGAAGG	cPCR for GldJ KO in <i>Flavobacterium IR1</i> , reverse
BG20154	AGAAGACATAGATCAACGTCGTGACTGGGAAAAGTCTATG TCTCA	LacZ spacer insertion through Golden Gate, forward
BG20155	TGAAGACATAGACTTTTCCAGTCACGACGTTGATCTATGT CTTCT	LacZ spacer insertion through Golden Gate, reverse
BG20271	AGAAGACATAGATGGCTGGTACCAGAACACGGTCTATG TCTCA	EndA spacer insertion through Golden Gate, forward
BG20272	TGAAGACATAGACC GTTGTCTGGTAGCCAGCCATCTATGT CTTCT	EndA spacer insertion through Golden Gate, reverse
BG21614	AGAAGACATAGATATTTCGAAGCCCAGCGCTGAGTCTATG TCTCA	FlgM spacer insertion through Golden Gate, forward
BG21615	TGAAGACATAGACTCAGCGCTGGGCTTCGAAATATCTATGT CTTCT	FlgM spacer insertion through Golden Gate, reverse
BG17871	AGGTCTCATAGATGATATTCTTACCAGGTTATGGTCTAAGA GACCA	SprF spacer insertion through Golden Gate, forward
BG17872	TGGTCTCTTAGACCATAACCTGGTAAGAATATCATCTATGA GACCT	SprF spacer insertion through Golden Gate, reverse
BG22036	AGGTCTCATAGATCCCATAGTAAACGTACCTCCGTCTAAGA GACCA	GldJ spacer insertion through Golden Gate, forward
BG22037	TGGTCTCTTAGACGGAGGTACGTTTACTATGGGATCTATGA GACCT	GldJ spacer insertion through Golden Gate, reverse
BG20140	AGCTGTTTCCTGTGTGAAAT	Upstream homology arm for LacZ, forward



## Chapter 6

BG20644	TAACGCTGCCGCGCCGGTAAGGCATCGTTCCCACTGCGAT	Upstream homology arm for LacZ, reverse
BG20643	TTAATTGGACCGCGGTCCGACCAACACAGCCAAACATCCG	Downstream homology arm for LacZ, forward
BG20139	ATTTCACACAGGAAACAGCTTAATAACCGGGCAGGCCATG	Downstream homology arm for LacZ, reverse
BG20266	CGTCTCAGCGCAGTCAATCTTCCTTCG	Upstream homology arm for EndA, forward
BG20267	CGTCTCAGTAGAGCAAAGAGCTGCAGCGGAT	Upstream homology arm for EndA, reverse
BG20264	CGTCTCATCCGAACCAGTAAAGTGCGGCCG	Downstream homology arm for EndA, forward
BG20265	CGTCTCAGCGCTCCTCAGGCCAGCGTTTGTA	Downstream homology arm for EndA, reverse
BG21610	CGTCTCATCCGACTTTTCCGCGACGTGGTG	Upstream homology arm for FlgM, forward
BG21611	CGTCTCAACGGGATCAGAAACCTCTGGGTATTTGG	Upstream homology arm for FlgM, reverse
BG21612	CGTCTCACCGTACGGCGCGCTGACTTC	Downstream homology arm for FlgM, forward
BG21613	CGTCTCAGTAGCGACACACGCAAGTAACGGC	Downstream homology arm for FlgM, reverse
BG17933	CCTCGAGATCTCCATGGACGCAACTAGACGTTACCAATGC	Upstream homology arm for SprF, forward
BG17934	CTTAAATGATCCTATTTTTCTGTTGGGGCAATCAATTGTTATC	Upstream homology arm for SprF, reverse
BG17935	ACGAAAAATAGGATCATTTAAG	Downstream homology arm for SprF, forward
BG17936	CCTCTAGAGTCGACGTCACGGTAATTTAGTCCAAAATGGC	Downstream homology arm for SprF, reverse
BG21734	AGTTGGGGCCTCGAGATCTCCATGGACGGAAGTTAAATTGCTTCCCGG	Upstream homology arm for GldJ, forward
BG21735	TGTTACAATTAATATATTGACTCATTCTTAGGTGATAAATTAGG	Upstream homology arm for GldJ, reverse
BG21736	CAATATATTTAATTGTAACAAAAGCCC	Downstream homology arm for GldJ, forward
BG21737	CCGGGGATCCTCTAGAGTCGACGTCACGCCTCGTGAAGTGGATTATC	Downstream homology arm for GldJ, reverse
BG5039	GATCTTAAGGATGTTTTGTTGGGTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [-7M]
BG5040	TGACTGCAGAATATTAACGGTAGCATTATGTTTCAGATAAGGTCTG	Reverse primer to construct WT 3' exonic flanking region of pEA001 series
BG5206	GATCTTAAGGATGTTTTCTCAGGTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PPP]
BG5207	GATCTTAAGGATGTTTTCTCGGGTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PPW]
BG5208	GATCTTAAGGATGTTTTCTCTGGTTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PPM]
BG5209	GATCTTAAGGATGTTTTCTTAGGTTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PWP]
BG5210	GATCTTAAGGATGTTTTCTTGGGTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PWW]
BG5211	GATCTTAAGGATGTTTTCTTTGGTTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PWM]
BG5212	GATCTTAAGGATGTTTTCTGAGGTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PMP]
BG5213	GATCTTAAGGATGTTTTCTGGGGTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PMW]
BG5214	GATCTTAAGGATGTTTTCTGTGGTTAAATTGAGGCCTGAGTATAAGGTG	Forward primer to construct pEA001 [PMM]

BG5215	GATCTTAAGGATGTTTTCCAGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MPP]
BG5216	GATCTTAAGGATGTTTTCCCGGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MPW]
BG5217	GATCTTAAGGATGTTTTCCCTGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MPM]
BG5218	GATCTTAAGGATGTTTTCTAGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MPW]
BG5219	GATCTTAAGGATGTTTTCTGCGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MWW]
BG5220	GATCTTAAGGATGTTTTCCCTGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MWM]
BG5221	GATCTTAAGGATGTTTTCCGAGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MMP]
BG5222	GATCTTAAGGATGTTTTCCCGGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MMW]
BG5223	GATCTTAAGGATGTTTTCCGTGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [MMM]
BG5304	GATCTTAAGGATGTTCTTTGGGTTAATTGAGGCCTGAGTA TAAGGTG	Forward primer to construct pEA001 [-7W]
BG5305	TGACTGCAGAATATTAACGGGAGCATTATGTTTCAGATAA GGTCG	Reverse primer to construct pEA001 [296P]
BG5306	TGACTGCAGAATATTAACGGAAGCATTATGTTTCAGATAA GGTCG	Reverse primer to construct pEA001 [296W]

[illegible]

## **Chapter 7**

### **Summary and general discussion**

## Thesis summary

Microbiology is an extraordinary field which has been, and will continue to be, central to the advancement of human cultures. In fact, human cultures and even our planet would not be as they are today if microorganisms and microbiology would not exist. Basic, but at the same time, complex microbiological processes play key roles in biodegradation and nutrient (re)cycling, climate change, disease and biotechnology. Due to its importance, a brief history about microbiology is provided in **Chapter 1** and the related (old or new) biotechnological applications are described. The importance of the microorganism-workhorses is highlighted and a “start-at-the-end” approach is recommended for choosing the appropriate microbe for the production of the final product. A brief introduction to clostridia species and their potential for the sustainable production of ester compounds is provided. Lastly, the importance of the revolutionary CRISPR-Cas genome engineering tools is highlighted.

**Chapter 2** reviews the state-of-the-art technologies used for the microbial production of short and medium chain esters. Several enzymatic processes are described with main focus on the AAT enzymes. In addition, several metabolic engineering processes are reviewed, along with the physical properties of the final ester product. Lastly, an outlook is provided on other products (e.g. alcohols, carboxylic acids and diols) that can be derived from (microbially produced) esters.

In **Chapter 3**, the unique Eat1 enzyme was further characterized using *in vitro* and *in vivo* approaches. Two new activities (alcoholysis and thiolysis) were added to the existing activities (AAT, esterase and thioesterase) of Eat1. Alcoholysis appears to be the main activity of the enzyme as it highly prefers to exchange the alcohol moiety of the ester molecule with another free alcohol. This discovery exposes the limitations for using Eat1 as an AAT but also provides insights into the portfolio of ester compounds produced by many yeasts.

An alternative approach for microbial ester production is described in **Chapter 4**. This approach uses the BMCs of *C. beijerinckii* NCIMB 8052 to encapsulate an AAT enzyme and produce propyl propionate as terminal product. As a proof of principle, the GFP protein is encapsulated in the BMCs by using N-terminal EPs. Unfortunately, due to time restrictions, the encapsulation of AATs and the successful utilization of *C. beijerinckii* BMCs to produce propyl propionate was not (yet) realized.

**Chapter 5** describes the development of a simple genome engineering tool for multiplex gene knock-out in *C. beijerinckii* NCIMB 8052. This tool is based on the CRISPR-FnCas12a system

which is only expressed in the presence of the xylose inducer. 25-100% knock-out efficiency was observed when targeting single genes for deletion whereas, 18% knock-out efficiency was apparent when targeting two genes for deletion. The spacer sequence and spacer order in a CRISPR array is also assessed, showing that it highly influences the knock-out efficiency.

Low HR efficiency along with strong CRISPR-Cas counterselection limit the application of CRISPR-Cas genome engineering tools in non-model prokaryotes. In **Chapter 6**, these limitations are mitigated through the development of SIBR. SIBR was developed and used as a widely applicable gene control system that can control the expression of virtually any gene of interest in any mesophilic prokaryote of interest. SIBR was used to control the expression of the FnCas12a protein (collectively termed SIBR-Cas) and allow more time for HR to occur before induction of counterselection. SIBR-Cas was applied in three different bacteria with high editing efficiencies demonstrating the wide applicability of SIBR.

## General discussion

Modern biotechnology and microbiology are revolutionary. The advances in genome sequencing, genome engineering, fermentation technologies and downstream processing have contributed to the rapid application of biotechnological methods to produce human-destined products in a more efficient and sustainable manner. In this thesis, several approaches are described that showcase the advances of biotechnology and microbiology. The characterization of the Eat1 enzyme along with the alternative approach to encapsulate Eat1 into the BMCs of *C. beijerinckii* NCIMB 8052, are examples of how fundamental knowledge can be transitioned into applied research. Moreover, the development of SIBR-Cas as a widely applicable genome engineering tool, brings the community of microbiologists a step closer to the engineering of any microorganism of interest.

Nonetheless, a “reality check” on esters, the Eat1 enzyme and the genome engineering tools should be performed. Are microbially produced esters an economically viable product? Is the Eat1 enzyme the appropriate catalyst? Why does Eat1 have so many activities? Can SIBR be the ultimate gene expression control system? What else can we do with SIBR? Outstanding questions, unpublished data and future recommendations are discussed in the following section.

### Is it industrially interesting to microbially produce short and medium chain esters?

Short and medium chain esters are inarguably an important family of molecules that find applications in myriads of products in the food and chemical industry (19,108,142,477,478). The production of esters used as bulk chemicals (solvents, lubricants, polymers and drop-in fuels) heavily relies on energy intensive petrochemical methods whereas esters found in food and beverage products are the result of the natural fermentation from yeasts and bacteria (34,108). Although the preceding sentences sound appealing and one may consider short and medium chain esters as an ideal product to produce through microbial conversion, developing this in an economically-efficient manner may not be straightforward.

Typically, for every biotechnologically produced product, the biggest competitor is the petrochemical industry. In other words, the production cost and final price of a biobased product should compete (or even outcompete) petrochemical-based products in order to be commercially viable and/or desirable by the customer/consumer. Therefore, to compete with petrochemical processes, the microbial conversion of substrates into ester products should be as efficient and as cheap as possible. If the production cost is higher for biobased compared to petrochemical processes, governmental subsidies should cover the price difference to render a product competitive

to the market. A way to finance such a subsidy instrument could be by increasing the taxation on CO<sub>2</sub> emitters and channel in parallel the acquired funds to biobased processes. Such an approach will have a double positive effect for the biobased industry. First, there will be an increase in the price of unsustainable products and second, biobased product will be more competitive price-wise. Governmental subsidies should, therefore, not only focus on funding and subsidies for research and development (R&D) but they should also follow and subsidize a biobased products to realize its entry to the market. Obviously, governmental subsidization should be reasonable and cover the price difference for products that are playing an important role in our economy (based on market value and size).

Although petrochemical-based esters are the main competitor of microbially-produced esters, other production processes also provide direct competition. Such processes include: 1) Ester extracts from plant material; 2) *In vitro* enzymatic catalysis using any of the enzymes capable of producing esters (AATs, esterases, HADHs, BVMOs, lipases) and; 3) Traditional Fischer-Speier esterification. The competition between microbial production and the three aforementioned processes is even more fierce as all of them can claim the final product to be “natural”. Even though the European Food Safety Administration (EFSA) does not have a definition for “natural”, the USA Food and Drug Administration (FDA) defines “natural” as “nothing artificial or synthetic...has been included in, or has been added to, a food that would not normally be expected to be in that food”. The definition is complex and contains terms such as “artificial” and “synthetic” or wording such as “normally” that further complicate the definition. Maybe this is the reason that EFSA does not define the word “natural” in the first place, despite the fact that many companies add the “natural” label to their F&B products. Anyhow, the question (and maybe the value) still lies at the premises of whether an ester is considered “natural” or not.

Ester extracts from plant material are considered “natural” only when the plant material is not genetically modified (based on EU regulations). For *in vitro* enzymatic catalysis, the source of the acyl donor and the acyl acceptor should be of “natural” origin (should not be derived from petrochemically-based carbon sources and should not contain GMOs or GM related DNA, RNA or proteins in the final product. Relevant information can be found at EFSA Journal 2011;9(6):2193, (EC) No 1331/2008, (EC) No 1829/2003). The use of unsustainable processes (e.g. petrochemical-based solvents) for *in vitro* enzymatic catalysis is irrelevant for the designation of the final ester compound as “natural”. Even for the traditional Fischer-Speier esterification, which relies on energy intensive petrochemical methods, if the alcohol and the carboxylic acid substrates are from “natural” origin, then the final ester product can be labelled as “natural”. It is therefore clear that competing with very efficient but still unsustainable processes just to have the same “natural” label



at the final product, is not a sustainable business model. However, even in the case where maximum titer, yield and productivity is achieved using microbial hosts to produce bulk esters, the very low market price and very narrow profit margins of such esters, make their production using microbial hosts not interesting unless great attention is attributed to the sustainability of the process and not to the “natural” label (34).

A slight business opportunity for microbial esters may arise for the production of more complex, more “exotic” “natural” esters. Such esters are referred to as fine chemical esters (Wikipedia definition for fine chemicals: “...complex, single, pure chemical substances, produced in limited quantities...sold for more than \$10 Kg<sup>-1</sup>). Fine chemical esters are rare esters usually composed by a complex alcohol moiety and/or a complex acyl moiety. Examples are presented in great reviews (142,478).

Extraction of fine chemical esters from plant material is sometimes not possible due to the rarity of the ester compound or the inefficient extraction processes. Production of “natural” fine chemical esters through *in vitro* enzymatic catalysis or Fisher-Speier esterification is (sometimes) possible but the substrates should be of “natural” origin. However, obtaining complex alcohol and acyl moieties of “natural” origin can be complicated. The most efficient alternative could be through microbial hosts, engineered to produce both the rare alcohol and the acyl moiety and combine them to an ester by means of an enzyme (AAT, esterase, HADH, BVMO, lipase). Some examples following this approach are reviewed in this thesis in **Chapter 2**. Important to note is the extensive metabolic engineering required for the production of both the alcohol and the acyl moiety but also the potential toxic effect the ester may have on the production host. Therefore, even for fine chemical esters, metabolic and toxicity hurdles may hamper their commercialization.

The microbial production of “natural” fine chemical esters is an interesting business proposition, however, two factors may prevent realization: 1) titer, yield and productivity and 2) market size. Although titer, yield and productivity are common benchmarks for every biotechnological process, market size is an important additional criterium for “natural” fine chemical esters. This is mainly due to the rarity and “exoticness” of fine chemical ester compounds which naturally translates to low familiarity of the consumer with such compounds. Therefore, due to low familiarity, low demand follows, inevitably. For this reason, building a start-up on a single, rare fine chemical ester is probably a risky approach. Nonetheless, if a portfolio of multiple fine chemical esters is built (always based on proper market analysis) then commercial viability is possible.

A perfect example to demonstrate the complexity and difficulty to commercialize microbially produced esters was published by Feng *et al.* (2021). In their paper, the authors performed a techno-

economic analysis to assess the economic competitiveness of microbially produced butyl acetate (32). In a nutshell, a total capital investment of \$472 million is required. Operating costs are approximately \$135.5 million year<sup>-1</sup> and the annual production of butyl acetate is estimated to be 83,332 metric tons (MT) year<sup>-1</sup>. According to the techno-economic analysis of the authors, the final production cost of 1 MT of butyl acetate is approximately \$1,364 MT<sup>-1</sup>. In 2019, the market price of butyl acetate (not specified whether this is the price for natural butyl acetate or not) was around \$1,200–1,400, putting the microbially produced esters at a similar but, not advantageous level as the existing production methods. Selling butyl acetate at \$1,364 MT<sup>-1</sup> leads to a revenue of approximately \$113.5 million year<sup>-1</sup> and a negative net profit of (–)\$20 million year<sup>-1</sup> (assuming a selling price equal to production cost). Obviously, this is not a sustainable business model. However, the authors indicate that other products derived from their fermentation processes such as butanol, isopropanol and electricity can be sold at a price of \$900 MT<sup>-1</sup>, \$1145 MT<sup>-1</sup> and \$65 MWh<sup>-1</sup>, respectively. Based on the production of 9,673 MT butanol, 13,555 MT isopropanol and 111,104 MWh surplus electricity, the authors estimate a revenue of \$31.5 million year<sup>-1</sup>. Subtracting the \$20 million year<sup>-1</sup> net loss to the additional \$31.5 million year<sup>-1</sup> revenue from butanol, isopropanol and electricity, results in an \$11.5 million net profit year<sup>-1</sup>. The net profit is used by the authors of the paper to decrease the production cost of butyl acetate from \$1,364 to \$986 year<sup>-1</sup>, making it more competitive to existing processes. Nevertheless, assuming a net profit of \$11.5 million year<sup>-1</sup>, approximately 41 years will be required until a return on investment is observed, something that makes this process not interesting for the investor. A solution to this can be governmental subsidies awarded to this technology based on its environmental and sustainability input. Unfortunately, the environmental impact of this process compared with existing approaches is not described by the authors.

In conclusion, the commercial production of “natural” bulk chemical esters using microbial hosts is currently inhibited by the final titer, yield and productivity and by focusing on the label “natural” instead of focusing on the sustainability of the process. The commercial production of “natural” fine chemical esters is determined by titer, yield and productivity and by the market size of the ester itself.

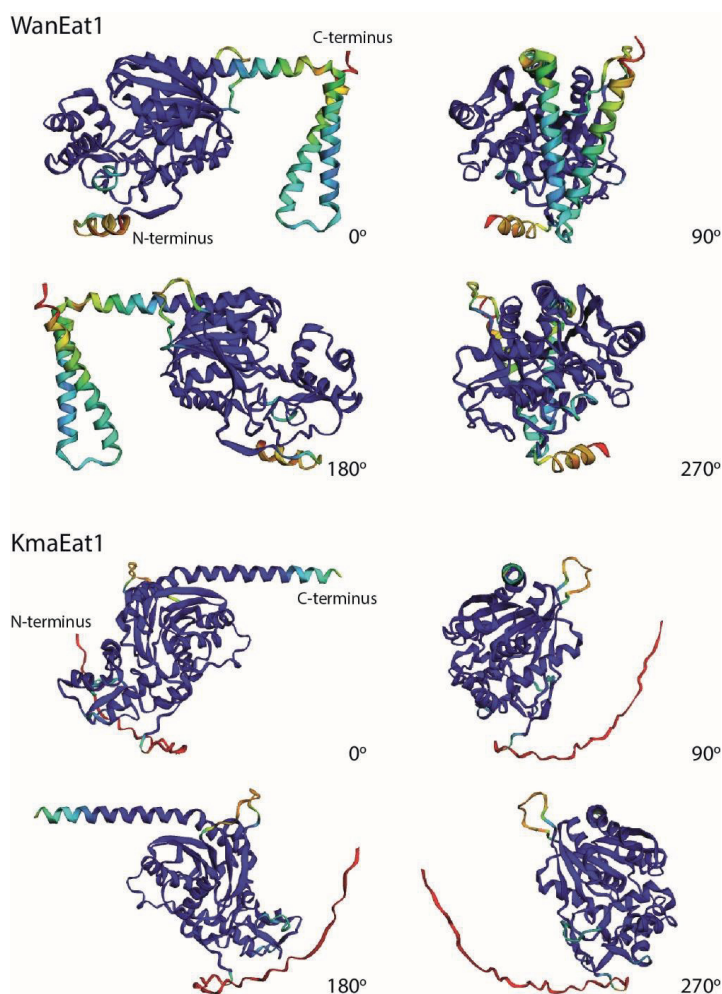
### **Unraveled and hidden mysteries of the Eat1 family: a biochemical and cellular point of view**

The Eat1  $\alpha/\beta$ -hydrolase fold family of enzymes is inarguably a very peculiar family because of its multiple catalytic activities but also due to its cellular localization in the yeast mitochondria. In

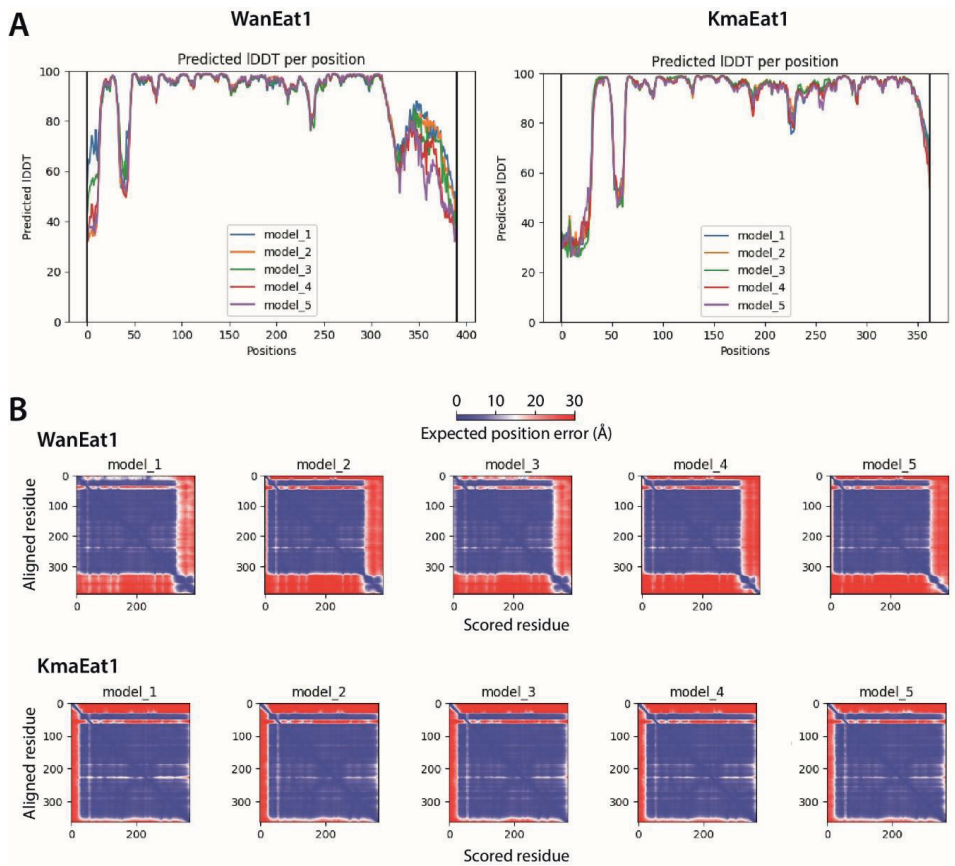
**Chapter 3** we expanded the catalytic portfolio of Eat1 by describing the discovery of its high alcoholysis and thiolysis activities. In total, five activities have been described for Eat1 (AAT, esterase, thioesterase, alcoholysis and thiolysis), all of which are performed by the same active site, typically consisting of a Ser-Asp-His catalytic triad (35,208,376). To accurately elucidate the mechanism that promotes one activity over the other, the crystal structure of Eat1 along with its substrate-protein intermediates is needed. Unfortunately, after multiple attempts by two different laboratories, the crystal structure of Eat1 could not be obtained. However, throughout the course of this PhD thesis, revolutionary protein prediction algorithms have been developed that can predict the 3D structure of proteins with an accuracy of approximately 90% (391,479). To this end, we used one of them (AlphaFold2) to predict the 3D structure of the *Wickerhamomyces anomalus* and *Kluyveromyces marxianus* Eat1 homologs (designated as WanEat1 and KmaEat1, respectively) and tried to elucidate their catalytic and accessory domains (Fig. 1 and 2).

### Eat1 C-terminus

From a first glance at the predicted 3D structure of WanEat1 and KmaEat1, the predicted local-distance difference test (pLDDT) and the predicted aligned error (PAE), we could differentiate a distorted C-terminal peptide which has low pLDDT and very high PAE scores (Fig. 1 and 2). Such distortion was not identified in previous attempts using, at that time, existing 3D structure prediction algorithms such as Phyre2 (35). Since the distorted C-terminus is observed for both WanEat1 and KmaEat1, we superimposed the two proteins to define their differences (Fig. 3). Superimposing revealed a difference in length and sequence of the C-terminal peptide for the two proteins. For WanEat1, the C-terminal peptide starts from the 312<sup>th</sup> aa and continues until the end (391<sup>th</sup> aa) of the protein (80 aa long) whereas a shorter, 32 aa long peptide, is observed at the C-terminus of KmaEat1. The observed distortion and the differences between the C-terminus of WanEat1 and KmaEat1 suggest that the C-terminus of the proteins does not take part in the protein activity and may be genus/species specific. In fact, blasting the 80 aa distorted domain of WanEat1 using BlastP (NCBI), returned hits only for the closely related *W. ciferrii* species but not for other yeast species harboring homologs of Eat1 such as *Kluyveromyces marxianus*, *Kluyveromyces lactis*, *Cyberlindnera jadinii*, *Cyberlindnera fabianii*, *Hanseniaspora uvarum*, *Eremothecium cymbalariae* and *Saccharomyces cerevisiae*. Similarly, blasting the 32 aa long distorted domain of KmaEat1, returned hits only for the closely related *K. lactis* species, further supporting that the C-terminus of Eat1 homologs is genus specific.



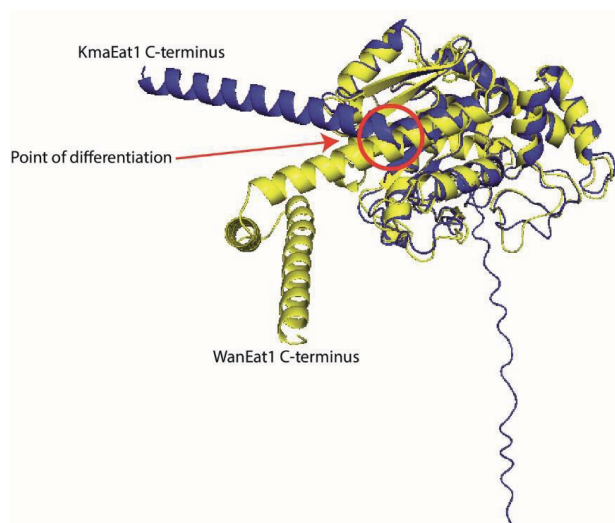
**Figure 1. Predicted 3D structures of WanEat1 and KmaEat1.** 3D structures were predicted and constructed using AlphaFold2. The different colors represent the associated pLDDT score as defined by AlphaFold2: red – very low, (<50); yellow – low (60); green – OK (70); turquoise – confident (80); blue – very high (>90).



**Figure 2. Predicted local-distance difference test (pLDDT) (A) and predicted aligned error (PAE) (B) by predicting the 3D structure of the WT WanEat1 and KmaEat1 through AlphaFold2.** (A) pLDDT shows the percentage of correctly predicted inter-atomic distances throughout the amino acids of the WT WanEat1 and KmaEat1 proteins. pLDDT is used to identify domains and potentially disordered regions. Scores of above 90% show high confidence for a structured domain and may indicate robust side chains or active sites. Confidence scores >70% demonstrate reasonable confidence in structure domains but may indicate mobile loops. Scores <60% indicate pieces of protein which may be unstructured in isolation. (B) PAE measures the confidence in the relative position of two residues. Low expected position error (Å) indicates high confidence in the relative position between pairs of residues whereas high expected position error (Å) indicates low confidence in the relative position between pairs.

Further analysis for the C-terminus of WanEat1 using Uniprot (BLAST against eukaryota) revealed sequence similarity between other, unrelated proteins such as the ubiquitinyl hydrolase 1 from multiple bird species, the protein phosphatase 1 regulatory subunit SDS22 from *Ceratomyces fimbriata* (fungal plant pathogen) and the remorin\_C domain-containing protein from multiple rice species (*Oryza*) (Table 1). Interestingly, sequence similarity is observed between the C-terminus of WanEat1 and the N-terminus of the ubiquitinyl hydrolase 1 protein. In contrast, the C-terminal sequence of WanEat1 is similar to the C-terminal domain of remorin (called remorin C-terminal

anchor or REM-CA) (480). REM-CA mediates the localization of Remorin to the plasma membrane and has been used as a C-terminus anchor to localize GFP into the plasma membrane (481). It is therefore tempting to speculate that the C-terminal sequence of *W. anomalus* Eat1 mediates the localization of Eat1 to the cell membrane or to membranous organelles in yeasts. Future experiments can assess this hypothesis by attaching the C-terminal sequence of Eat1 to GFP and assess its localization in the yeast cell.



**Figure 3. WanEat1-KmaEat1 superimpose using PyMol.** WanEat1 is shown in yellow and KmaEat1 is shown in blue. The point of differentiation for the C-terminus of the two proteins is indicated with a red circle which is essentially the 312<sup>th</sup> aa for WanEat1 and the 331<sup>st</sup> aa for KmaEat1.

The biological role for the potential localization of WanEat1 to the yeast cell membrane or to a membranous organelle is not clear. Intriguingly, the Atf1 and Atf2 AATs from *S. cerevisiae* are known to localize to lipid droplets and to the endoplasmic reticulum through N- and C-terminal sequences (185,190,192,193,482-484). Similarly, the biological role/function of their localization to the lipid droplets or to the endoplasmic reticulum is not clear but a role in the acetylation of sterols and the detoxification from acids was proposed (192).

**Table 1. Selected top hits using the UniProt BLAST tool against the C-terminal distorted sequence (300-391 aa) of *W. anomalus* Eat1.**

Entry	Protein name	Info
A0A1E3P8S6	Ethanol acetyltransferase 1 ( <i>Wickerhamomyces anomalus</i> (st...))	E-value: 220E-57; Score: 442; Ident.: 100.0%
K0KPV8	Ethanol acetyltransferase 1 ( <i>Wickerhamomyces ciferrii</i> (st...))	E-value: 7.4E-15; Score: 175; Ident.: 42.7%
R0LJK7	Ubiquitinyl hydrolase 1 ( <i>Anas platyrhynchos</i> )	E-value: 0.028; Score: 90; Ident.: 33.3%
A0A7L0TPS9	Ubiquitinyl hydrolase 1 ( <i>Chordeiles acutipennis</i> )	E-value: 0.04; Score: 89; Ident.: 31.9%
A0A6J3CGJ0	Ubiquitinyl hydrolase 1 ( <i>Aythya fuligula</i> )	E-value: 0.04; Score: 89; Ident.: 33.3%
A0A0F8CYR2	Protein phosphatase 1 regulatory subunit SDS22 ( <i>Ceratocystis fimbriata</i> f. sp...)	E-value: 0.058; Score: 88; Ident.: 32.4%
A0A2C5XF08	Protein phosphatase 1 regulatory subunit SDS22 ( <i>Ceratocystis fimbriata</i> CBS 1...)	E-value: 0.058; Score: 88; Ident.: 32.4%
I1P2T4	Remorin_C domain-containing protein ( <i>Oryza glaberrima</i> )	E-value: 0.082; Score: 87; Ident.: 32.9%
A0A0D3F8G1	Remorin_C domain-containing protein ( <i>Oryza barthii</i> )	E-value: 0.086; Score: 87; Ident.: 32.9%
A0A0D9YVZ0	Remorin_C domain-containing protein ( <i>Oryza glumipatula</i> )	E-value: 0.087; Score: 87; Ident.: 32.9%
A0A0E0GAP2	Remorin_C domain-containing protein ( <i>Oryza nivara</i> )	E-value: 0.087; Score: 87; Ident.: 32.9%
A0A0E0NIR2	Remorin_C domain-containing protein ( <i>Oryza rufipogon</i> )	E-value: 0.087; Score: 87; Ident.: 32.9%
B9F1H3	Remorin_C domain-containing protein ( <i>Oryza sativa</i> subsp. japonica)	E-value: 0.087; Score: 87; Ident.: 32.9%

Analysis of the C-terminal sequence of KmaEat1 using Uniprot (BLAST against eukaryota), resulted in different hits than WanEat1 (Table 2). Top hits included the RRM domain-containing protein from *Sporisorium graminicola* and an RRP7 related protein from *Ustilago trichophora* (amongst others). The resulting hits were very confound to generate a possible hypothesis for the role of the KmaEat1 C-terminus sequences. This might be due to the short sequence (32 aa) used for BLAST analysis which may yielded unrelated hits.

In summary, both WanEat1 and KmaEat1 have distorted C-terminal sequences which appear to be genus specific. A role in the localization to the yeast cell membrane or to membranous organelles is proposed but, further experimentation is required to reveal the exact role of the C-terminal sequences. We propose sequential truncation of the C-terminal sequences for both WanEat1 and KmaEat1 and assessment of their function (inhibition or enhancement) through alcoholysis or AAT by performing *in vitro* and *in vivo* experiments. Fusion of the WanEat1 and KmaEat1 C-terminal sequences to the C-terminus of GFP is also recommended to assess the role in cellular localization. Lastly, it is likely that the C-terminus of the Eat1 proteins inhibits successful crystallization.

Therefore, for future attempts we recommend truncating the C-terminal sequence from the Eat1 proteins before attempting to crystallize it.

**Table 2. Selected top hits using the UniProt BLAST tool against the C-terminal distorted sequence (331-363 aa) of *K. marxianus* Eat1.**

Entry	Protein name	Info
W0T4A7	Ethanol acetyltransferase 1 ( <i>Kluyveromyces marxianus</i> (str...))	E-value: 9.1E-24; Score: 235; Ident.: 100.0%
A0A4U7KMZ6	RRM domain-containing protein ( <i>Sporisorium graminicola</i> )	E-value: 0.5; Score: 81; Ident.: 51.6%
A0A5C3E8H2	Related to RRP7 - essential protein involved in rRNA processing and ribosome biogenesis ( <i>Ustilago trichophora</i> )	E-value: 0.7; Score: 80; Ident.: 50.0%
Q6CLY8	Ethanol acetyltransferase 1 ( <i>Kluyveromyces lactis</i> (strain...))	E-value: 0.98; Score: 79; Ident.: 78.6%
A0A5P2UC39	AB hydrolase-1 domain-containing protein ( <i>Kluyveromyces lactis</i> )	E-value: 0.98; Score: 79; Ident.: 78.6%

### Eat1 N-terminus

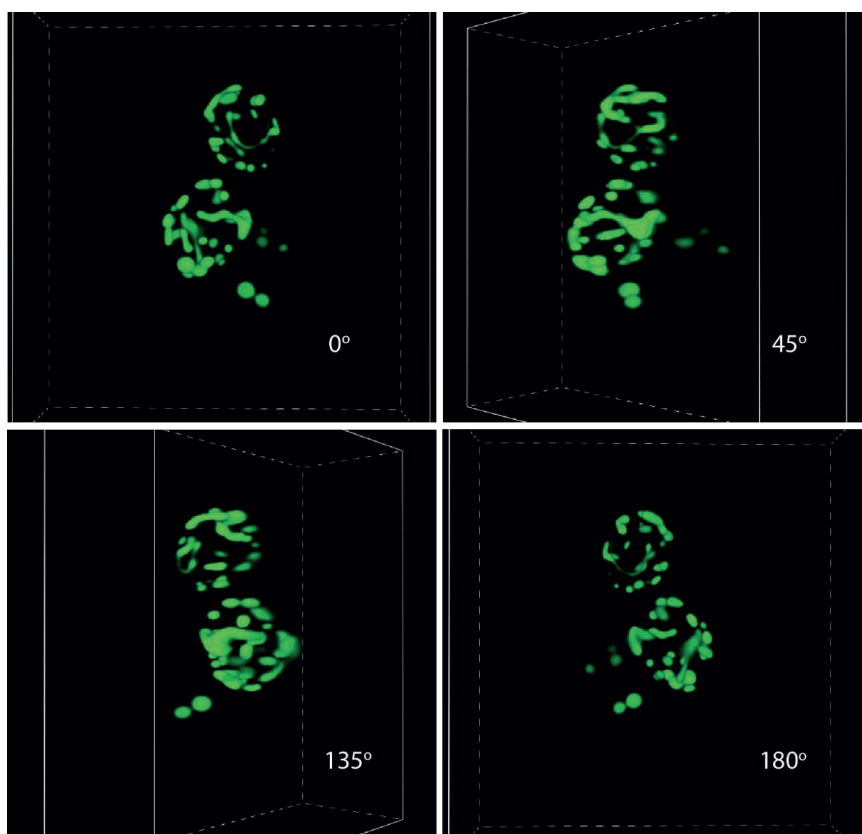
Although the localization of Eat1 to the cell membrane or to membranous organelles has not been investigated yet, the localization of the Eat1 enzyme to the yeast mitochondria is established (at least for some variants) (117,191,404). Localization to the mitochondria is mediated by N-terminal sequences which are predicted to be present in all (but the *S. cerevisiae* Eat1) known Eat1 homologs (404). The characteristic amphipathic helix region which destabilizes the protein and facilitates cross-membrane transport (485), is present at the N-terminus of Eat1 homologs and is visible in the predicted 3D structure of WanEat1 and KmaEat1 (Fig. 1). The N-terminus of WanEat1 and KmaEat1 have low pLDDT and high PAE scores (Fig. 1 and 2), suggesting that they do not interact with the core protein. WanEat1 does not contain predicted MPP/lcp55/Oct1 cleavage sites to remove part of the mitochondrial signal peptide whereas other homologs (*K. marxianus*, *K. lactis*, *C. jadinii*, *C. fabianii* and *H. uvarum*) do (404). Interestingly, the length of the predicted N-terminal mitochondrial sequences varies amongst Eat1 homologs as some are long (36 aa for *K. lactis*) and some others are short (16 aa for *W. anomalus*). Collectively, although the sequence is not conserved amongst species, the N-terminus of Eat1 homologs represents a mitochondrial signal sequence with (in some cases) the possibility of cleavage after transfer of Eat1 into the yeast mitochondria.

The localization of the Eat1 enzyme in the mitochondria of *K. lactis* or *K. marxianus*, was shown in the studies of Kruis *et al.* (2018) and Löbs *et al.* (2018) by fusing GFP to the C-terminus of the Eat1 protein (117,191). Although Löbs *et al.* (2018) created a series of N-terminal mutants ( $\Delta$ 1-14 and  $\Delta$ 1-14) to illustrate the importance and functionality of the sequence for mitochondrial



localization, the repurposing of the Eat1 N-terminal sequence to translocate foreign proteins to the yeast mitochondria was never investigated. To this end, we performed a series of experiments and showed localization of yEGFP (yeast enhanced GFP) to the mitochondria of *K. marxianus* when the N-terminus (34 aa) of the *K. marxianus* Eat1 was attached to the N-terminus of yEGFP (Fig. 4). Transformants bearing an empty vector control did not show fluorescence and transformants constitutively expressing yEGFP showed full cell fluorescence (data not shown). Our results illustrate the feasibility of the 34 aa N-terminal KmaEat1 sequence to be used as a targeting signal for protein import in the mitochondria of *K. marxianus*.

The biological role for the localization of Eat1 in the yeast mitochondria is not fully elucidated yet. However, several studies have shown that increased ethyl acetate production is linked to iron and oxygen limitation in Crabtree-negative yeasts which is associated with an inhibition of the TCA cycle and subsequent accumulation of acetyl-CoA in the mitochondria (88,117,118,127,189). Thomas and Dawson (1978) were the first to propose that an AAT is responsible for a relieve of the acetyl-CoA accumulation and replenishing the CoA pool through the formation of ethyl acetate (189). Eat1 seems to play a role in the aforementioned observations and hypotheses, supported by its localization in the yeast mitochondria (*K. marxianus*, *K. lactis*), its overexpression during oxygen-limited conditions (*W. anomalous*) and the huge decrease in ethyl acetate production in the  $\Delta Eat1$  *K. lactis* mutants (35,117). Despite the supporting data generated by Kruis *et al.* (2017 and 2018), puzzling questions remain unanswered: Why the growth and phenotype of the  $\Delta Eat1$  *K. lactis* mutant is not altered? Why Eat1 relieves the accumulation of acetyl-CoA through its AAT activity instead of its more simple thioesterase activity? Does the production of ethyl acetate have an ancillary function? When Eat1 is deleted, are there other genes/proteins overexpressed to cover the inhibitory effects of acetyl-CoA accumulation in the yeast mitochondria? If another AAT with higher affinity to acetyl-CoA (e.g. Atf1) is transferred into the yeast mitochondria, will it increase the production of ethyl acetate? The answers to these questions will increase our understanding in yeast biology and will aid towards the improvement of biotechnological applications.



**Figure 4.** Localization of yEGFP to the mitochondria of *K. marxianus* when the N-terminus (34 aa) of the *K. marxianus* Eat1 was attached to the N-terminus of yEGFP. Two *K. marxianus* cells are shown next to each other. Pictures were captured using a Nikon Eclipse Ti2 Confocal Fluorescence Microscope and assembled into 3D images using the associated software.

### Catalytic domain

The Eat1 protein homologs are members of the  $\alpha/\beta$ -hydrolase fold enzymes (376). As such, Eat1 is able to catalyze many different reactions by using the same core catalytic machinery (208,376). The catalytic machinery of  $\alpha/\beta$ -hydrolase fold enzymes typically includes a cap domain and a catalytic domain composed by the catalytic triad and the oxyanion hole (208). Kruis *et al.* (2017) identified the position of the catalytic triad Ser-His-Asp by aligning several Eat1 homologs and showed that the serine moiety of the catalytic triad is located in a conserved GX<sub>1</sub>SX<sub>2</sub>G motif which is called the catalytic elbow. The oxyanion hole and the cap domain residues have not been assigned, yet, to any of the Eat1 homologs. Because the nature of the oxyanion hole and the cap domain can affect the substrate binding and specificity of the enzyme, we sought to discover the domains and residues that may affect the enzyme's catalytic capacity.

### *Oxyanion hole*

The oxyanion hole is involved in the stabilization of the substrates. It is characterized by two residues which interact with the oxygen of the carbonyl group of the substrate and stabilize the tetrahedral intermediate formed during the reaction (486). The first residue is generally located at the N-terminus of the protein while the second residue is always located at the C-terminus of the catalytic serine moiety in the catalytic elbow. Based on the surrounding of the first residue, two different types of oxyanion holes can be distinguished,  $GY_1$  and  $GGGY_1$ , where  $Y_1$  represents the first residue involved in the oxyanion stabilization. The second residue of the oxyanion hole is the  $X_2$  residue of the  $GX_1SX_2G$  catalytic elbow. Furthermore, according to the oxyanion hole type, the enzyme's substrate specificity can be determined. In particular, enzymes with a  $GY_1$  oxyanion hole type are usually more specific to medium and long chain substrates, while  $GGGY_1$  types are more specific to short chain substrates (487). To visualize the oxyanion hole, all the Gly residues present in WanEat1 were checked. While no  $GGGY_1$  motif was apparent, searching for single Gly residues present in close proximity (in the structural model) to the catalytic Ser residue (121) and essentially to the  $X_2$  residue, the Gly (54) and Ile (55) residues were selected as the putative residues of the WanEat1 oxyanion hole with a  $GY_1$  motif.

Once the putative oxyanion hole was identified in WanEat1, multiple sequence alignment was performed between fifteen Eat1 homologs. The catalytic triad and the oxyanion hole residues derived from the sequence alignment are shown in Figure 5. Overall, both the first ( $Y_1$ ) and the second ( $X_2$ ) residues are conserved amongst the members of the Eat1 family. In particular, the first residue is either a Leu or an Ile, with Leu being more conserved than Ile. Additionally, the second residue located in the catalytic elbow is a highly conserved Leu in all homologs, apart from EcyEat1 and SceEat1 which show a methionine.

	Oxyanion hole residues (GY <sub>1</sub> )	Catalytic triad Catalytic elbow (GX <sub>1</sub> SX <sub>2</sub> G)	Asp	His
WanEat1	--FVHG FGSKKNY--	--LIGYSLGAKIC--	--IDNS--	--ATHFI--
WciEat1	--FVHG FGSKKNY--	--LIGYSLGAKIC--	--IDNS--	--ATHFI--
CjaEat1	--FLHG FGSKKSY--	--LVGYSLGAKIS--	--IDNA--	--S SHDI--
CfaEat1	--FVHG FGSKKSY--	--LVGYSLGAKVS--	--IDNA--	--SAHDI--
HguEat1	--FVHGLFGSKKNY--	--LIGYSLGAKVS--	--IDNS--	--ATHFI--
HopEat1	--FXHGLFGSKKNY--	--LIGYSLGAKVS--	--IDNS--	--ATHFI--
HuvEat1	--FVHGLFGSKKNY--	--LIGYSLGAKVS--	--IDNS--	--ATHFI--
HvaEat1	--FVHGLFGSKRNY--	--LIGYSLGAKVS--	--IDNS--	--ATHFI--
HosEat1	--FVHGLYGSKRNY--	--LIGYSLGAKIS--	--IDNS--	--ATHFI--
KmaEat1	--FYHGLLGSKRNY--	--LVGYSLGAKVA--	--IDNS--	--ATHFI--
KlaEat1	--FFHGLLGSKRNY--	--LIGYSLGAKVG--	--IDNA--	--ATHFV--
LfeEat1	--FLHGLFGSKRNY--	--VVGYSLGAKIA--	--IDNS--	--ATHFI--
EcyEat1	--FIHGLFGWKRFY--	--LIGYSMGKMS--	--IDNA--	--TTHNV--
PtaEat1	--FIPGLFGSTKMY--	--VSGFSLGGKVA--	--VDIS--	--TYHLI--
SceEat1	--LIHGLFGNKLNN--	--LIGHSMGKVA--	--IDNA--	--AGHWV--

**Figure 5. Multiple sequence alignment of 15 Eat1 homologs identified in yeasts.** The catalytic triad is shown on the right while the GY<sub>1</sub> residues of the oxyanion hole is shown on the left. The sequences are reported from the N-terminus to the C-terminus. Asterisks indicate the position of the first and second residue of the oxyanion hole. Abbreviations: WanEat1: *W. anomalus*; WciEat1: *W. ciferrii*; CjaEat1: *C. jadonii*; Cfa: *C. fabianii*; HguEat1: *H. guilliermondii*; HopEat1: *H. opuntiae*; HuvEat1: *H. uvarum*; HvaEat1: *H. valbyensis*; HosEat1: *H. osmophila*; KmaEat1: *K. marxianus*; KlaEat1: *K. lactis*; LfeEat1: *L. fermentati*; EcyEat1: *E. cymbalariae*; PtaEat1: *P. tannophilus*; SceEat1: *S. cerevisiae*.

### Cap domain

The lid or cap domain varies widely amongst the  $\alpha/\beta$ -hydrolase fold family of enzymes. Some enzymes have a long (~100-355 aa) cap domain (488), some others have a short (~40 aa) cap domain (489) or, it can be missing completely from the enzyme (208). Also, the position where it occurs in the protein sequence varies, since some enzymes have the cap domain between the  $\beta 6$  and  $\beta 7$  strands of the catalytic domain whereas some others have it at the N-terminus (208,490). Moreover, some cap domains contain residues of the catalytic domain whereas some others just form tunnels that mediate substrate specificity. Therefore, it is not straightforward to define the cap domain of an  $\alpha/\beta$ -hydrolase fold enzyme solely by looking at the protein sequence. However, a typical characteristic of the cap domain are the consecutive helical secondary structures and the positioning of the cap domain. Hence, the predicted 3D structure of WanEat1 and KmaEat1 were analyzed to define the cap domain.

Both WanEat1 and KmaEat1 share a homologous ~100 aa long sequence between  $\beta 5$  and  $\beta 6$  which is composed of five consecutive helices; a typical characteristic of cap domains (Fig. 6). None of the residues of the chosen sequence seems to be taking part in the catalytic domain (catalytic triad or oxyanion hole) but it appears that the selected domain takes part in the formation of the catalytic pocket (Fig. 7). As mentioned previously, the cap domain may affect substrate specificity as it takes part in the formation of the tunnel that leads to the catalytic pocket. The residues that shape the

## Chapter 7

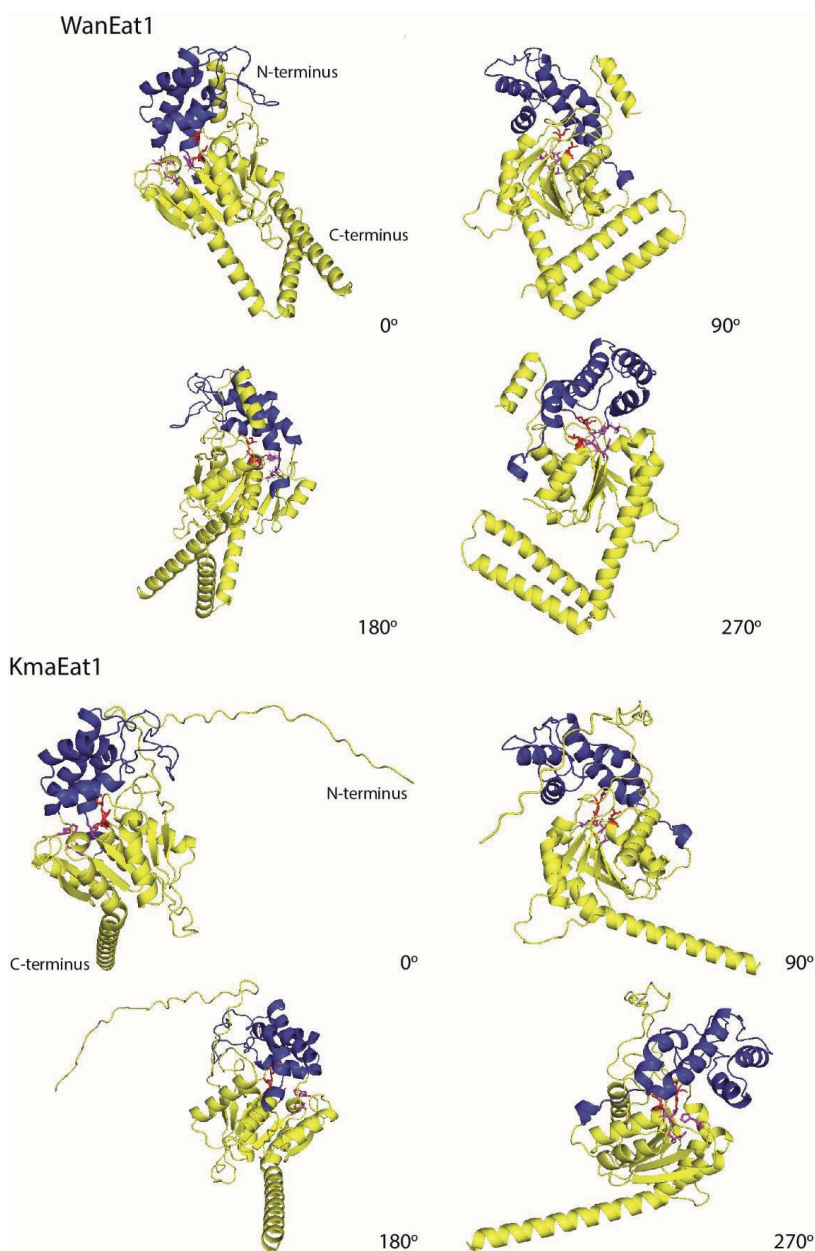
tunnel for WanEat1 are IEIFLTQFI and IPNGGI and for KmaEat1 are IKPLL TALV and IPDAGI. The cap domain can be both in an open and closed conformation but, unfortunately, this can be defined only by X-ray structures.

### *Tunnel / Catalytic pocket*

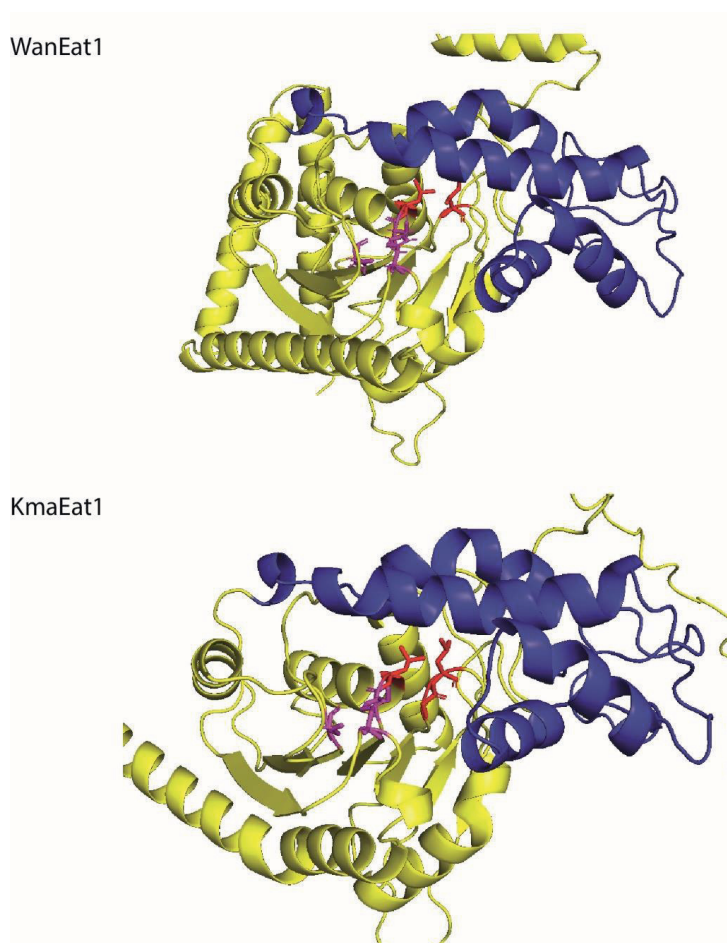
In **Chapter 3** we have shown that the WanEat1 protein has high alcoholysis activity in an aqueous environment (376). The alcoholysis activity is almost ten times higher than the typical esterase activity of  $\alpha/\beta$ -hydrolase fold enzymes, demonstrating that alcohols are preferred over water as acyl acceptors, even in an aqueous environment. To favor the binding of an alcohol over water, the catalytic pocket of many  $\alpha/\beta$ -hydrolase fold enzymes is hydrophobic (208). Therefore, it is fair to presume that the catalytic pocket of Eat1 enzymes is also hydrophobic as it favors alcohols over water.

After defining the catalytic triad, oxyanion hole and cap domain of WanEat1 and KmaEat1, we could assess whether the hydrophobicity of the catalytic pocket supports our observations. Interestingly, the oxyanion hole residues Y<sub>1</sub> (Ile or Leu) and X<sub>2</sub> (Leu) are both hydrophobic. Moreover, the X<sub>1</sub> residue (Tyr) of the catalytic elbow is also hydrophobic. Following, the residues of the cap domain that shape the tunnel are highly hydrophobic (66.67% and 50% hydrophobic amino acids for the WanEat1 IEIFLTQFI and IPNGGI residues, respectively; 77.78% and 66.67% for the KmaEat1 IKPLL TALV and IPDAGI residues, respectively). Altogether, the combination of our observations in **Chapter 3** and the analysis performed on the predicted structure of Eat1 proteins, show that the catalytic pocket of Eat1 proteins is highly hydrophobic.

In summary, by using state-of-the-art protein structure prediction algorithms, we were able to predict the 3D structure of WanEat1 and KmaEat1 along with their distorted N- and C-terminal sequences, the oxyanion hole and the cap domain of the two proteins and also the residues of the cap domain that create a hydrophobic tunnel that mediates substrate specificity. This knowledge can be used in the future to elucidate the biological role of Eat1 in yeasts, increase or widen its substrate specificity through mutagenesis and successfully crystallize the protein.



**Figure 6. Predicted cap domain of WanEat1 and KmaEat1.** The predicted cap domain is shown in blue whereas the rest of the protein is shown in yellow. The catalytic triad is shown in magenta sticks and the oxyanion hole residues are shown in red sticks. 3D structures were predicted using AlphaFold2 and constructed using PyMol.



**Figure 7. Close-up on the predicted tunnel of WanEat1 and KmaEat1.** The predicted cap domain is shown in blue whereas the rest of the protein is shown in yellow. The catalytic triad is shown in magenta sticks and the oxyanion hole residues are shown in red sticks. 3D structures were predicted using AlphaFold2 and constructed using PyMol.

## Expanding the CRISPR-Cas12a toolbox in *Clostridium beijerinckii* NCIMB 8052

In **Chapter 5** we expanded the CRISPR-Cas toolbox for *C. beijerinckii* NCIMB 8052 by showing single and double gene knock-outs using CRISPR-FnCas12a. We also sought to expand the CRISPR-Cas toolbox in *C. beijerinckii* NCIMB 8052 for gene knock-down but also for gene disruption through CRISPR-Cas mediated base editing.

### Gene knock-down

To achieve gene knock-down, the WT FnCas12a was replaced with a catalytically inactive FnCas12a (referred to as dead; dFnCas12a) which has the amino acid substitutions D917A and E1006A. As a target gene for silencing, the *Spo0A* gene was chosen since it shows distinct morphological changes including lack of spore formation and elongation of the cells. Transformed *C. beijerinckii* NCIMB 8052 cells harboring plasmids with either a non-targeting spacer or an *Spo0A*-targeting spacer, were cultured under inducing (xylose) or non-inducing (glucose) conditions for 72 h. Every 24 h, samples were recovered and viewed under the microscope to assess morphological changes. Important to note is that the *Spo0A*-targeting spacer sequence (5'-TAGAAGTAATATATTGTGTT-3') was targeting the -35 and -10 box of the *Spo0A* promoter sequence and was selected following the teachings of Bikard *et al.*, (2013) (491).

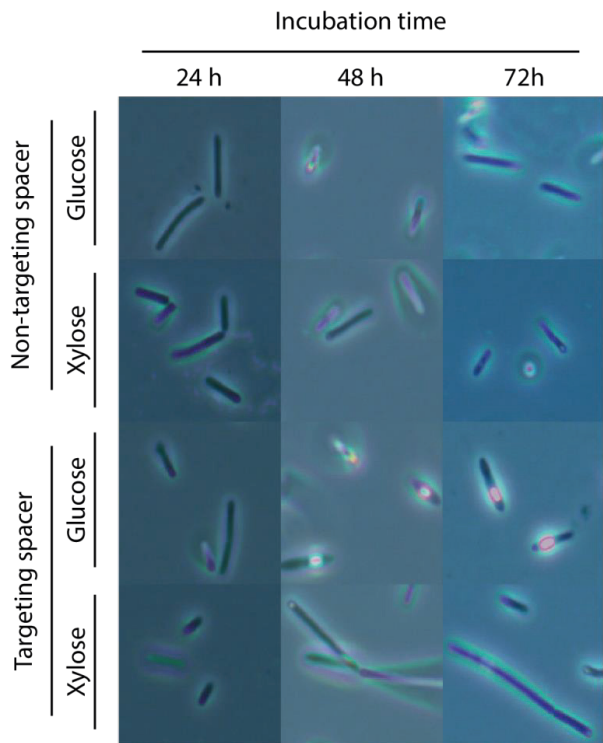
The results of this experiment are illustrated in Figure 8. Elongated, non-sporulating cells were observed only when the transformed cells expressed the *Spo0A* targeting spacer and were grown in medium containing xylose. In contrast, cells expressing the *Spo0A* targeting spacer and grown in medium containing glucose, formed single cells. Control transformants expressing the non-targeting spacer did not show the elongated cell phenotype. To confirm and quantify the downregulation of the *Spo0A* gene, future experiments should include qPCR assays.

### Gene disruption through base editing

Homologous recombination-based CRISPR-Cas tools (e.g. the one developed in **Chapter 5 and 6**) rely on the native homologous recombination machinery of the bacterium which is regularly inefficient, requires the introduction of homologous arms on the replicative plasmid and is often laborious. To circumvent these limitations, CRISPR-Cas base-editors have been developed which make use of a cytidine deaminase (rAPOBEC1) fused to a catalytically inactive Cas9 or Cas12a (492,493). The Cas protein is used to navigate the cytidine deaminase to the target site which in turn substitutes a nearby cytidine base with a uracil base. The uracil is then substituted to a thymine upon replication. To inhibit the cellular repair mechanism from correcting the cytidine substitution, a uracil glycosylase inhibitor (UGI) is fused to the Cas-deaminase fusion protein. By using a



CRISPR-Cas base-editor, the disruption of the gene of interest (e.g. through the introduction of stop codons) can be achieved circumventing all of the aforementioned limitations exhibited by homologous recombination-based approaches.

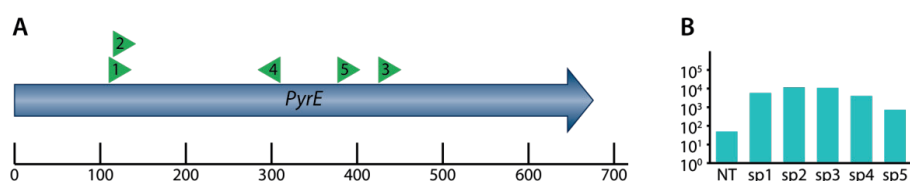


**Figure 8. *Spo0A* knock-down experiment using CRISPR-FnCas12a.** *C. beijerinckii* NCIMB 8052 cells were transformed with a plasmid expressing dFnCas12a and either an *Spo0A* targeting or a non-targeting spacer. Transformed cells were incubated in inducing (xylose) or non-inducing (glucose) medium for 72 h. The morphology of the cells was assessed every 24 h. A representative portion of the image captured with the microscope is shown for each culture.

An rAPOBEC1-nCas9-UGI base editor was previously used in *C. beijerinckii* NCIMB 8052 to disrupt the orotate phosphoribosyltransferase (*PyrE*) and *Spo0A* genes (409). Although successful, this tool relied on single genetic alterations at a time. We sought to improve this by using CRISPR-FnCas12a and achieve multiplex base editing in one transformation event. As a proof of concept, we used the rAPOBEC-dFnCas12a-UGI fusion protein and targeted the *PyrE* gene for disruption due to its easy counterselection phenotype (resistant to 5-FOA). Similar to our previous endeavors, the xylose inducible promoter was used to control the expression of the rAPOBEC-dFnCas12a-UGI fusion protein. Moreover, five spacers were selected to knockout the *PyrE* gene. Spacer 1, 2

and 3 were designed to introduce a premature stop codon and spacer 4 and 5 were designed to introduce an amino acid substitution in the substrate binding sites of the protein (Fig. 9A).

To disrupt the *PyrE* gene, *C. beijerinckii* NCIMB 8052 cells were transformed with a plasmid expressing the rAPOBEC-dFnCas12a-UGI fusion protein and one of the five selected spacers. For control, a non-targeting spacer was used. Transformed cells were grown for 72 h in selective medium containing xylose for induction of the rAPOBEC-dFnCas12a-UGI fusion protein. The cells were then plated on selective medium containing 5-FOA to select for cells where disruption of the *PyrE* occurred. A clear difference between the non-targeting and targeting spacers was observed, indicating that base-editing caused disruption of the *PyrE* gene using the different spacers (Fig. 9B). 16 colonies were picked from each plate, and the *PyrE* gene was PCR amplified and sequenced through Sanger sequencing (Macrogen).



**Figure 9. Base-editing spacers and survival test.** (A) The ruler indicates the nucleotide position in the *PyrE* gene. The position and the orientation of spacers 1 to 5 is indicated with green arrow heads. (B) Colony count obtained after transforming *C. beijerinckii* NCIMB 8052 with the five base-editing constructs and the non-targeting (NT) control, growing the transformants for 72 h in inducing (xylose) medium and then plating them on plates containing 5-FOA (n=1).

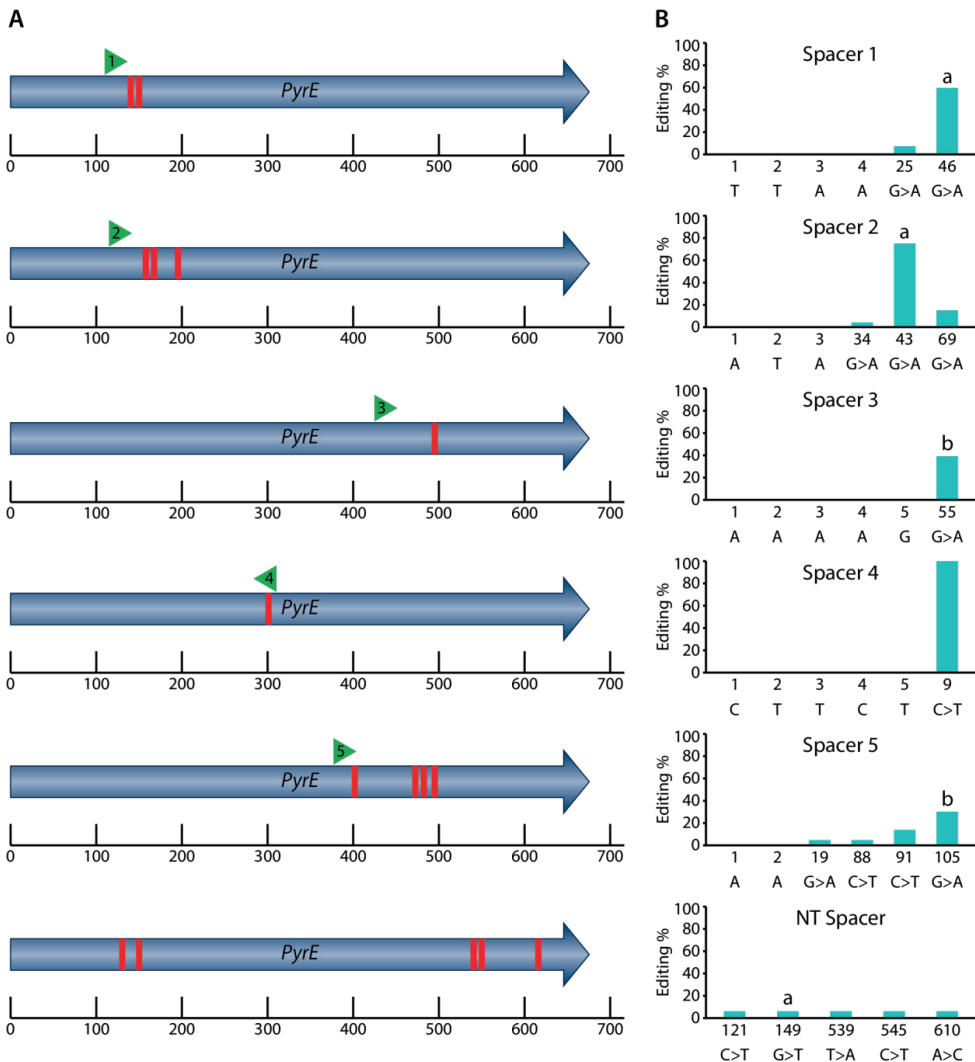
All base-editing constructs showed a base-editing efficiency ranging from 3% to 100% (Fig. 10). However, an irregular editing window of target positions 9-105 was observed, in contrast to the 8-13 editing window observed by LbCas12a base editors (493). Spacer 4 was the only one that showed editing within the expected editing window at a single nucleotide (9<sup>th</sup> nt of the spacer) with 100% efficiency. Spacer 5 showed editing within the spacer region (19<sup>th</sup> nt of the spacer) but with very low efficiency (~5%). Also, spacer 5 showed editing of other bases at the 88<sup>th</sup> (~3%), 91<sup>st</sup> (~15%) and 105<sup>th</sup> (~25%) nts from the start of the spacer. The edited 105<sup>th</sup> nt of spacer 5 was also edited by spacer 3 (55<sup>th</sup> nt) at a frequency of ~40%. Interestingly, this was the only edited base observed for spacer 3. The 46<sup>th</sup> nt of spacer 1 was also edited by spacer 2 (43<sup>rd</sup> nt) with ~55% and ~75% editing efficiency, respectively.

With the current dataset, we cannot draw firm conclusions about the editing window of our rAPOBEC-dFnCas12a-UGI construct as it appears to be large (9-105) and often outside of the spacer sequence region. However, it is clear that the base editing events were associated with the

position of the spacer and its directionality. For example, spacer 1 and 2 show base editing at the 120-200 nt region of *PyrE* whereas spacer 3 and 5 show base editing only at the 400-500 nt region of *PyrE*. Intriguingly, spacer 4 shows only one single base mutation with 100% efficiency. Moreover, the NT spacer control shows low (<~10%) base editing at random locations all over the *PyrE* gene which are often not related to the characteristic G>A or C>T base editing substitution of the cytidine deaminase base editor.

Although base editing was observed on the screened colonies, culture PCR did not show base editing for any of the tested spacers (data not shown). This result suggested that the base editing efficiency was very low and could be observed only after selection on medium containing 5-FOA. Therefore, we cannot draw conclusions on the editing efficiency at a population level since we were selecting for mutants on selective medium. Nonetheless, editing using the 5 spacers was higher than the NT spacer control as indicated by the number of obtained colonies when plated on medium containing 5-FOA and by the low editing efficiency of single bases.

To conclude, future experiments should include more target genes and more sites within a gene to determine the editing window and the editing efficiency. Also, to determine the editing efficiency of a non-screen-able gene, whole culture sequencing could be performed instead of single colony PCR followed by sequencing as performed in this study. Moreover, different constructs should be tested including variable linker length between the UGI, Cas protein and base editor, the test of different base editors but also the test of the more established AsFnCas12a and LbFnCas12a (492-494).



**Figure 10. *PyrE* disruption through CRISPR-FnCas12a-mediated base-editing.** (A) Position of mutant bases along the *PyrE* gene using the 5 different targeting spacers and the non-targeting (NT) spacer. (B) Base editing percentage at the designated bases. For the targeting spacers, the base numbering starts at the start of the spacer whereas for the NT spacer the base numbering starts at the beginning of the gene. A base substitution is indicated by “>” where the base at the left of “>” is the WT base and the base at the right of “>” is the resulting mutant base. “a” and “b” indicate the same base position along the *PyrE* gene.

## SIBR-Cas: The first-to-try CRISPR-Cas tool for mesophilic prokaryotes

In **Chapter 6** we successfully knocked-out several genes in three WT bacterial species (*E. coli* MG1655, *P. putida* KT2440 and *Flavobacterium* IR1) using SIBR-Cas. Based on the simplicity and wide applicability of SIBR-Cas, we believe that SIBR-Cas should be the first-to-try CRISPR-Cas tool in WT mesophilic prokaryotes. Hereby, we provide a step-by-step protocol to be able to transfer SIBR-Cas to any mesophilic prokaryote of interest.

Before using SIBR-Cas in the prokaryote of interest, the below check boxes should be ticked-off:

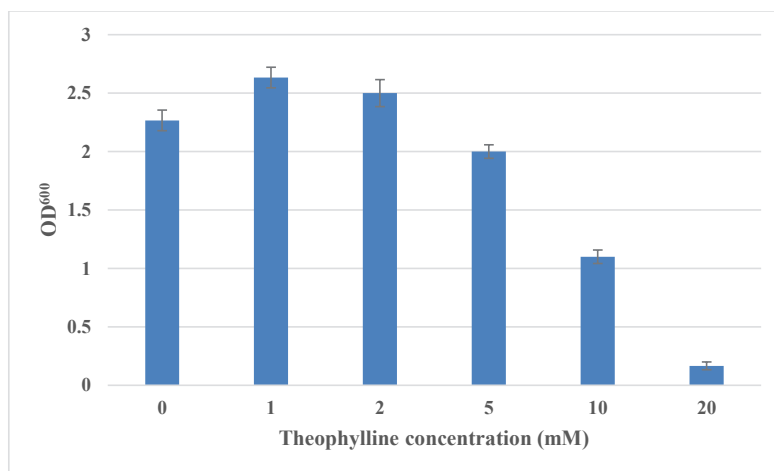
- ☐ Is the prokaryote transformable? (through natural competency, electroporation, heat shock, conjugation or transduction)
- ☐ Does a replicative plasmid exist? (high copy number plasmids are preferred)
- ☐ Is the prokaryote a mesophile? (20-37°C)
- ☐ Are inducible promoters lacking, not yet characterized or leaky?
- ☐ Is homologous recombination very inefficient in the prokaryote of interest?

Important to note are the potential T4 *td* intron inhibitors reported by various studies (465-475). The effect of these inhibitors has been confirmed *in vitro* but, *in vivo* studies are lacking. Therefore, caution should be taken when these inhibitors are present in the growth medium and should be avoided when possible. Compounds that inhibit T4 *td* intron functionality include: Novobiocin, Spectinomycin, Neomycin B, Gentamicin, Paromomycin, Streptomycin, Chloramphenicol, Neamine, Apramycin, Hygromycin B, Garamine, Gentamicin C, 2'' Phospho-gentamicin C, Gentamicin B, G-418, NAD<sup>+</sup> NADH, ADP-ribose, Deamido-NAD<sup>+</sup>, 3'-NADP<sup>+</sup>, NADP<sup>+</sup> NADP<sup>+</sup>-dialdehyde, NADPH, 1,N6-etheno-NADP<sup>+</sup>, Pyridoxal phosphate, Pyridoxine, Pyridoxamine, Pyridoxal, Pyridoxic acid, Thiamine Pyrophosphate, Thiamine monophosphate, Thiamine and Thiochrome.

To transfer SIBR-Cas to the prokaryote of interest, the following stepwise guideline can be followed:

1. **Test theophylline uptake and toxicity.** A growth assay where theophylline is present at different concentrations in the growth medium should be performed to define the theophylline uptake and toxicity. Theophylline concentrations used for prokaryotes range between 0 and 20 mM. Typically, growth reduction is observed as theophylline is toxic to prokaryotes at high concentrations. The concentration of theophylline that is appropriate to use in follow-up experiments is the one that slightly reduces the growth of the prokaryote of interest (compared to the 0 mM theophylline concentration). An example is provided in

Figure 11 where a theophylline concentration of 2-5 mM is the appropriate concentration to use for induction.



**Figure 11. Fictional example for theophylline uptake and toxicity test in the bacterium of interest.** The prokaryote of interest is grown in medium containing 0, 2, 5, 10 and 20 mM theophylline. Initial OD<sub>600</sub> was equal to 0.1 and the cultures were grown for 24 h. An appropriate theophylline concentration is considered between 2-5 mM as it does not cause significant cell death compared to the 0 mM theophylline control.

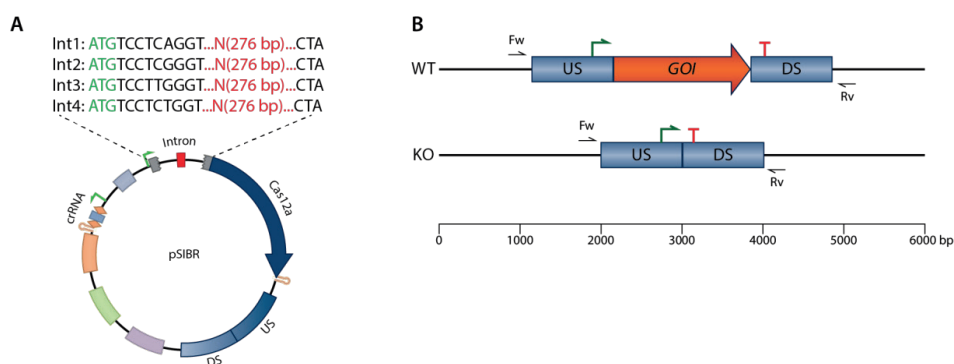
2. **Test the targeting efficiency of CRISPR-Cas.** Before trying any of the SIBR variants, the targeting efficiency of a constitutively expressed CRISPR-Cas system shall be tested to ensure functionality of the system. To do this, a replicative plasmid which constitutively expresses the Cas protein of interest (e.g. Cas9 or Cas12a) along with a constitutively expressed targeting CRISPR array should be constructed (for choosing and inserting an appropriate targeting spacer, see the spacer selection and spacer insertion protocol in the supplementary information of **Chapter 5**). As a control, a non-targeting spacer should replace the targeting spacer to avoid cell death. Preferably, the promoter chosen to drive the expression the CRISPR array should be stronger than the promoter chosen to drive the expression of the Cas protein. However, in the case where promoter strength is not known, any promoter should be sufficient. We strongly recommend the use of a codon harmonized Cas protein. Protocols and information on codon harmonization can be found in the Galaxy server (<https://galaxyproject.org/use/codon-harmonizer/>) created by Claassens *et al.* (2017) (419). Regularly a 10<sup>3</sup> fold reduction should be observed in the obtained colony forming units (CFUs) when a targeting CRISPR-array is used.
3. **Test the inducibility of SIBR-Cas.** Four intron variants (Int1, 2, 3 and 4) have been used in **Chapter 6** to control the expression of FnCas12a (495). The intron variants vary in

splicing efficiency with Int1 having the lowest and Int4 the highest splicing efficiency. Since the splicing efficiency may differ in the prokaryote of interest, we highly recommend testing all four intron variants. The intron variants (Table 3) should be introduced directly after the ATG start codon of the Cas gene as demonstrated in Fig. 12A. Inducibility of SIBR-Cas should be tested following the below steps:

- a. Transform the SIBR-Cas variants containing the targeting CRISPR array in the prokaryote of interest. Triplicate transformation is recommended for each SIBR-Cas variant.
  - b. Serially dilute the transformants (up to the  $10^{-5}$ ) and perform a spot assay (3  $\mu$ L drops) on solid medium containing the appropriate antibiotic (to maintain the SIBR-Cas plasmid) and the theophylline inducer (the theophylline concentration is defined at step 1). As a control, repeat the spot assay on plates without theophylline. Reduction of CFUs should be visible in the medium containing theophylline. Note 1: in case where the transformation efficiency is low, we recommend culturing the transformants overnight to increase their biomass and, hence, be able to perform the spot assays described above. Note 2: for the splicing of the intron we recommend a 20-30°C incubation temperature. Note 3: choose the two best splicers to perform the following experiments.
4. **Assess the knock-out efficiency using SIBR-Cas.** To knock-out the gene of interest (GOI), homologous arms corresponding to upstream and downstream genomic sequences of the GOI should be introduced on the SIBR-Cas plasmid. The length of the homologous arms can vary (100 to 2000 bp) but we recommend using 1000 bp as a starting point. As a control, the homology arms should be introduced on the plasmid constructed in step 2 which constitutively expresses the Cas protein and contains a targeting CRISPR-array. The knock-out efficiency should be tested following the below steps:
- a. Transform the SIBR-Cas and control plasmids containing the targeting CRISPR array and the homology arms in the prokaryote of interest. Triplicate transformation is recommended for each SIBR-Cas variant.
  - b. Allow enough time for homologous recombination to occur. This step is complex since different prokaryotes may require different time length for homologous recombination to occur. For *E. coli* MG1655, *P. putida* KT2440 and *Flavobacterium* IR1, 1 h, 2 h and 96 h were used, respectively (495). Therefore, a trial and error approach should be followed by testing induction directly after transformation and recovery, after overnight growth or after prolonged (24 to 96 h) incubation.

Induction can be performed either in liquid or on solid medium followed by incubation at the appropriate temperature (we recommend 20–30°C).

- c. Obtained colonies should be screened using colony PCR (cPCR) for WT or knock-out genotype with primers that bind outside the homologous arms upstream and downstream of the GOI (see Fig. 12B).



**Figure 12. Use of SIBR to control the *FnCas12a* gene (A) and example knock-out of the gene of interest (GOI) (B).** (A) Example plasmid using SIBR to control the *FnCas12a* gene. The four SIBR intron variants are introduced directly after the ATG start codon of the *FnCas12a* gene. (B) 1000 bp upstream (US) and downstream (DS) homology arms are used to knock-out the GOI. Complete knock-out of the GOI is shown. Forward (Fw) and reverse (Rv) primers were designed outside the homology arms to screen for wild type (WT) or knock-out (KO) genotypes.

5. **Targeting any genomic locus for gene knock-out or knock-in.** When the previous steps lead to a successful outcome, other genomic loci can be targeted for gene knock-out or knock-in. Also, in the case where Cas12a is used as the Cas nuclease, multiple genes can be targeted for deletion simultaneously by inserting two or more spacers at the CRISPR array and by inserting two or more pairs of homology arms to the SIBR-Cas plasmid.



**Table 3.** Sequence of the four SIBR intron variants. The 5' and 3' exonic sequences are in bold letters. The theophylline aptamer sequence is underlined. The intron variants should be introduced between the ATG start codon and the second codon of the GOI.

SIBR intron variant	Sequence (5' > 3')
Int1	<b>TCCTCAGGTTAATTGAGGCCTGAGTATAAGGTGACTTATACTTGT</b> AATCTATCTAAACGGGGAACCTCTCTAGTAGACAATCCCGTGCTAA ATTG <u>ATACCAGCATCGTCTTGATGCCCTTGGCAG</u> CATAAATGCCTA ACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACTCGAAACGATA GACAACTTGCTTTAACAAGTTGGAGATATAGTCTGCTCTGCATGGT GACATGCAGCTGGATATAATTCCGGGGTAAGATTAACGACCTTAT CTGAACATAATGCTA
Int2	<b>TCCTCGGGTTAATTGAGGCCTGAGTATAAGGTGACTTATACTTGT</b> AATCTATCTAAACGGGGAACCTCTCTAGTAGACAATCCCGTGCTAA ATTG <u>ATACCAGCATCGTCTTGATGCCCTTGGCAG</u> CATAAATGCCTA ACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACTCGAAACGATA GACAACTTGCTTTAACAAGTTGGAGATATAGTCTGCTCTGCATGGT GACATGCAGCTGGATATAATTCCGGGGTAAGATTAACGACCTTAT CTGAACATAATGCTA
Int3	<b>TCCTTGGGTTAATTGAGGCCTGAGTATAAGGTGACTTATACTTGT</b> AATCTATCTAAACGGGGAACCTCTCTAGTAGACAATCCCGTGCTAA ATTG <u>ATACCAGCATCGTCTTGATGCCCTTGGCAG</u> CATAAATGCCTA ACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACTCGAAACGATA GACAACTTGCTTTAACAAGTTGGAGATATAGTCTGCTCTGCATGGT GACATGCAGCTGGATATAATTCCGGGGTAAGATTAACGACCTTAT CTGAACATAATGCTA
Int4	<b>TCCTCTGGTTAATTGAGGCCTGAGTATAAGGTGACTTATACTTGT</b> AATCTATCTAAACGGGGAACCTCTCTAGTAGACAATCCCGTGCTAA ATTG <u>ATACCAGCATCGTCTTGATGCCCTTGGCAG</u> CATAAATGCCTA ACGACTATCCCTTTGGGGAGTAGGGTCAAGTGACTCGAAACGATA GACAACTTGCTTTAACAAGTTGGAGATATAGTCTGCTCTGCATGGT GACATGCAGCTGGATATAATTCCGGGGTAAGATTAACGACCTTAT CTGAACATAATGCTA

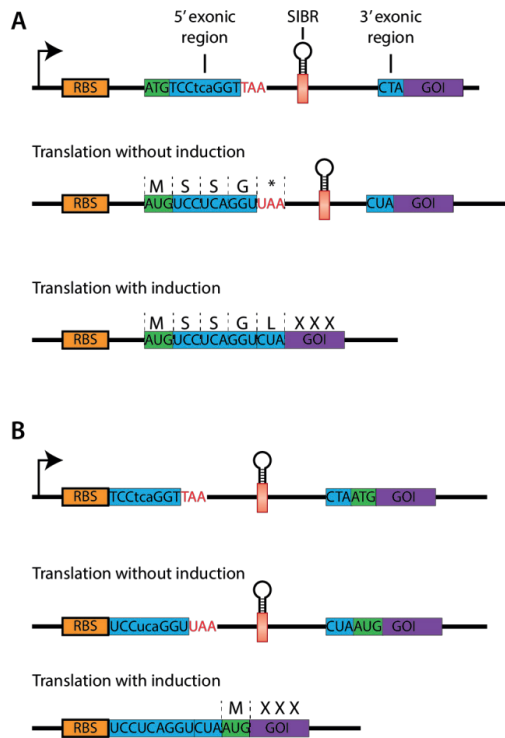
## Transferring SIBR before the ATG start codon to avoid the attachment of N-terminal tags

In **Chapter 6** we introduced SIBR directly after the ATG start codon. This design ensures simplicity and universality as it does not require any complex design and as it is applicable for virtually any GOI. Our design also ensures that the 5' and 3' exonic sequences of SIBR are retained since they are important for its self-splicing efficiency. However, this design adds a 12 nucleotide tag (9 nts from the 5' and 3 nts from the 3' exonic sequences) at the 5' end of the GOI leading to the addition of 4 amino acids at the N-terminus of the translated protein. Whilst a 4 amino acid N-terminal tag is considered very small to impact the functionality of the protein, one cannot exclude potential unpredictable inhibitory effects that an N-terminal tag can have on the functionality of a protein.

As an alternative, SIBR can be introduced directly before the ATG start codon (Fig. 13). According to this design, the presence of the intron causes a physical separation of the ribosome binding site (RBS) and the ATG start codon of the GOI, thereby preventing the translation of the GOI. In contrast, when the intron is excised, the proximity of the RBS and the ATG start codon is shortened, allowing for the translation of the GOI.

Our design was inspired from a natural group I intron discovered in the bacterium *Clostridium difficile* (446,447). This natural group I intron-based riboswitch resides between the RBS and the ATG start codon of the adjacent gene, preventing translation initiation. After induction by intracellular GTP or c-di-GMP, the intron is self-spliced, resulting in appropriate re-positioning of the RBS upstream the start codon, initiating protein translation as a consequence. Although this intron is a natural example of controlling gene expression, its requirement for specific endogenous inducers (GTP and c-di-GMP) as well as its dependency on specific secondary structures (including both the ribozyme and the coding sequence) complicates its general applicability. Therefore, SIBR can be a simple and widely applicable alternative.

Important to note is the “left-over” 12 nt sequence (9 nts from the 5' and 3 nts from the 3' exonic sequences of the intron) between the RBS and the ATG start codon in our design. Currently, we do not know the impact of the “left-over” sequence on protein translation but it can be simply tested by adding different (e.g. the Int1/2/3/4 exonic sequences) 12 nt “left-over” sequences between the RBS and the ATG start codon of a reported protein (e.g. GFP) and test for variations in protein translation.



**Figure 13. Introduction of SIBR after and before the ATG start codon. (A)** SIBR after the ATG start codon. SIBR is introduced directly after the ATG start codon, resulting in a “left-over” 4 amino acid N-terminal tag upon induction. **(B)** SIBR before the ATG start codon. SIBR is introduced directly before the ATG start codon, resulting in a protein without an N-terminal tag sequence upon induction. Note: It is not known how the resulting sequence between the RBS and the ATG start codon will affect protein translation efficiency

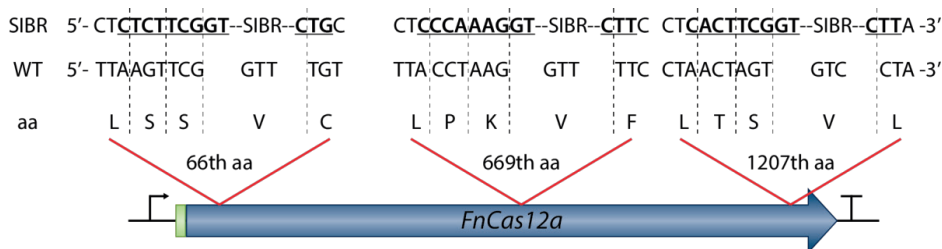
**Transferring SIBR within the ORF of the GOI to avoid the attachment of N-terminal tags**

Another alternative to avoid the “left-over” N-terminal sequence produced by SIBR, is to introduce the intron in the open reading frame (ORF) of the GOI. In fact, the T4 *td* intron used in **Chapter 6**, originally resides in the ORF of the *thymidylate synthase* gene of *E. coli* (496-498). As mentioned in **Chapter 6**, the 5' and 3' exonic sequences interact with the intron and are crucial to preserve the secondary structure of the P1 and P10 stems of the T4 *td* intron. Therefore, transferring the intron along with the flanking exonic regions to the ORF of another gene will disrupt the coding sequence of the target gene. On the other hand, changing the exonic flanking regions of the intron to preserve the coding sequence of the target gene may affect (or completely inhibit) the splicing activity of the intron. Nonetheless, from the data obtained from **Chapter 6**, we now know that the 5' and 3' exonic sequences of the T4 *td* intron are amenable to modifications. This knowledge let us to define the

modifications permitted at the 5' and 3' exonic sequences and consequently allows us to introduce SIBR in the ORF of virtually any GOI.

To introduce the T4 *td* intron in the ORF of the GOI, insertion positions should be identified based on the sequence of the GOI and the parameters defined by **Chapter 6**. Important to note is that the codon frame and the translated amino acid sequence should not be altered as these will lead to a non-sense/non-functional protein. Based on the aforementioned parameters, Dr. Sjoerd C.A. Creutzburg wrote a Python script (named SIBR-ISF) that identifies insertion positions at the GOI and returns a list with the insertion position along with a score that indicates the splicing efficiency of the intron. The script can be found in the Appendix section of **Chapter 7**.

The outcome of SIBR-ISF when the WT *FnCas12a* sequence is used as input, is shown in Appendix Table 1. Twenty two insertion sites have been identified and 109 possible 5' and 3' exonic sequences have been generated. An example of inserting the SIBR in the ORF of *FnCas12a* is shown in Figure 14. For optimal result (i.e. splicing efficiency), the score for splicing should be the highest. Also, to prevent any site/residual activities of the protein of interest, the intron should be introduced closer to the 5' end of the gene of interest to terminate translation as soon as possible.



**Figure 14.** Example of three different SIBR insertion sites in the WT *FnCas12a* gene generated by the SIBR-ISF script. On top, the mutated 5' and 3' exonic sequences (bold and underlined) of SIBR at the insertion site are shown. The wild type (WT) sequence is also shown, indicating the resulting mutations which are necessary for the maintenance of the 5' and 3' exonic sequences. The resulting amino acid sequence after the splicing of SIBR is indicated. The insertion positions are also indicated (e.g. 66<sup>th</sup>, 669<sup>th</sup> and 1207<sup>th</sup> aa).

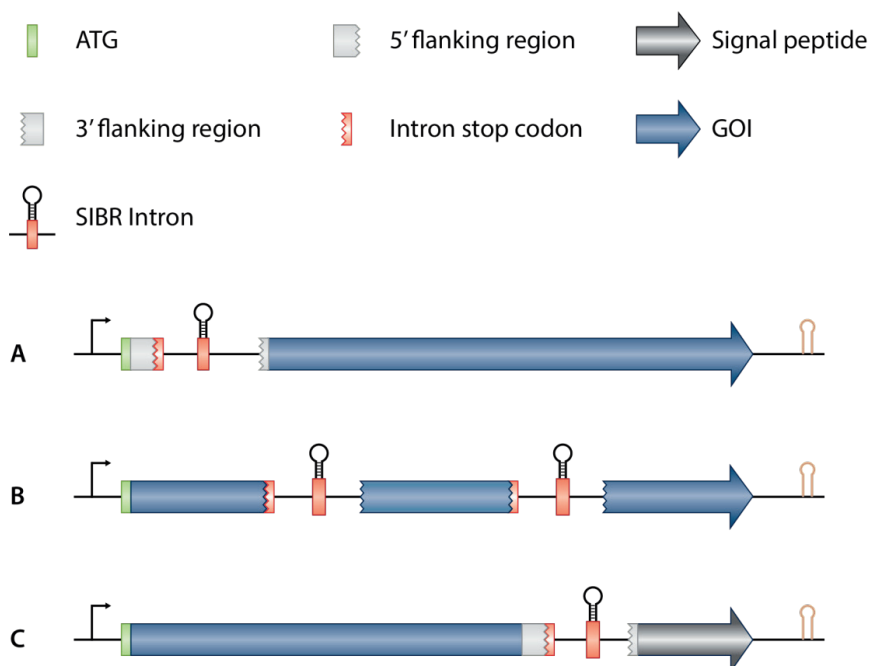
### SIBR-X: a swiss army knife

In **Chapter 6** we demonstrated the use of SIBR as an OFF to ON switch for *FnCas12a*. We also suggested that SIBR can have multiple other uses than just an OFF to ON switch. Examples include ON to OFF switch through the attachment of C-terminal degradation signals and polycistronic

operon control. Here, we further expand the potential applications of SIBR, turning it into a “swiss army knife”.

**Stringent OFF to ON:** Based on the data obtained in **Chapter 6**, SIBR can be used to induce the expression of any GOI in any mesophilic prokaryote of interest. To achieve this, an identical design can be used as the one in **Chapter 6** (i.e. introduction of SIBR directly after the ATG start codon) and as illustrated in Figure 15A. Alternatively, SIBR can be introduced in the ORF of the GOI as discussed above (see “Transferring SIBR within the ORF of the GOI to avoid the attachment of N-terminal tags” section in **Chapter 7**). Moreover, in case where very stringent control of the gene of interest is required, two SIBRs can be introduced into the GOI as illustrated in Figure 15B. This design will avoid any potential leakiness of SIBR since two splicing events should occur for the translation of a functional protein, reducing consequently the chances of leakiness. Such design can be used for toxic proteins (e.g. CRISPR-Cas proteins).

**Translocation:** Inducible translocation of proteins is currently not feasible in prokaryotic and eukaryotic organisms. SIBR can provide an elegant solution to this problem as it can induce the translocation of protein molecules from their production to their destination place. This can be achieved by attaching SIBR in-between the GOI and the signal peptide of interest (Fig. 15C). Based on this design, when the theophylline inducer is absent from the growth medium, complete translation of the GOI will occur until it encounters the stop codon of the SIBR. In contrast, when the theophylline inducer is included in the growth medium, excision of SIBR will occur leading to the attachment of the signal peptide to the protein of interest. The signal peptide can be any peptide specific in transport or localization to the nucleus, mitochondria, chloroplasts, lysosomes, vacuoles, peroxisomes, golgi apparatus, plasma membrane, endoplasmic reticulum, cytoskeleton, outside of the cell (i.e. extracellular) or any other cellular compartment including bacterial microcompartments (BMCs).



**Figure 15. Potential applications of SIBR-X.** (A) Inducible OFF to ON switch as described in Chapter 6. (B) Stringent inducible OFF to ON switch using two SIBR systems introduced in the ORF of the GOI. (C) Inducible translocation of the GOI upon induction and, consequently, the attachments of a C-terminal signal peptide.

## SIBR in eukaryotes

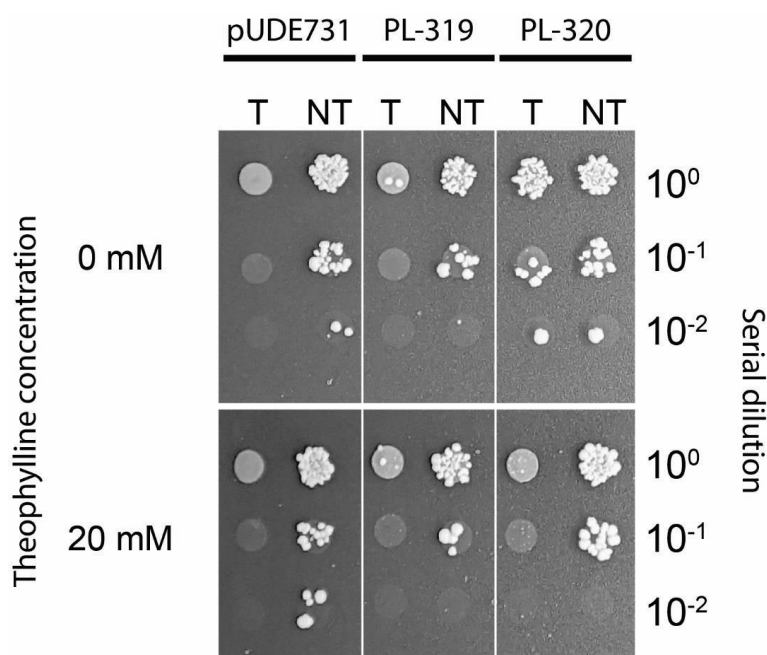
To show the functionality and applicability of SIBR into eukaryotic systems, we transferred SIBR into the eukaryotic model organism *Saccharomyces cerevisiae* (baker's yeast) and controlled the expression of the FnCas12a protein (B. Adiego Perez, C. Patinios *et al*, unpublished). Although *S. cerevisiae* has a very efficient HR system, it has a very inefficient NHEJ system. Therefore, dsDNA cleavage of the *S. cerevisiae* genome by FnCas12a will kill the cell (i.e. counter-selection) when a repair template is not provided, making the screening for induction and functionality of SIBR easy.

To control the expression of FnCas12a, we sought to disrupt the coding sequence of the *FnCas12a* gene through the placement of SIBR directly after the ATG start codon (identical design as the one from **Chapter 6**). However, this approach was not successful (data not shown). We believe that the premature stop codon present in the T4 *td* intron triggers non-sense mediated decay (NMD) as demonstrated in other studies (499-502).

Assuming that NMD was triggered by the premature stop codon of SIBR, we reasoned that changing the position of the intron towards the 3' end of the *FnCas12a* gene, may bypass NMD (503,504). Therefore, SIBR was introduced at the amino acid position 859 of FnCas12a to inhibit the translation of the RuvC I, II and III domains of FnCas12a when SIBR is not spliced out of the transcript. Following the SIBR-ISF script, the 5' exonic sequence of the intron was changed to 5'-AAAGAGTCGGT-3' and the 3' exonic sequence of the intron was changed to 5'-CTT-3'. Two plasmids were constructed which contained the modified T4 *td* intron (either with or without the theophylline aptamer) at the amino acid position 859 of FnCas12a. The FnCas12a + intron gene constructs were expressed by the constitutive TEF1 promoter. Those plasmids were named PL-319 (FnCas12a + intron without aptamer) and PL-320 (FnCas12a + intron with aptamer). pUDE731 was used as a positive control for targeting (constitutive expression of FnCas12a; Addgene plasmid # 103008).

PL-319, PL-320 or pUDE731 were co-transformed in the yeast *S. cerevisiae* with either a plasmid containing a non-targeting spacer (PL-207) or a plasmid containing the *ADE2* targeting spacer (PL-074). The LiAc/SS carrier DNA/PEG method by Gietz and Schiestl (2007) was used to transform the yeast cells using 500 ng of each plasmid (505). After transformation, the transformed yeast cells were recovered in YPD medium for 3 h at 30°C and then serially diluted in PBS (pH 7.4). 3 µl were used for spot assays on drop-out (omitting uracil; for the selection of PL-319, PL-320 or pUDE731) minimal agar medium (1.7 g L<sup>-1</sup> bacto-yeast nitrogen base w/o amino acids and w/o ammonium sulfate; 1 g L<sup>-1</sup> monosodium glutamate; 20 g L<sup>-1</sup> glucose; 20 g L<sup>-1</sup> agar) containing 200 µg mL<sup>-1</sup> geneticin (for the selection of PL-074 or PL-207) in the presence or absence of 20 mM theophylline.

The results of this experiment are depicted at Figure 16. *S. cerevisiae* cells co-transformed with PL-319, PL-320 or pUDE731 plasmids and the PL-207 non-targeting plasmid showed colony formation up to the  $10^{-2}$  dilution regardless of the presence or absence of theophylline. As expected, when the pUDE731 was co-transformed with the PL-074 targeting plasmid, almost no colonies were formed regardless of the presence or absence of theophylline. Similarly, when PL-319 was co-transformed with the PL-074 targeting plasmid, reduced number of colonies were formed regardless of the presence or absence of theophylline. This indicated that the intron is able to self-splice out of the formed FnCas12a mRNA and code for a functional FnCas12a that is able to target and cleave the target site. In contrast, when PL-320 was co-transformed with the PL-074 targeting plasmid, normal size colonies were formed only when theophylline was omitted from the solid agar medium. However, when theophylline was included in the solid agar medium, a reduced number of colonies was observed. This result indicated that inducible excision of SIBR can be achieved in eukaryotes such as *S. cerevisiae*.



**Figure 16. SIBR-Cas is functional in the yeast *S. cerevisiae*.** pUDE731, PL-319 or PL-320 were co-transformed either with a plasmid containing a non-targeting (NT) or a targeting (T) crRNA. Transformants were serially diluted ( $10^0$ ,  $10^{-1}$ ,  $10^{-2}$ ) and plated on selective medium in the presence or absence of the theophylline inducer.

Hereby, we demonstrated that SIBR is transferable and functionable in the eukaryotic yeast *S. cerevisiae*. We anticipate that SIBR will also function in other mesophilic yeasts but also in other



mesophilic eukaryotic systems (e.g. human cells). Applications of SIBR in eukaryotes may include inducible genome editing, inducible protein expression or degradation and inducible localization.

### Turning every Group I intron to SIBR

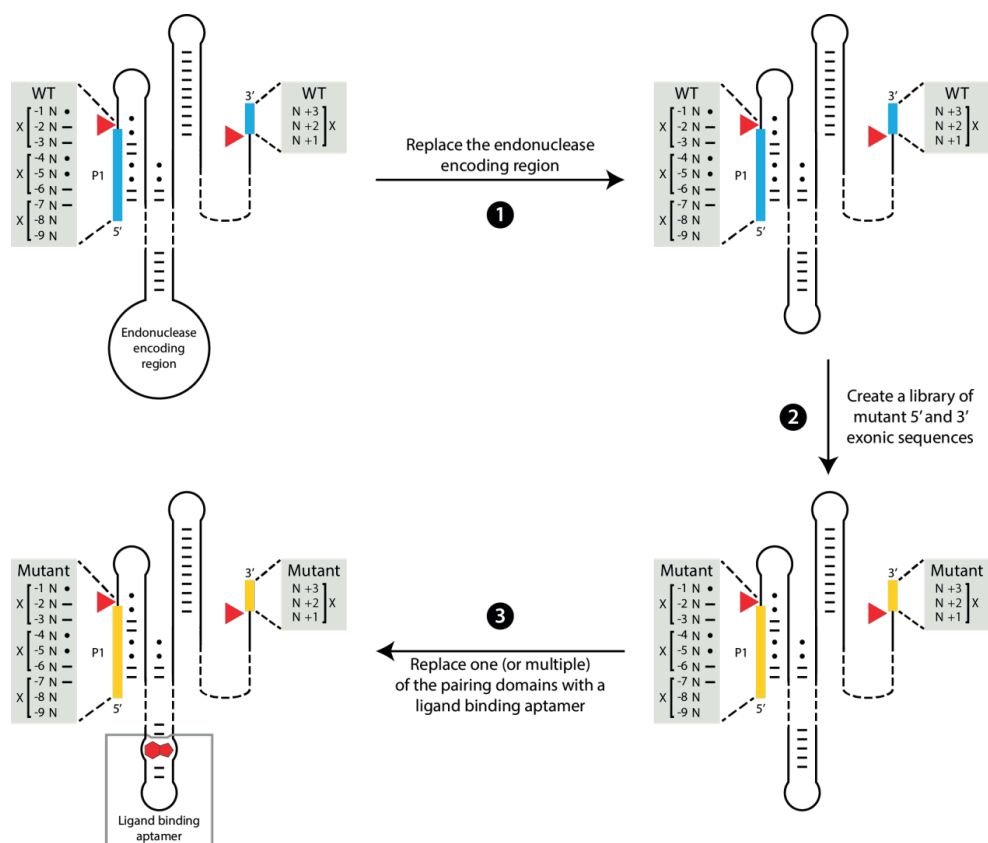
Group I introns, like the T4 *td* intron, form core secondary structures consisting of multiple paired regions. In principle, to turn any Group I intron into a Self-splicing Intron Based Riboswitch (SIBR), a stepwise approach can be followed, similar to the one described in **Chapter 6** and illustrated in Figure 17.

Briefly, any intron-encoded genes (e.g. endonucleases) should be removed from the intron sequence to avoid any unwanted integration events. Following, since the 5' and 3' exonic sequences of Group I introns interact with the intron sequence and affect the secondary and tertiary structure of the intron, a library of mutant 5' and 3' exonic sequences should be developed. This mutant library will serve as the basis to define the effect of the 5' and 3' exonic sequences on the splicing efficiency of the intron. Moreover, this library will likely contain introns with a range (low to high) of splicing efficiency. It is very likely that in the mutant intron library, introns with better splicing efficiency than the wild type intron will be generated. Also, the intron library will allow the transfer of the intron of interest to the open reading frame of any GOI without disturbing the amino acid sequence of the target protein. To test the intron library, an *in vivo* assay similar to **Chapter 6** can be performed.

Next, to achieve inducible control over the splicing of the intron, an aptamer moiety which responds to specific small molecules (e.g. theophylline) shall be introduced at one or multiple pairing (P) domains of the intron. For example, as described by Thompson *et al.*, 2002, and also applied in **Chapter 6**, the theophylline aptamer was introduced at the P6 domain of the T4 *td* intron, turning it into an inducible self-splicing gene regulator (448). Another example is described by Kertsburg and Soukup, 2002, where they turned the *Tetrahymena* group I intron into an inducible self-splicing intron by replacing the P6 or P8 or both P6 and P8 domains with a theophylline aptamer (506). Similar approaches (to that of Thompson *et al.*, 2002 and that of Kertsburg and Soukup, 2002) can be taken for any other Group I intron where one of their P domains is altered to contain an aptamer moiety that responds to specific small molecules and can consequently control the splicing of the intron.

After generating the mutant intron library (mutations at the 5' and 3' exonic sequences) and achieving inducible control over the splicing of the intron (through the introduction of an aptamer

in one of the P domains of the intron), the generated intron variants can be moved directly after or before the ATG start codon of the GOI. Alternatively, the intron can be inserted within the ORF of the GOI. To achieve insertion within the ORF, a script similar to SIBR-ISF has to be developed. The script shall consider the self-splicing efficiency and the constraints defined for the intron mutants.



**Figure 17. Step-wise approach to turn every Group I intron into SIBR.** 1) The endonuclease encoding region (or any other intron-encoded genes) should be replaced. 2) A library of mutant 5' and 3' exonic sequences shall be generated to assess the amenability of the exonic sequences to modifications. 3) One or multiple pairing domains should be replaced with a ligand binding aptamer to make the self-splicing of the intron inducible. Red triangles indicate the splicing sites. Blue rectangles indicate the WT exonic sequences. Yellow rectangles indicate the mutant exonic sequences.

## Turning group II introns to SIBR

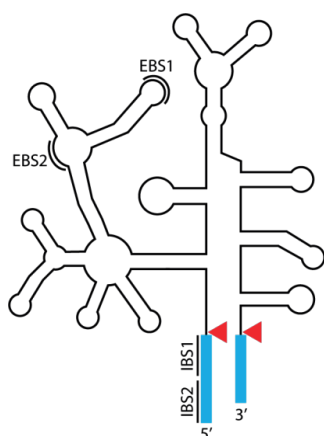
Group II introns are ribozymes that occur in higher eukaryotes (plants), lower eukaryotes (fungi and protists) as well as in bacteria and their phages (507-509). Similar to Group I introns, group II introns reside in genes (separating the coding region into 5' and 3' exons) which upon transcription and excision from the gene's mRNA transcript results in restoration of the gene's reading frame and hence the formation of a functional protein. Group II introns can self-splice (509-512), although in some cases intron-encoded proteins (IEPs) may facilitate splicing by stabilizing the intron RNA structure (513). To achieve splicing, the 5' and 3' exonic sequences of the Group II introns (called intron-binding site or IBS), interact with conserved domains of the intron (called exon-binding site or EBS) to form long-range tertiary interactions (508,514) (Fig. 18). The intron-exon interactions are necessary for splicing as they bring the intron at the active site of the exons in order to facilitate the typical transesterification reaction that mediates the excision of the intron. The necessity of intron-exon interactions for splicing, implies a limitation for transferring any group II intron into any GOI other than the WT sequence where it resides into, as the exon sequences need to be conserved. To overcome this, a similar approach as the one developed in **Chapter 6** for the T4 *td* Group I intron can be used.

First, any intron-encoded genes (e.g. endonucleases) should be removed from the Group II intron to avoid any unwanted events. Then, a mutant library of Group II introns can be generated in which the exon sequences (IBS1 and IBS2 for 5' exon and IBS3 for 3' exon) are mutated. Since the IBS and EBS sites are necessary for splicing, when an IBS site is mutated, the EBS site shall be mutated as well to facilitate base pairing. The generated library can then be assessed for the efficiency of the self-splicing activity of the intron, by following a similar approach as that for LacZ as described in **Chapter 6**. Important to note is that the self-splicing efficiency can be assessed by any other *in vitro* or *in vivo* methods (other than LacZ) as long as it can distinguish the formation of spliced from un-spliced products, or the formation (qualitatively and quantitatively) of active from inactive proteins.

In the case that the Group II intron mutant library assay will be through a protein assay (similar to that of LacZ; **Chapter 6**) then, for convenience and to maintain the coding sequence of the protein, the Group II intron can be transferred directly after the ATG start codon in order to maintain the coding sequence of the protein. This approach was described in **Chapter 6** for FnCas12a. The outcome of the Group II intron mutant library assay is expected to yield a range with good and bad splicing introns which can then be used to modulate/tune the expression of the gene/RNA/protein of interest.

After establishing the requirements for splicing as defined by the IBS-EBS interactions, a script similar to SIBR-ISF can be developed that allows for transferring the mutant Group II intron to virtually any gene of interest. Next, an aptamer moiety which responds to specific small molecules (e.g. theophylline) should be introduced at one or multiple pairing (P) domains of the intron. To achieve this, the approaches developed and applied by Thompson *et al.* (2002) and Kertsburg and Soukup (2002) can be used (448,506).

After generating the mutant Group II intron library (mutations at the 5' and 3' exonic sequences) and achieving inducible control over the splicing of the Group II intron (through the introduction of an aptamer in one of the P domains of the intron), the generated Group II intron variants can be moved at or 5' of the start of the polynucleotide portion encoding the POI, or within the polynucleotide portion encoding the POI.



**Figure 18. Cartoon example of a self-splicing Group II intron.** The intron-binding sites (IBS) 1 and 2, interact with the exon-binding sites (EBS) 1 and 2, respectively, to achieve splicing. The exonic sequences are indicated with blue rectangles and the splicing sites are indicated with red triangles.

## Are BMCs the appropriate metabolic medium?

Bacterial microcompartments (BMCs) have triggered the interest of many international groups due to their prevalence in bacterial species and their potential use for biotechnological applications. Several great reviews summarize the advances in deciphering the physiology and role of BMCs and their involvement in bacterial survival, growth and infection but also in the potential of BMCs as medium for metabolic processes (358,359,363,364,515-517). Still, many fundamental questions remain unanswered.

The assembly mechanism of BMCs and the encapsulation of enzymes in the lumen of BMCs remain unknown. Two assembly theories are known where either the core enzymes of the BMC are assembled first followed by the assembly of the shell proteins around the core enzymes (core-first assembly) or, the core enzymes and the shell proteins are assembled concomitantly (concomitant assembly) (359). To date, none of the two theories prevails but the observation of aggregates formed by the core enzymes may favor the core-first assembly (518). Also, several (but not all) core enzymes, contain N-terminal sequences (termed as encapsulation peptides; EPs) which mediate their encapsulation in the lumen of the BMCs (369,370,388,390,395,396). The absence of distinct N-terminal EPs to some of the core enzymes may indicate that some of the core enzymes are “piggybacking” their encapsulation through interactions with other core enzymes that contain EP. Such interactions are not described yet but state-of-the-art protein-protein interaction prediction software (e.g. AlphaFold2 and Rosetta) may shed some light on those interactions (391,479). Moreover, the encapsulation efficiency and the number of encapsulated protein molecules is still not very well known. Such information is critical in defining the suitability of an EP to direct protein encapsulation in the lumen of the BMC.

One of the most urging mysteries is whether and how the pores in the center of the BMC shell proteins facilitate (selective) compound transfer inside and outside the BMCs. Several computational studies suggest that some pores are more specific for certain compounds than others (360,361,519,520). However, experimental results on the hypothesized BMC pore specificity are entirely lacking or very indirect (i.e. inferred from *in vivo* experiments) (360,521,522). Therefore, future studies should focus on unravelling (experimentally) the selectivity of the BMC pores at the single pore level; something that will allow one to tune the efficiency and selectivity of the BMC metabolism.

The formation of BMCs is usually coupled to the presence of certain metabolites (e.g. ethanolamine or 1,2-propanediol). Therefore, to convert BMCs into universal media for metabolic processes (i.e. have any metabolism of interest in the lumen of BMCs), the formation of BMCs should be

uncoupled from the inducing metabolites. Such approach requires either the complete deregulation of the endogenous BMC operon through the change of the metabolite induced promoter sequence, or the transfer of the entire BMC operon in a non-native host and the expression of the BMC operon under a metabolite-independent promoter. Both approaches are complicated to pursue since the regulation of BMCs is usually associated to survival and growth and the BMC operon is regularly large (~ 20 kb). Minimal BMCs have been assembled but their encapsulation capacity is not known (523,524). Therefore, changing the expression, the metabolism and the size of the BMCs remain a challenge for repurposing BMCs as the appropriate medium for engineered metabolisms.

In summary, although the studies on BMCs have been intensified the last decade, many fundamental and practical questions remain unanswered. The complete re-purposing of BMCs into a universal metabolic medium is not a straightforward approach and the *in vivo* engineering of BMCs may present several metabolic burdens. Nevertheless, BMCs can be used in various other *ex vivo* applications (516).

## Concluding remarks

Overall, in line with spectacular developments in biotechnology (microbial genomics, genome editing and bioprocess engineering), this thesis describes further advancements to enable microbial ester production through the engineering of enzymes, pathways and genomes. As described in **Chapter 2**, esters are important constituents of the modern industry (food and beverage, chemical and pharma). Unfortunately, the production of esters is heavily relied on unsustainable methods, providing (as a consequence) an interesting opportunity to explore and develop more sustainable production methods. Microbial ester production is an interesting alternative to produce esters in a sustainable manner. The advancements for microbial ester production are reviewed in **Chapter 2** and, the opportunities but also the limitations of microbially produced esters are indicated. Since esters vary in complexity, market size and production method, it is clear that a technoeconomic and life-cycle analysis should be performed to assess the feasibility for the successful commercialization of microbially produced esters. Governmental subsidies are (or should be) an important incentive to drive innovations in microbial ester production and should follow an ester product from its conception, all the way to its commercialization.

The first aim of this thesis was to further characterize the newly discovered Eat1 enzyme. In **chapter 3**, we discovered that Eat1 can perform alcoholysis and thiolysis in addition to the AAT, esterase and thioesterase activities. Alcoholysis is the preferred activity of Eat1, making it relevant

when using Eat1 as the appropriate enzyme to produce esters in microbial hosts. This is more apparent when the ester product is retained in the fermentation medium and can serve as a substrate for the alcoholysis activity of Eat1. Therefore, the removal of the ester product(s) from the fermentation medium is essential to achieve maximum titers when Eat1 is used as the catalyst. Also, yeasts that contain and express Eat1 homologs, may affect the quantity and quality of esters produced by yeasts, due to the exchange of the alcohol moiety between endogenously produced esters and alcohols. Therefore, the alcoholysis activity of Eat1 should be investigated further to elucidate its role in the quality of the final fermented product.

The second aim of this thesis was to engineer the industrially relevant *C. beijerinckii* NCIMB 8052 to produce ester compounds. An alternative approach was used in **Chapter 4**, to bring AAT enzymes and substrates in close proximity in a confined space. To do this, we used the BMCs of *C. beijerinckii* NCIMB 8052 as an appropriate medium and attempted to re-purpose them to produce propyl propionate as a proof of principle. Unfortunately, this approach appeared to be cumbersome. Nonetheless, we have set-up the foundations of our vision through the encapsulation of GFP in the lumen of the *C. beijerinckii* NCIMB 8052 BMCs. It is obvious that fundamental knowledge for the successful encapsulation of enzymes in the lumen of BMCs is lacking and future studies should focus on understanding the biochemical and biophysical attributes of the shell and core proteins of BMCs. Such knowledge will expand the understanding of the existence and evolution of BMCs, the physiology of the bacteria that produce BMCs and the applicability of BMCs for metabolic engineering.

The third and final aim of this thesis was to create simple and efficient CRISPR-Cas-based genome engineering tools in bacteria. To this end, in **Chapter 5** we developed the first multiplex CRISPR-Cas12a genome engineering tool for *C. beijerinckii*. The efficiency of our tool for single gene knock-outs was high (up to 100%) but the efficiency for the simultaneous knock-out of two genes was 18%. Despite the low editing efficiency for multiplex gene knock-out, our tool can generate double knock-outs in a relatively short period of time. In **Chapter 6** we advanced even further the CRISPR-Cas genome engineering toolbox for bacteria and developed the widely applicable SIBR-Cas tool. SIBR-Cas uses an elegant gene expression control mechanism which is based on the T4 *td* intron. This attribute makes it easy to apply and transfer SIBR-Cas in virtually any mesophilic prokaryote of interest. We tested the functionality of SIBR-Cas in several wild type prokaryotes and showed the simplicity and efficiency of SIBR-Cas. We foresee that SIBR-Cas will be the “first-to-try” genome engineering approach in wild type prokaryotes.

To conclude, the developments reported in this thesis illustrate the potential and complexity of microbially produced esters. Reflection is given on relevant ester-producing enzymes and ester-producing hosts, with an emphasis on the recently discovered Eat1 enzyme and the industrially relevant *C. beijerinckii*. Moreover, the CRISPR-Cas tools developed in this thesis are expected to simplify and accelerate the genome engineering of wild type mesophilic prokaryotes.



## Appendix

### SIBR Insertion Site Finder (SIBR-ISF): Python script to find insertion sites for SIBR in the gene of interest

Note: this script was developed by Dr. Sjoerd C.A. Creutzburg. The author of this thesis does not claim authorship of the script. Also, to run the script, additional files are needed (Codon.csv, NEB RE.csv and TdScoreTable.csv). The files can be provided upon request.

```
def findbindingtype(Q, S):
    a = "M"

    if Q == "T" and S == "A": a = "P"
    if Q == "T" and S == "G": a = "W"
    if Q == "C" and S == "G": a = "P"
    if Q == "A" and S == "T": a = "P"
    if Q == "G" and S == "C": a = "P"
    if Q == "G" and S == "T": a = "W"
    if Q == "N" or S == "N": a = "X"

    return a

def permutatelist(X):
    Y = [""]
    for i in X:
        Z = Y
        Y = []
        for j in i:
            for l in range(len(Z)):
                Y.append(Z[l]+j)
    return Y

def matchesense(q, s):
    a = 100.000
    if q == s: a = 1.000
    if q == "R" and (s == "A" or s == "G"): a = 1.000
    if q == "Y" and (s == "C" or s == "T"): a = 1.000
    if q == "S" and (s == "G" or s == "C"): a = 1.000
    if q == "W" and (s == "A" or s == "T"): a = 1.000
    if q == "K" and (s == "G" or s == "T"): a = 1.000
    if q == "M" and (s == "A" or s == "C"): a = 1.000
    if q == "B" and s != "A": a = 1.000
    if q == "D" and s != "C": a = 1.000
    if q == "H" and s != "G": a = 1.000
    if q == "V" and s != "T": a = 1.000
    if q == "N": a = 1.000
    return a

def RevComp(sequence):
    RC = ""
```

```

for i in sequence:
    j = ""

    if i == "T": j = "A"
    if i == "C": j = "G"
    if i == "A": j = "T"
    if i == "G": j = "C"

    if i == "Y": j = "R"
    if i == "W": j = "W"
    if i == "K": j = "M"
    if i == "M": j = "K"
    if i == "S": j = "S"
    if i == "R": j = "Y"

    if i == "H": j = "G"
    if i == "B": j = "A"
    if i == "D": j = "C"
    if i == "V": j = "T"

    if i == "N": j = "N"

    RC = j + RC

return RC

#----- Input Subject -----

SubDNA = []

FN = raw_input('DNA File name = ')
with open(FN, 'r+') as f:
    for r in f:
        for c in r:
            if c.upper() == "G" or c.upper() == "A" or c.upper() == "T" or c.upper() == "C":
                SubDNA.append(c.upper())

#----- Input Query -----

#      -8 -7 -6 -5 -4 -3 -2 -1  1  2  3  4
QueDNA = ["N","T","G","A","G","T","C","C","G","G","A","G","T"] #Native

QueDNA = ["N","N","G","A","G","T","C","C","G","G","A","G"] #Simplified

#"Query Functions:
#(P)air, (W)obbly, (N)onbinding")
#(F)ixed (P)air, (F)ixed (W)obbly, (F)ixed (N)onbinding

#----- Import Codon Table -----

AA = []

```

## Chapter 7

```
Codon = []

import csv
with open('Codon.csv') as csvfile:
    f = csv.reader(csvfile, delimiter=',', quotechar='"')
    for r in f:
        Codon.append(r[0])
        AA.append(r[1])

#----- Translate Subject -----

SubPro = []

for i in range(int(len(SubDNA)/3)):
    qcodon = SubDNA[3*i]+SubDNA[3*i+1]+SubDNA[3*i+2]
    for j in range(64):
        if Codon[j] == qcodon:
            SubPro.append(AA[j])

#----- Define Peptide -----

TBS_DNA = []
TBS_Pos = []

for s in range(len(SubPro)-int((len(QueDNA)+4)/3)-1):           #Subject Protein Start
    Peptide = SubPro[s:int((len(QueDNA)+4)/3)]
#----- Reverse Translate -----
    mRNA = []
    for i in Peptide:
        Cdn = []
        for j in range(64):
            if AA[j] == i:
                Cdn.append(Codon[j])
        mRNA.append(Cdn)
    RevTrans = permutatelist(mRNA)
    TBS_DNA = TBS_DNA + RevTrans
    for k in RevTrans:
        TBS_Pos.append(s)

#----- Test binding type -----

print "Binding Type analysis"

ScoreTable_BP = []
ScoreTable_Score = []

import csv
with open('TdScoreTable.csv') as csvfile:
    f = csv.reader(csvfile, delimiter=',', quotechar='"')
    for r in f:
        ScoreTable_BP.append(r[0])
        ScoreTable_Score.append(r[1])
```

```

DNA_List = []
Pos_List = []
Score_List = []
Pep_List = []
Frame_List = []

Query = ["N","N","N"] + QueDNA + ["N","N","N"]
for i in range(len(TBS_DNA)):
    Subject = TBS_DNA[i]
    for f in range(3):
        BP = ""
        for j in range(len(QueDNA)):
            BP = BP + findbindingtype(QueDNA[j],Subject[j+f])

        Score = 0
        for k in range(len(ScoreTable_BP)):
            if BP == ScoreTable_BP[k]:
                Score = ScoreTable_Score[k]
                DNA_List.append(TBS_DNA[i])
                Pos_List.append(TBS_Pos[i]+1)
                Score_List.append(Score)
                Frame_List.append(f+1)
                Peptide = ""
                for l in SubPro[TBS_Pos[i]:TBS_Pos[i]+int((len(QueDNA)+4)/3)]:
                    Peptide = Peptide + l
                Pep_List.append(Peptide)

#----- Select 5 best -----

Result = []
for row in range(len(Pos_List)):
    Result.append([Pos_List[row], DNA_List[row], Pep_List[row], Score_List[row],
Frame_List[row]])

Result2 = sorted(sorted(Result, key=lambda A: A[3], reverse=True), key=lambda A: A[0])

A = Result2[0][0]
count = 0
Result3 = []

for r in range(len(Result2)):
    if Result2[r][0] == A and count <= 5:
        Result3.append(Result2[r])
        count = count + 1
    if Result2[r][0] != A:
        A = Result2[r][0]
        count = 1

```

## Chapter 7

```
print Result3
```

```
#----- Export to CSV -----
```

```
print "Writing intron sites to file"
```

```
import csv
```

```
Writer = csv.writer(open(FN+".csv", 'wb'), delimiter=',')
```

```
Writer.writerow(["Position", "DNA", "Protein", "Score", "Frame"])
```

```
with open(FN+".csv", 'ab') as F:
```

```
    for row in range(len(Result3)):
```

```
        Writer = csv.writer(F, delimiter=',')
```

```
        Writer.writerow(Result3[row])
```

```
#-----Find the Restriction Enzymes-----
```

```
#-----Import RE Sequences-----
```

```
PreREName = []
```

```
PreRESeq = []
```

```
import csv
```

```
with open('NEB RE.csv') as csvfile:
```

```
    f = csv.reader(csvfile, delimiter=',', quotechar='"')
```

```
    for r in f:
```

```
        PreREName.append(r[0])
```

```
        PreRESeq.append(r[1])
```

```
#-----Find RE in DNA Sequence-----
```

```
print("Analysing current Restriction endonulease sites")
```

```
REName = []
```

```
RESeq = []
```

```
REPresent = []
```

```
PreREPresent = []
```

```
for e in range(len(PreRESeq)):
```

```
    QueDNA = []
```

```
    s = 0
```

```
    for c in PreRESeq[e]:
```

```
        QueDNA.append(c)
```

```
    for i in range(len(SubDNA)-len(QueDNA)+1):
```

```
        score = 0
```

```
        for j in range(len(QueDNA)):
```

```
            score = score + matchsense(QueDNA[j],SubDNA[i+j])
```

```
        if score == len(QueDNA):
```

```

    if s == 0:
        Result = str(int((i+3)/3))
    if s > 0 and s <= 3:
        Result = Result + ", " + str(int((i+3)/3))
    s = s + 1
if s > 0:
    PreREPresent.append(Result)
if s == 0:
    PreREPresent.append("N/A")

if s <= 3:
    REPresent.append(PreREPresent[e])
    REName.append(PreREName[e])
    RESeq.append(PreRESeq[e])

#-----Find RE is protein Sequence-----

ResultName = []
ResultPos = []
ResultDNA = []
ResultProtein = []
ResultFrame = []
ResultOtherRE = []

import csv
Writer = csv.writer(open(FN+"-RE.csv", 'wb'), delimiter=',')
Writer.writerow(["Enzyme", "Position", "DNA Seq", "AA Seq", "Frame", "RE sites present"])

print("Analysing potential Restriction endonuclease sites")

for e in range(len(RESeq)):
    QueDNA = ["N", "N", "N"]
    for c in RESeq[e]:
        QueDNA.append(c)
    QueDNA = QueDNA+["N"]+["N"]+["N"]+["N"]

    for f in range(3):
        for i in range(len(SubPro)-int((len(QueDNA)-3)/3)+1): #Subject amino acid start
            currentDNA = ""
            currentProtein = ""
            currentScore = 0
            m = 0 #Maxscore counter
            for j in range(int((len(QueDNA)-3)/3)): #Subject amino acid
                SubAA = SubPro[i+j]
                maxscore = 0
                for k in range(64): #64 codons/AA
                    if SubAA == AA[k]:
                        score = 0
                        QueCodon = Codon[k]
                        for l in range(3):
                            score = score + matchesense(QueDNA[3*j+3+l-f], QueCodon[l])

```

## Chapter 7

```
        if score<100 and score>maxscore:
            maxscore = score
            maxcodon = k
    if maxscore>0:
        m = m + 1
        currentDNA = currentDNA + Codon[maxcodon]
        currentProtein = currentProtein + AA[maxcodon]
        currentScore = currentScore + maxscore

    if m == int((len(QueDNA)-3)/3):
        ResultName.append(REName[e])
        ResultPos.append(i+1)
        ResultDNA.append(currentDNA)
        ResultProtein.append(currentProtein)
        ResultFrame.append(f+1)
        ResultOtherRE.append(REPresent[e])

print("Writing RE sites to file")

import csv
with open(FN+"-RE.csv", 'ab') as F:
    for i in range(len(ResultName)):
        Writer = csv.writer(F, delimiter=',')
        Writer.writerow([ResultName[i], ResultPos[i], ResultDNA[i], ResultProtein[i],
ResultFrame[i], ResultOtherRE[i]])

print "Done"
```

**Appendix table 1. Outcome of SIBR-ISF using the WT sequence of *FnCas12a*. Position indicates the insertion site at the amino acid level. DNA indicates the resulting 5' and 3' exonic sequences after splicing. Protein indicates the resulting amino acid sequence after splicing of the intron. Score indicates the splicing efficiency of the intron (higher is better). Frame indicates whether the intron is introduced in frame (1) or shifted by one nt (2) or by two nts (3) in relation to the codons of the gene of interest.**

Position	DNA (5' to 3')	Protein	Score	Frame
33	GCCAGAGGTCTAATA	ARGLI	0.767369	1
	GCCCCGAGGTCTAATA	ARGLI	0.767369	1
	GCCAGAGGTCTGATA	ARGLI	0.767369	1
	GCCCCGAGGTCTGATA	ARGLI	0.767369	1
	GCCAGAGGTCTAATC	ARGLI	0.767369	1
	GCCCCGAGGTCTAATC	ARGLI	0.767369	1
66	CTCTCTTCGGTCTGC	LSSVC	1.299541	3
	CTGTCTTCGGTCTGC	LSSVC	1.299541	3
	CTTTCTTCGGTCTGC	LSSVC	1.299541	3
	TTATCTTCGGTCTGC	LSSVC	1.299541	3
	TTGTCTTCGGTCTGC	LSSVC	1.299541	3
254	AATCAACGGGTCTTC	NQRVF	0.303609	3
	AACCAGCGGGTCTTC	NQRVF	0.303609	3
	AATCAGCGGGTCTTC	NQRVF	0.303609	3
	AACCAACGGGTCTTT	NQRVF	0.303609	3
	AATCAACGGGTCTTT	NQRVF	0.303609	3
260	CTCGACGAGGTCTTC	LDEVF	0.368337	3
	CTGGACGAGGTCTTC	LDEVF	0.368337	3
	CTTGACGAGGTCTTC	LDEVF	0.368337	3
	TTAGACGAGGTCTTC	LDEVF	0.368337	3
	TTGGACGAGGTCTTC	LDEVF	0.368337	3
320	AAGATGTCGGTCTTA	KMSVL	0.572077	3
	AAAATGTCGGTCTTG	KMSVL	0.572077	3
	AAGATGTCGGTCTTG	KMSVL	0.572077	3
403	AGTCAACAGGTCTTC	SQQVF	0.257101	3
	TCACAACAGGTCTTC	SQQVF	0.257101	3
	TCCCAACAGGTCTTC	SQQVF	0.257101	3
	TCGCAACAGGTCTTC	SQQVF	0.257101	3
	TCTCAACAGGTCTTC	SQQVF	0.257101	3
414	GGCACTGCGGTCTTA	GTAVL	0.872721	3
	GGGACTGCGGTCTTA	GTAVL	0.872721	3
	GGTACTGCGGTCTTA	GTAVL	0.872721	3
	GGAAGTACGGGTCTTG	GTAVL	0.872721	3
	GGCACTGCGGTCTTG	GTAVL	0.872721	3



# Chapter 7

561	TTTTATCTGGTCTTC	FYLVF	0.660894	3
	TTCTATCTGGTCTTT	FYLVF	0.660894	3
	TTTTATCTGGTCTTT	FYLVF	0.660894	3
	TTCTATTTGGTCTTC	FYLVF	0.62378	3
	TTTTATTTGGTCTTC	FYLVF	0.62378	3
669	CTCCCAAAGGTCTTC	LPKVF	0.613895	3
	CTGCCAAAGGTCTTC	LPKVF	0.613895	3
	CTTCCAAAGGTCTTC	LPKVF	0.613895	3
	TTACCAAAGGTCTTC	LPKVF	0.613895	3
	TTGCCAAAGGTCTTC	LPKVF	0.613895	3
699	AAGAATGGGTCTCCA	KNGSP	0.324639	2
	AAAAATGGGTCTCCC	KNGSP	0.324639	2
	AAGAATGGGTCTCCC	KNGSP	0.324639	2
	AAAAATGGGTCTCCG	KNGSP	0.324639	2
	AAGAATGGGTCTCCG	KNGSP	0.324639	2
818	CAGGATGTGGTCTAC	QDVVY	0.654541	3
	CAAGATGTGGTCTAT	QDVVY	0.654541	3
	CAGGATGTGGTCTAT	QDVVY	0.654541	3
	CAAGACGTGGTCTAC	QDVVY	0.427474	3
	CAGGACGTGGTCTAC	QDVVY	0.427474	3
859	AAGGAATCGGTCTTC	KESVF	0.429057	3
	AAAGAGTCGGTCTTC	KESVF	0.429057	3
	AAGGAGTCGGTCTTC	KESVF	0.429057	3
	AAAGAATCGGTCTTT	KESVF	0.429057	3
	AAGGAATCGGTCTTT	KESVF	0.429057	3
1001	GCCATTGTGGTCTTC	AIVVF	0.698177	3
	GCGATTGTGGTCTTC	AIVVF	0.698177	3
	GCTATTGTGGTCTTC	AIVVF	0.698177	3
	GCAATTGTGGTCTTT	AIVVF	0.698177	3
	GCCATTGTGGTCTTT	AIVVF	0.698177	3
1020	GAGAAACAGGTCTAC	EKQVY	0.321376	3
	GAAAAGCAGGTCTAC	EKQVY	0.321376	3
	GAGAAGCAGGTCTAC	EKQVY	0.321376	3
	GAAAAACAGGTCTAT	EKQVY	0.321376	3
	GAGAAACAGGTCTAT	EKQVY	0.321376	3
1036	AATTATCTGGTCTTC	NYLVF	0.660894	3
	AACTATCTGGTCTTT	NYLVF	0.660894	3
	AATTATCTGGTCTTT	NYLVF	0.660894	3
	AACTATTTGGTCTTC	NYLVF	0.62378	3
	AATTATTTGGTCTTC	NYLVF	0.62378	3

1048	ACCGGTGGGGTCTTA	TGGVL	0.305534	3
	ACGGGTGGGGTCTTA	TGGVL	0.305534	3
	ACTGGTGGGGTCTTA	TGGVL	0.305534	3
	ACAGGTGGGGTCTTG	TGGVL	0.305534	3
	ACCGGTGGGGTCTTG	TGGVL	0.305534	3
1099	TATGAATCGGTCTCA	YESVS	0.321793	3
	TACGAGTCGGTCTCA	YESVS	0.321793	3
	TATGAGTCGGTCTCA	YESVS	0.321793	3
	TACGAATCGGTCTCC	YESVS	0.321793	3
	TATGAATCGGTCTCC	YESVS	0.321793	3
1146	AGTTTGGGTCTAGA	SFGSR	0.8	2
	TCATTGGGTCTAGA	SFGSR	0.8	2
	TCCTTGGGTCTAGA	SFGSR	0.8	2
	TCGTTGGGTCTAGA	SFGSR	0.8	2
	TCTTTGGGTCTAGA	SFGSR	0.8	2
1165	ACCAGAGAGGTCTAC	TREVY	0.460422	3
	ACGAGAGAGGTCTAC	TREVY	0.460422	3
	ACTAGAGAGGTCTAC	TREVY	0.460422	3
	ACAAGGGAGGTCTAC	TREVY	0.460422	3
	ACCAGGGAGGTCTAC	TREVY	0.460422	3
1207	CTCACTTCGGTCTTA	LTSVL	1.039633	3
	CTGACTTCGGTCTTA	LTSVL	1.039633	3
	CTTACTTCGGTCTTA	LTSVL	1.039633	3
	TTAACTTCGGTCTTA	LTSVL	1.039633	3
	TTGACTTCGGTCTTA	LTSVL	1.039633	3
1261	CACATAGGTCTGAAA	HIGLK	0.951825	1
	CACATAGGTCTAAAG	HIGLK	0.951825	1
	CACATAGGTCTGAAG	HIGLK	0.951825	1
	CACATCGGTCTAAAA	HIGLK	0.89387	1
	CACATTGGTCTAAAA	HIGLK	0.89387	1
1264	CTCAAAGGTCTGATG	LKGLM	0.767369	1
	CTTAAAGGTCTAATG	LKGLM	0.613895	1
	CTTAAAGGTCTGATG	LKGLM	0.613895	1
	CTCAAAGGTCTTATG	LKGLM	0.613895	1
	CTCAAGGTCTAATG	LKGLM	0.599524	1

## REFERENCES

## References

1. Liu, L., Wang, J., Rosenberg, D., Zhao, H., Lengyel, G. and Nadel, D. (2018) Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. *Journal of Archaeological Science: Reports*, **21**, 783-793.
2. Tamang, J.P., Shin, D.-H., Jung, S.-J. and Chae, S.-W. (2016) Functional properties of microorganisms in fermented foods. *Frontiers in microbiology*, **7**, 578.
3. Mogensen, G., Salminen, S., CRITTENDEN, R., BIANCHI SALVADORI, B. and ZINK, R. (2002) Inventory of microorganisms with a documented history of use in food. *Bulletin-International Dairy Federation*, 10-19.
4. Bernardeau, M., Guguen, M. and Vernoux, J.P. (2006) Beneficial lactobacilli in food and feed: long-term use, biodiversity and proposals for specific and realistic safety assessments. *FEMS Microbiology Reviews*, **30**, 487-513.
5. Bourdichon, F., Casaregola, S., Farrokh, C., Frisvad, J.C., Gerds, M.L., Hammes, W.P., Harnett, J., Huys, G., Laulund, S. and Ouwehand, A. (2012) Food fermentations: microorganisms with technological beneficial use. *International journal of food microbiology*, **154**, 87-97.
6. Thapa, N. and Tamang, J.P. (2015) Functionality and therapeutic values of fermented foods. *Health Benefits of Fermented Foods*, ed. JP Tamang (New York: CRC Press), 111-168.
7. Wainwright, M. (2003) An alternative view of the early history of microbiology. *Advances in applied microbiology*, **52**, 333-356.
8. Schlenk, F. (1985) Early Research on Fermentation: A Story of Missed Opportunities. *Trends in Biochemical Sciences*, **10**, 252-254.
9. Cornish-Bawden, A. (1997) *New Beer in an Old Bottle. Eduard Buchner and the Growth of Biochemical Knowledge*. Universitat de València.
10. Gonzalo, D.G. and Lavandera, I. (2021) Enzymatic Commercial Sources. *Biocatalysis for Practitioners: Techniques, Reactions and Applications*, 467-485.
11. insight, F.b.
12. Borrelle, S.B., Ringma, J., Law, K.L., Monnahan, C.C., Lebreton, L., McGivern, A., Murphy, E., Jambeck, J., Leonard, G.H. and Hilleary, M.A. (2020) Predicted growth in plastic waste exceeds efforts to mitigate plastic pollution. *Science*, **369**, 1515-1518.
13. Lau, W.W., Shiran, Y., Bailey, R.M., Cook, E., Stuchtey, M.R., Koskella, J., Velis, C.A., Godfrey, L., Boucher, J. and Murphy, M.B. (2020) Evaluating scenarios toward zero plastic pollution. *Science*, **369**, 1455-1461.
14. MacLeod, M., Arp, H.P.H., Tekman, M.B. and Jahnke, A. (2021) The global threat from plastic pollution. *Science*, **373**, 61-65.
15. Cucina, M., de Nisi, P., Tambone, F. and Adani, F. (2021) The role of waste management in reducing bioplastics' leakage into the environment: a review. *Bioresource Technology*, 125459.
16. Triantis, V., Bode, L. and van Neerven, R.J. (2018) Immunological effects of human milk oligosaccharides. *Frontiers in pediatrics*, **6**, 190.

## References

17. Bych, K., Mikš, M.H., Johanson, T., Hederos, M.J., Vignæs, L.K. and Becker, P. (2019) Production of HMOs using microbial hosts—from cell engineering to large scale production. *Current opinion in biotechnology*, **56**, 130-137.
18. Whittall, D.R., Baker, K.V., Breitling, R. and Takano, E. (2020) Host systems for the production of recombinant spider silk. *Trends in Biotechnology*.
19. Kruis, A.J., Bohnenkamp, A.C., Patinios, C., van Nuland, Y.M., Levisson, M., Mars, A.E., van den Berg, C., Kengen, S.W.M. and Weusthuis, R.A. (2019) Microbial production of short and medium chain esters: Enzymes, pathways, and applications. *Biotechnol Adv*, **37**, 107407.
20. Gabriel, C. (1928) Butanol fermentation process1. *Industrial & engineering chemistry*, **20**, 1063-1067.
21. Gibbs, D. (1983) The rise and fall (... and rise?) of acetone/butanol fermentations. *Trends in Biotechnology*, **1**, 12-15.
22. Killeffer, D. (1927) Butanol and acetone from corn1: a description of the fermentation process. *Industrial & Engineering Chemistry*, **19**, 46-50.
23. Spivey, M. (1978) Acetone/butanol/ethanol fermentation. *Process Biochem.:(United Kingdom)*, **13**.
24. Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D. and Hitti, J. (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *Journal of bacteriology*, **183**, 4823-4838.
25. Noh, H.J., Lee, S.Y. and Jang, Y.-S. (2019) Microbial production of butyl butyrate, a flavor and fragrance compound. *Applied microbiology and biotechnology*, 1-8.
26. Fang, D., Wen, Z., Lu, M., Li, A., Ma, Y., Tao, Y. and Jin, M. (2020) Metabolic and process engineering of *Clostridium beijerinckii* for butyl acetate production in one step. *Journal of Agricultural and Food Chemistry*, **68**, 9475-9487.
27. Noh, H.J., Woo, J.E., Lee, S.Y. and Jang, Y.-S. (2018) Metabolic engineering of *Clostridium acetobutylicum* for the production of butyl butyrate. *Applied microbiology and biotechnology*, **102**, 8319-8327.
28. van den Berg, C., Heeres, A.S., van der Wielen, L.A. and Straathof, A.J. (2013) Simultaneous clostridial fermentation, lipase-catalyzed esterification, and ester extraction to enrich diesel with butyl butyrate. *Biotechnology and bioengineering*, **110**, 137-142.
29. Xin, F., Basu, A., Yang, K.-L. and He, J. (2016) Strategies for production of butanol and butyl-butylate through lipase-catalyzed esterification. *Bioresource technology*, **202**, 214-219.
30. Seo, S.O., Wang, Y., Lu, T., Jin, Y.S. and Blaschek, H.P. (2017) Characterization of a *Clostridium beijerinckii* spo0A mutant and its application for butyl butyrate production. *Biotechnology and bioengineering*, **114**, 106-112.
31. Zhang, Z.T., Taylor, S. and Wang, Y. (2017) In situ esterification and extractive fermentation for butyl butyrate production with *Clostridium tyrobutyricum*. *Biotechnology and bioengineering*, **114**, 1428-1437.
32. Feng, J., Zhang, J., Ma, Y., Feng, Y., Wang, S., Guo, N., Wang, H., Wang, P., Jiménez-Bonilla, P. and Gu, Y. (2021) Renewable fatty acid ester production in *Clostridium*. *Nature Communications*, **12**, 1-13.

33. Fischer, E. and Speier, A. (1924), *Untersuchungen aus Verschiedenen Gebieten*. Springer, pp. 285-291.
34. Kruis, A.J., Bohnenkamp, A.C., Patinios, C., van Nuland, Y.M., Levisson, M., Mars, A.E., van den Berg, C., Kengen, S.W. and Weusthuis, R.A. (2019) Microbial production of short and medium chain esters: Enzymes, pathways, and applications. *Biotechnology advances*.
35. Kruis, A.J., Levisson, M., Mars, A.E., van der Ploeg, M., Daza, F.G., Ellena, V., Kengen, S.W., van der Oost, J. and Weusthuis, R.A. (2017) Ethyl acetate production by the elusive alcohol acetyltransferase from yeast. *Metabolic Engineering*, **41**, 92-101.
36. Kruis, A.J., Gallone, B., Jonker, T., Mars, A.E., van Rijswijk, I.M., Wolkers–Rooijackers, J., Smid, E.J., Steensels, J., Verstrep, K.J. and Kengen, S.W. (2018) Contribution of Eat1 and other alcohol acyltransferases to ester production in *Saccharomyces cerevisiae*. *Frontiers in microbiology*, **9**, 3202.
37. Layton, D.S. and Trinh, C.T. (2016) Expanding the modular ester fermentative pathways for combinatorial biosynthesis of esters from volatile organic acids. *Biotechnology and bioengineering*, **113**, 1764-1776.
38. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology*, **169**, 5429-5433.
39. Mojica, F.J., Juez, G. and Rodriguez-Valera, F. (1993) Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Molecular microbiology*, **9**, 613-621.
40. Jansen, R., Embden, J.D.v., Gastra, W. and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, **43**, 1565-1575.
41. Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**, 2551-2561.
42. Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J. and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, **60**, 174-182.
43. Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, **151**, 653-663.
44. Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct*, **1**, 1-26.
45. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.
46. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and Van Der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960-964.

## References

47. Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V. and Van der Oost, J. (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*, **353**.
48. Mougiakos, I., Bosma, E.F., de Vos, W.M., van Kranenburg, R. and van der Oost, J. (2016) Next generation prokaryotic engineering: the CRISPR-Cas toolkit. *Trends in biotechnology*, **34**, 575-587.
49. Wang, F. and Qi, L.S. (2016) Applications of CRISPR genome engineering in cell biology. *Trends in cell biology*, **26**, 875-888.
50. Kirchner, M. and Schneider, S. (2015) CRISPR-Cas: From the Bacterial Adaptive Immune System to a Versatile Tool for Genome Engineering. *Angewandte Chemie International Edition*, **54**, 13508-13514.
51. Selle, K. and Barrangou, R. (2015) Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends in microbiology*, **23**, 225-232.
52. Finger-Bou, M., Orsi, E., van der Oost, J. and Staals, R.H. (2020) CRISPR with a happy ending: Non-templated DNA repair for prokaryotic genome engineering. *Biotechnology journal*, **15**, 1900404.
53. Jakočiūnas, T., Jensen, M.K. and Keasling, J.D. (2016) CRISPR/Cas9 advances engineering of microbial cell factories. *Metabolic engineering*, **34**, 44-59.
54. Afzal, S., Sirohi, P. and Singh, N.K. (2020) A review of CRISPR associated genome engineering: application, advances and future prospects of genome targeting tool for crop improvement. *Biotechnology Letters*, 1-22.
55. Shanmugam, S., Ngo, H.-H. and Wu, Y.-R. (2020) Advanced CRISPR/Cas-based genome editing tools for microbial biofuels production: A review. *Renewable Energy*, **149**, 1107-1119.
56. Lau, A., Ren, C.L. and Lee, L.P. (2020) Critical review on where CRISPR meets molecular diagnostics. *Progress in Biomedical Engineering*.
57. Huang, C.-H., Lee, K.-C. and Doudna, J.A. (2018) Applications of CRISPR-Cas enzymes in cancer therapeutics and detection. *Trends in cancer*, **4**, 499-512.
58. Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W., Essletzbichler, P., Dy, A.J., Joung, J., Verdine, V., Donghia, N., Daringer, N.M. and Freije, C.A. (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, **356**, 438-442.
59. Nethery, M.A., Korvink, M., Makarova, K.S., Wolf, Y.I., Koonin, E.V. and Barrangou, R. (2021) CRISPRclassify: Repeat-Based Classification of CRISPR Loci. *The CRISPR journal*, **4**, 558-574.
60. Wu, W.Y., Lebbink, J.H., Kanaar, R., Geijsen, N. and Van Der Oost, J. (2018) Genome editing by natural and engineered CRISPR-associated nucleases. *Nature chemical biology*, **14**, 642-651.
61. Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*, **31**, 233-239.
62. Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J. and Yang, S. (2015) Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Applied and environmental microbiology*, **81**, 2506-2514.
63. Zhao, D., Yuan, S., Xiong, B., Sun, H., Ye, L., Li, J., Zhang, X. and Bi, C. (2016) Development of a fast and easy method for Escherichia coli genome editing with CRISPR/Cas9. *Microbial cell factories*, **15**, 1-9.

64. Sun, J., Wang, Q., Jiang, Y., Wen, Z., Yang, L., Wu, J. and Yang, S. (2018) Genome editing and transcriptional repression in *Pseudomonas putida* KT2440 via the type II CRISPR system. *Microbial cell factories*, **17**, 1-17.
65. Zhang, J., Hong, W., Zong, W., Wang, P. and Wang, Y. (2018) Markerless genome editing in *Clostridium beijerinckii* using the CRISPR-Cpf1 system. *Journal of biotechnology*, **284**, 27-30.
66. Wasels, F., Jean-Marie, J., Collas, F., López-Contreras, A.M. and Ferreira, N.L. (2017) A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *Journal of microbiological methods*, **140**, 5-11.
67. Diallo, M., Hocq, R., Collas, F., Chartier, G., Wasels, F., Wijaya, H.S., Werten, M.W., Wolbert, E.J., Kengen, S.W. and van der Oost, J. (2020) Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*. *Methods*, **172**, 51-60.
68. Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J.A. and Liu, D.R. (2015) Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nature chemical biology*, **11**, 316-318.
69. Truong, D.-J.J., Kühner, K., Kühn, R., Werfel, S., Engelhardt, S., Wurst, W. and Ortiz, O. (2015) Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic acids research*, **43**, 6450-6458.
70. Zetsche, B., Volz, S.E. and Zhang, F. (2015) A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature biotechnology*, **33**, 139-142.
71. Nihongaki, Y., Kawano, F., Nakajima, T. and Sato, M. (2015) Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature biotechnology*, **33**, 755-760.
72. Liu, K.I., Ramli, M.N.B., Woo, C.W.A., Wang, Y., Zhao, T., Zhang, X., Yim, G.R.D., Chong, B.Y., Gowher, A. and Chua, M.Z.H. (2016) A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nature chemical biology*, **12**, 980.
73. Zhao, J., Inomata, R., Kato, Y. and Miyagishi, M. (2020) Development of aptamer-based inhibitors for CRISPR/Cas system. *Nucleic Acids Research*.
74. Cañadas, I.S.C., Groothuis, D., Zygiouropoulou, M., Rodrigues, R. and Minton, N.P. (2019) RiboCas: a universal CRISPR-based editing tool for *Clostridium*. *ACS synthetic biology*, **8**, 1379-1390.
75. Tang, W., Hu, J.H. and Liu, D.R. (2017) Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nature communications*, **8**, 1-8.
76. Siu, K.-H. and Chen, W. (2019) Riboregulated toehold-gated gRNA for programmable CRISPR-Cas9 function. *Nature chemical biology*, **15**, 217-220.
77. Kundert, K., Lucas, J.E., Watters, K.E., Fellmann, C., Ng, A.H., Heineike, B.M., Fitzsimmons, C.M., Oakes, B.L., Qu, J. and Prasad, N. (2019) Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nature communications*, **10**, 1-11.
78. Moroz-Omori, E.V., Satyapertiwi, D., Ramel, M.-C., Høgset, H.k., Sunyovszki, I.K., Liu, Z., Wojciechowski, J.P., Zhang, Y., Grigsby, C.L. and Brito, L. (2020) Photoswitchable gRNAs for Spatiotemporally Controlled CRISPR-Cas-Based Genomic Regulation. *ACS Central Science*.
79. Jain, P.K., Ramanan, V., Schepers, A.G., Dalvie, N.S., Panda, A., Fleming, H.E. and Bhatia, S.N. (2016) Development of Light-Activated CRISPR Using Guide RNAs with Photocleavable Protectors. *Angewandte Chemie International Edition*, **55**, 12440-12444.



## References

80. Wang, Y., Liu, Y., Xie, F., Lin, J. and Xu, L. (2020) Photocontrol of CRISPR/Cas9 function by site-specific chemical modification of guide RNA. *Chemical science*, **11**, 11478-11484.
81. Zhou, W., Brown, W., Bardhan, A., Delaney, M., Ilk, A.S., Rauen, R.R., Kahn, S.I., Tsang, M. and Deiters, A. (2020) Spatiotemporal control of CRISPR/Cas9 function in cells and zebrafish using light-activated guide RNA. *Angewandte Chemie*, **132**, 9083-9088.
82. Saerens, S.M., Delvaux, F.R., Verstrepen, K.J. and Thevelein, J.M. (2010) Production and biological function of volatile esters in *Saccharomyces cerevisiae*. *Microbial biotechnology*, **3**, 165-177.
83. Berger, R.G. (2009) Biotechnology of flavours—the next generation. *Biotechnology letters*, **31**, 1651-1659.
84. Carroll, A.L., Desai, S.H. and Atsumi, S. (2016) Microbial production of scent and flavor compounds. *Current opinion in biotechnology*, **37**, 8-15.
85. Rodriguez, G.M., Tashiro, Y. and Atsumi, S. (2014) Expanding ester biosynthesis in *Escherichia coli*. *Nature chemical biology*, **10**, 259-265.
86. Białecka-Florjańczyk, E. and Florjańczyk, Z. (2007), *Thermodynamics, solubility and environmental issues*. Elsevier, pp. 397-408.
87. Durrans, T.H. and Davies, E.H. (1971) *Solvents*.
88. Löser, C., Urit, T. and Bley, T. (2014) Perspectives for the biotechnological production of ethyl acetate by yeasts. *Applied microbiology and biotechnology*, **98**, 5397-5415.
89. Chuck, C.J. and Donnelly, J. (2014) The compatibility of potential bioderived fuels with Jet A-1 aviation kerosene. *Applied Energy*, **118**, 83-91.
90. Kalscheuer, R., Stölting, T. and Steinbüchel, A. (2006) Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology*, **152**, 2529-2536.
91. Lange, J.P., Price, R., Ayoub, P.M., Louis, J., Petrus, L., Clarke, L. and Gosselink, H. (2010) Valeric biofuels: a platform of cellulosic transportation fuels. *Angewandte Chemie International Edition*, **49**, 4479-4483.
92. Jyoti, G., Keshav, A., Anandkumar, J. and Bhoi, S. (2018) Homogeneous and heterogeneous catalyzed esterification of acrylic acid with ethanol: reaction kinetics and modeling. *International Journal of Chemical Kinetics*, **50**, 370-380.
93. Liu, Y., Lotero, E. and Goodwin Jr, J.G. (2006) Effect of water on sulfuric acid catalyzed esterification. *Journal of Molecular Catalysis A: Chemical*, **245**, 132-140.
94. Saba, T., Estephane, J., El Khoury, B., El Khoury, M., Khazma, M., El Zakhem, H. and Aouad, S. (2016) Biodiesel production from refined sunflower vegetable oil over KOH/ZSM5 catalysts. *Renewable Energy*, **90**, 301-306.
95. Vyas, A.P., Verma, J.L. and Subrahmanyam, N. (2010) A review on FAME production processes. *Fuel*, **89**, 1-9.
96. Donate, P.M. (2014) Green synthesis from biomass. *Chemical and Biological Technologies in Agriculture*, **1**, 1-8.
97. Haveren, J.V., Scott, E.L. and Sanders, J. (2008) Bulk chemicals from biomass. *Biofuels, Bioproducts and Biorefining: Innovation for a sustainable economy*, **2**, 41-57.

98. Vennestrøm, P., Osmundsen, C.M., Christensen, C. and Taarning, E. (2011) Beyond petrochemicals: the renewable chemicals industry. *Angewandte Chemie International Edition*, **50**, 10502-10509.
99. Werpy, T. and Petersen, G. (2004). National Renewable Energy Lab., Golden, CO (US).
100. Quax, W.J. (2006) Bacterial enzymes. *Prokaryotes*, **1**, 777-796.
101. Council, N.R. (2015) *Industrialization of biology: a roadmap to accelerate the advanced manufacturing of chemicals*. National Academies Press.
102. Marella, E.R., Holkenbrink, C., Siewers, V. and Borodina, I. (2018) Engineering microbial fatty acid metabolism for biofuels and biochemicals. *Current opinion in biotechnology*, **50**, 39-46.
103. Dien, V.S. (2013) From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. *Current opinion in biotechnology*, **24**, 1061-1068.
104. Corbion. (2019).
105. BioSolvents, V. (2019).
106. Banella, M.B., Gioia, C., Vannini, M., Colonna, M., Celli, A. and Gandini, A. (2016) A sustainable route to a terephthalic acid precursor. *ChemSusChem*, **9**, 942-945.
107. Lee, S.Y., Kim, H.U., Chae, T.U., Cho, J.S., Kim, J.W., Shin, J.H., Kim, D.I., Ko, Y.-S., Jang, W.D. and Jang, Y.-S. (2019) A comprehensive metabolic map for production of bio-based chemicals. *Nature Catalysis*, **2**, 18-33.
108. Liu, S.-Q., Holland, R. and Crow, V. (2004) Esters and their biosynthesis in fermented dairy products: a review. *International Dairy Journal*, **14**, 923-945.
109. Dzialo, M.C., Park, R., Steensels, J., Lievens, B. and Verstrepen, K.J. (2017) Physiology, ecology and industrial applications of aroma formation in yeast. *FEMS microbiology reviews*, **41**, S95-S128.
110. Tucci, S., Vacula, R., Krajcovic, J., Proksch, P. and Martin, W. (2010) Variability of wax ester fermentation in natural and bleached *Euglena gracilis* strains in response to oxygen and the elongase inhibitor flufenacet. *Journal of Eukaryotic Microbiology*, **57**, 63-69.
111. Beijerinck, M. (1892) Zur Ernährungsphysiologie des Kahmpilzes. *Centralblatt für Bakteriologie und Parasitenkunde*, **11**, 68-75.
112. Armstrong, D.W., Martin, S.M. and Yamazaki, H. (1984) Production of ethyl acetate from dilute ethanol solutions by *Candida utilis*. *Biotechnology and bioengineering*, **26**, 1038-1041.
113. Meersman, E., Steensels, J., Struyf, N., Paulus, T., Saels, V., Mathawan, M., Allegaert, L., Vrancken, G. and Verstrepen, K.J. (2016) Tuning chocolate flavor through development of thermotolerant *Saccharomyces cerevisiae* starter cultures with increased acetate ester production. *Applied and environmental microbiology*, **82**, 732-746.
114. Tabachnick, J. and Joslyn, M. (1953) FORMATION OF ESTERS BY YEAST I: The Production of Ethyl Acetate by Standing Surface Cultures of *Hansenula Anomala*. *Journal of Bacteriology*, **65**, 1-9.
115. Tabachnick, J. and Joslyn, M. (1953) Formation of esters by yeast. II. Investigations with cellular suspensions of *Hansenula anomala*. *Plant physiology*, **28**, 681.

## References

116. Rijswijk, V.I.M., Wolkers–Rooijackers, J.C., Abee, T. and Smid, E.J. (2017) Performance of non-conventional yeasts in co-culture with brewers' yeast for steering ethanol and aroma production. *Microbial biotechnology*, **10**, 1591-1602.
117. Kruis, A.J., Mars, A.E., Kengen, S.W., Borst, J.W., van der Oost, J. and Weusthuis, R.A. (2018) Alcohol acetyltransferase Eat1 is located in yeast mitochondria. *Appl. Environ. Microbiol.*, **84**, e01640-01618.
118. Urit, T., Stukert, A., Bley, T. and Löser, C. (2012) Formation of ethyl acetate by *Kluyveromyces marxianus* on whey during aerobic batch cultivation at specific trace element limitation. *Applied microbiology and biotechnology*, **96**, 1313-1323.
119. Fredlund, E., Beerlage, C., Melin, P., Schnürer, J. and Passoth, V. (2006) Oxygen and carbon source-regulated expression of PDC and ADH genes in the respiratory yeast *Pichia anomala*. *Yeast*, **23**, 1137-1149.
120. Kallel-Mhiri, H., Engasser, J.-M. and Miclo, A. (1993) Continuous ethyl acetate production by *Kluyveromyces fragilis* on whey permeate. *Applied microbiology and biotechnology*, **40**, 201-205.
121. Löser, C., Urit, T., Nehl, F. and Bley, T. (2011) Screening of *Kluyveromyces* strains for the production of ethyl acetate: design and evaluation of a cultivation system. *Engineering in Life Sciences*, **11**, 369-381.
122. Willetts, A. (1989) Ester formation from ethanol by *Candida pseudotropicalis*. *Antonie van Leeuwenhoek*, **56**, 175-180.
123. Medeiros, A.B., Pandey, A., Christen, P., Fontoura, P.S., de Freitas, R.J. and Soccol, C.R. (2001) Aroma compounds produced by *Kluyveromyces marxianus* in solid state fermentation on a packed bed column bioreactor. *World Journal of Microbiology and Biotechnology*, **17**, 767-771.
124. Fonseca, G.G., Heinze, E., Wittmann, C. and Gombert, A.K. (2008) The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Applied microbiology and biotechnology*, **79**, 339-354.
125. Urit, T., Manthey, R., Bley, T. and Löser, C. (2013) Formation of ethyl acetate by *Kluyveromyces marxianus* on whey: Influence of aeration and inhibition of yeast growth by ethyl acetate. *Engineering in Life Sciences*, **13**, 247-260.
126. Urit, T., Li, M., Bley, T. and Löser, C. (2013) Growth of *Kluyveromyces marxianus* and formation of ethyl acetate depending on temperature. *Applied microbiology and biotechnology*, **97**, 10359-10371.
127. Fredlund, E., Blank, L.M., Schnürer, J., Sauer, U. and Passoth, V. (2004) Oxygen- and glucose-dependent regulation of central carbon metabolism in *Pichia anomala*. *Appl. Environ. Microbiol.*, **70**, 5905-5911.
128. Fredlund, E., Druvefors, U.Ä., Olstorpe, M.N., Passoth, V. and Schnürer, J. (2004) Influence of ethyl acetate production and ploidy on the anti-mould activity of *Pichia anomala*. *FEMS microbiology letters*, **238**, 133-137.
129. Fredlund, E., Druvefors, U., Boysen, M.E., Lingsten, K.-J. and Schnürer, J. (2002) Physiological characteristics of the biocontrol yeast *Pichia anomala* J121. *FEMS Yeast Research*, **2**, 395-402.
130. Iwaki, H., Hasegawa, Y., Wang, S., Kayser, M.M. and Lau, P.C. (2002) Cloning and characterization of a gene cluster involved in cyclopentanol metabolism in *Comamonas* sp. strain NCIMB 9872 and biotransformations effected by *Escherichia coli*-expressed

- cyclopentanone 1, 2-monooxygenase. *Applied and environmental microbiology*, **68**, 5671-5684.
131. Kotani, T., Yurimoto, H., Kato, N. and Sakai, Y. (2007) Novel acetone metabolism in a propane-utilizing bacterium, *Gordonia* sp. strain TY-5. *Journal of bacteriology*, **189**, 886-893.
  132. Christiaens, J.F., Franco, L.M., Cools, T.L., De Meester, L., Michiels, J., Wenseleers, T., Hassan, B.A., Yaksi, E. and Verstrepen, K.J. (2014) The fungal aroma gene ATF1 promotes dispersal of yeast cells through insect vectors. *Cell reports*, **9**, 425-432.
  133. Onwueme, K.C., Vos, C.J., Zurita, J., Ferreras, J.A. and Quadri, L.E. (2005) The dimycocerosate ester polyketide virulence factors of mycobacteria. *Progress in lipid research*, **44**, 259-302.
  134. Ishige, T., Tani, A., Sakai, Y. and Kato, N. (2003) Wax ester production by bacteria. *Current opinion in microbiology*, **6**, 244-250.
  135. Wältermann, M. and Steinbüchel, A. (2005) Neutral lipid bodies in prokaryotes: recent insights into structure, formation, and relationship to eukaryotic lipid depots. *Journal of bacteriology*, **187**, 3607-3619.
  136. Müller, M., Mentel, M., van Hellemond, J.J., Henze, K., Woehle, C., Gould, S.B., Yu, R.-Y., van der Giezen, M., Tielens, A.G. and Martin, W.F. (2012) Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiology and Molecular Biology Reviews*, **76**, 444-495.
  137. Flamholz, A., Noor, E., Bar-Even, A. and Milo, R. (2012) eQuilibrator—the biochemical thermodynamics calculator. *Nucleic acids research*, **40**, D770-D775.
  138. Sherkanov, S., Korman, T.P., Clarke, S.G. and Bowie, J.U. (2016) Production of FAME biodiesel in *E. coli* by direct methylation with an insect enzyme. *Scientific reports*, **6**, 24239.
  139. Menendez-Bravo, S., Comba, S., Sabatini, M., Arabolaza, A. and Gramajo, H. (2014) Expanding the chemical diversity of natural esters by engineering a polyketide-derived pathway into *Escherichia coli*. *Metabolic engineering*, **24**, 97-106.
  140. Menendez-Bravo, S., Comba, S., Gramajo, H. and Arabolaza, A. (2017) Metabolic engineering of microorganisms for the production of structurally diverse esters. *Applied Microbiology & Biotechnology*, **101**.
  141. Bornscheuer, U.T. (2018) Enzymes in lipid modification. *Annual review of food science and technology*, **9**, 85-103.
  142. Khan, N.R. and Rathod, V.K. (2015) Enzyme catalyzed synthesis of cosmetic esters and its intensification: A review. *Process Biochemistry*, **50**, 1793-1806.
  143. Mukdsi, M.C.A., Medina, R.B., Alvarez, M.d.F. and González, S.N. (2009) Ester synthesis by lactic acid bacteria isolated from goat's and ewe's milk and cheeses. *Food Chemistry*, **117**, 241-247.
  144. Mukdsi, M.C.A., Maillard, M.-B., Medina, R.B. and Thierry, A. (2018) Ethyl butanoate is synthesised both by alcoholysis and esterification by dairy lactobacilli and propionibacteria. *LWT*, **89**, 38-43.
  145. Costello, P., Siebert, T., Solomon, M. and Bartowsky, E. (2013) Synthesis of fruity ethyl esters by acyl coenzyme A: alcohol acyltransferase and reverse esterase activities in *Oenococcus oeni* and *Lactobacillus plantarum*. *Journal of Applied Microbiology*, **114**, 797-806.

## References

146. Hong, Q., Liu, X., Hang, F., Zhao, J., Zhang, H. and Chen, W. (2018) Screening of adjunct cultures and their application in ester formation in Camembert-type cheese. *Food microbiology*, **70**, 33-41.
147. Kashima, Y., Iijima, M., Nakano, T., Tayama, K., Koizumi, Y., Udaka, S. and Yanagida, F. (2000) Role of intracellular esterases in the production of esters by *Acetobacter pasteurianus*. *Journal of bioscience and bioengineering*, **89**, 81-83.
148. Rojas, V., Gil, J.V., Manzanares, P., Gavara, R., Piñaga, F. and Flors, A. (2002) Measurement of alcohol acetyltransferase and ester hydrolase activities in yeast extracts. *Enzyme and Microbial Technology*, **30**, 224-230.
149. Kashima, Y., Iijima, M., Okamoto, A., Koizumi, Y., Udaka, S. and Yanagida, F. (1998) Purification and characterization of intracellular esterases related to ethylacetate formation in *Acetobacter pasteurianus*. *Journal of fermentation and bioengineering*, **85**, 584-588.
150. Kusano, M., Sakai, Y., Kato, N., Yoshimoto, H., Sone, H. and Tamai, Y. (1998) Hemiacetal dehydrogenation activity of alcohol dehydrogenases in *Saccharomyces cerevisiae*. *Bioscience, biotechnology, and biochemistry*, **62**, 1956-1961.
151. Kusano, M., Sakai, Y., Kato, N., Yoshimoto, H. and Tamai, Y. (1999) A novel hemiacetal dehydrogenase activity involved in ethyl acetate synthesis in *Candida utilis*. *Journal of bioscience and bioengineering*, **87**, 690-692.
152. Park, Y.C., Shaffer, C.E.H. and Bennett, G.N. (2009) Microbial formation of esters. *Applied microbiology and biotechnology*, **85**, 13-25.
153. Sakai, Y., Murdanoto, A.P., Sembiring, L., Tani, Y. and Kato, N. (1995) A novel formaldehyde oxidation pathway in methylotrophic yeasts: methylformate as a possible intermediate. *FEMS microbiology letters*, **127**, 229-234.
154. Kunjapur, A.M. and Prather, K.L. (2015) Microbial engineering for aldehyde synthesis. *Applied and environmental microbiology*, **81**, 1892-1901.
155. Yurimoto, H., Kato, N. and Sakai, Y. (2005) Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *The Chemical Record*, **5**, 367-375.
156. Murdanoto, A.P., Sakai, Y., Sembiring, L., Tani, Y. and Kato, N. (1997) Ester synthesis by NAD<sup>+</sup>-dependent dehydrogenation of hemiacetal: production of methyl formate by cells of methylotrophic yeasts. *Bioscience, biotechnology, and biochemistry*, **61**, 1391-1393.
157. Yurimoto, H., Lee, B., Yasuda, F., Sakai, Y. and Kato, N. (2004) Alcohol dehydrogenases that catalyse methyl formate synthesis participate in formaldehyde detoxification in the methylotrophic yeast *Candida boidinii*. *Yeast*, **21**, 341-350.
158. Löbs, A.-K., Engel, R., Schwartz, C., Flores, A. and Wheeldon, I. (2017) CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *Biotechnology for biofuels*, **10**, 1-14.
159. Park, Y.-C., San, K.-Y. and Bennett, G.N. (2007) Characterization of alcohol dehydrogenase 1 and 3 from *Neurospora crassa* FGSC2489. *Applied microbiology and biotechnology*, **76**, 349-356.
160. Fraaije, M.W., Kamerbeek, N.M., van Berkel, W.J. and Janssen, D.B. (2002) Identification of a Baeyer–Villiger monooxygenase sequence motif. *FEBS letters*, **518**, 43-47.

161. Mascotti, M.L., Lapadula, W.J. and Juri Ayub, M. (2015) The origin and evolution of Baeyer—Villiger monooxygenases (BVMOs): An ancestral family of flavin monooxygenases. *PloS one*, **10**, e0132689.
162. Frank, B., Wenzel, S.C., Bode, H.B., Scharfe, M., Blöcker, H. and Müller, R. (2007) From genetic diversity to metabolic unity: Studies on the biosynthesis of aurafurones and aurafuron-like structures in myxobacteria and streptomycetes. *Journal of molecular biology*, **374**, 24-38.
163. Wen, Y., Hatabayashi, H., Arai, H., Kitamoto, H.K. and Yabe, K. (2005) Function of the cypX and moxY genes in aflatoxin biosynthesis in *Aspergillus parasiticus*. *Applied and Environmental Microbiology*, **71**, 3192-3198.
164. Britton, L.N. and Markovetz, A. (1977) A novel ketone monooxygenase from *Pseudomonas cepacia*. Purification and properties. *Journal of Biological Chemistry*, **252**, 8561-8566.
165. Werf, V.D.M.J. and Boot, A.M. (2000) Metabolism of carveol and dihydrocarveol in *Rhodococcus erythropolis* DCL14. *Microbiology*, **146**, 1129-1141.
166. Onaca, C., Kieninger, M., Engesser, K.-H. and Altenbuchner, J. (2007) Degradation of alkyl methyl ketones by *Pseudomonas veronii* MEK700. *Journal of bacteriology*, **189**, 3759-3767.
167. Völker, A., Kirschner, A., Bornscheuer, U.T. and Altenbuchner, J. (2008) Functional expression, purification, and characterization of the recombinant Baeyer-Villiger monooxygenase MekA from *Pseudomonas veronii* MEK700. *Applied microbiology and biotechnology*, **77**, 1251-1260.
168. Kamerbeek, N.M., Janssen, D.B., van Berkel, W.J. and Fraaije, M.W. (2003) Baeyer–Villiger monooxygenases, an emerging family of flavin-dependent biocatalysts. *Advanced Synthesis & Catalysis*, **345**, 667-678.
169. Leisch, H., Morley, K. and Lau, P.C. (2011) Baeyer–Villiger monooxygenases: more than just green chemistry. *Chemical reviews*, **111**, 4165-4222.
170. Pazmiño, D.E.T., Dudek, H.M. and Fraaije, M.W. (2010) Baeyer–Villiger monooxygenases: recent advances and future challenges. *Current opinion in chemical biology*, **14**, 138-144.
171. Gonzalo, D.G., Mihovilovic, M.D. and Fraaije, M.W. (2010) Recent developments in the application of Baeyer-Villiger monooxygenases as biocatalysts. *ChemBioChem*, **11**, 2208-2231.
172. Beekwilder, J., Alvarez-Huerta, M., Neef, E., Verstappen, F.W., Bouwmeester, H.J. and Aharoni, A. (2004) Functional characterization of enzymes forming volatile esters from strawberry and banana. *Plant Physiology*, **135**, 1865-1878.
173. Aharoni, A., Keizer, L.C., Bouwmeester, H.J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A.M., De Vos, R.C. and van der Voet, H. (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *The Plant Cell*, **12**, 647-661.
174. Holland, D., Larkov, O., Bar-Ya'akov, I., Bar, E., Zax, A., Brandeis, E., Ravid, U. and Lewinsohn, E. (2005) Developmental and varietal differences in volatile ester formation and acetyl-CoA: alcohol acetyl transferase activities in apple (*Malus domestica* Borkh.) fruit. *Journal of Agricultural and Food Chemistry*, **53**, 7198-7203.
175. Nagasawa, N., Bogaki, T., Iwamatsu, A., HAMACHI, M. and KUMAGAI, C. (1998) Cloning and nucleotide sequence of the alcohol acetyltransferase II gene (ATF2) from

## References

- Saccharomyces cerevisiae Kyokai No. 7. *Bioscience, biotechnology, and biochemistry*, **62**, 1852-1857.
176. Saerens, S.M., Verstrepen, K.J., Van Laere, S.D., Voet, A.R., Van Dijck, P., Delvaux, F.R. and Thevelein, J.M. (2006) The Saccharomyces cerevisiae EHT1 and EEB1 genes encode novel enzymes with medium-chain fatty acid ethyl ester synthesis and hydrolysis capacity. *Journal of Biological Chemistry*, **281**, 4446-4456.
177. Shalit, M., Katzir, N., Tadmor, Y., Larkov, O., Burger, Y., Shalekhet, F., Lastochkin, E., Ravid, U., Amar, O. and Edelstein, M. (2001) Acetyl-CoA: alcohol acetyltransferase activity and aroma formation in ripening melon fruits. *Journal of Agricultural and Food Chemistry*, **49**, 794-799.
178. Stribny, J., Querol, A. and Pérez-Torrado, R. (2016) Differences in enzymatic properties of the Saccharomyces kudriavzevii and Saccharomyces uvarum alcohol acetyltransferases and their impact on aroma-active compounds production. *Frontiers in microbiology*, **7**, 897.
179. Mastrigt, V.O., Abee, T., Lillevang, S.K. and Smid, E.J. (2018) Quantitative physiology and aroma formation of a dairy Lactococcus lactis at near-zero growth rates. *Food microbiology*, **73**, 216-226.
180. Shi, S., Valle-Rodríguez, J.O., Khoomrung, S., Siewers, V. and Nielsen, J. (2012) Functional expression and characterization of five wax ester synthases in Saccharomyces cerevisiae and their utility for biodiesel production. *Biotechnology for biofuels*, **5**, 7.
181. Röttig, A. and Steinbüchel, A. (2013) Acyltransferases in bacteria. *Microbiology and Molecular Biology Reviews*, **77**, 277-321.
182. Stöveken, T., Kalscheuer, R., Malkus, U., Reichelt, R. and Steinbüchel, A. (2005) The wax ester synthase/acyl coenzyme A: diacylglycerol acyltransferase from Acinetobacter sp. strain ADP1: characterization of a novel type of acyltransferase. *Journal of bacteriology*, **187**, 1369-1376.
183. Fujii, T., Nagasawa, N., Iwamatsu, A., Bogaki, T., Tamai, Y. and Hamachi, M. (1994) Molecular cloning, sequence analysis, and expression of the yeast alcohol acetyltransferase gene. *Applied and environmental microbiology*, **60**, 2786-2792.
184. Verstrepen, K.J., Van Laere, S.D., Vanderhaegen, B.M., Derdelinckx, G., Dufour, J.-P., Pretorius, I.S., Winderickx, J., Thevelein, J.M. and Delvaux, F.R. (2003) Expression levels of the yeast alcohol acetyltransferase genes ATF1, Lg-ATF1, and ATF2 control the formation of a broad range of volatile esters. *Applied and environmental microbiology*, **69**, 5228-5237.
185. Tiwari, R., Köffel, R. and Schneider, R. (2007) An acetylation/deacetylation cycle controls the export of sterols and steroids from S. cerevisiae. *The EMBO journal*, **26**, 5109-5119.
186. Tehlivets, O., Scheuringer, K. and Kohlwein, S.D. (2007) Fatty acid synthesis and elongation in yeast. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, **1771**, 255-270.
187. Holt, S., Trindade de Carvalho, B., Foulquié-Moreno, M.R. and Thevelein, J.M. (2018) Polygenic analysis in absence of major effector ATF1 unveils novel components in yeast flavor ester biosynthesis. *MBio*, **9**, e01279-01218.
188. Löser, C., Urit, T., Keil, P. and Bley, T. (2015) Studies on the mechanism of synthesis of ethyl acetate in Kluyveromyces marxianus DSM 5422. *Applied microbiology and biotechnology*, **99**, 1131-1144.

189. Thomas, K. and Dawson, P. (1978) Relationship between iron-limited growth and energy limitation during phased cultivation of *Candida utilis*. *Canadian journal of microbiology*, **24**, 440-447.
190. Huh, W.-K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S. and O'Shea, E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686-691.
191. Löbs, A.-K., Schwartz, C., Thorwall, S. and Wheeldon, I. (2018) Highly multiplexed CRISPRi repression of respiratory functions enhances mitochondrial localized ethyl acetate biosynthesis in *Kluyveromyces marxianus*. *ACS Synthetic Biology*, **7**, 2647-2655.
192. Lin, J.-L. and Wheeldon, I. (2014) Dual N-and C-terminal helices are required for endoplasmic reticulum and lipid droplet association of alcohol acetyltransferases in *Saccharomyces cerevisiae*. *PLoS One*, **9**, e104141.
193. Verstrepen, K.J., Van Laere, S.D., Vercammen, J., Derdelinckx, G., Dufour, J.P., Pretorius, I.S., Winderickx, J., Thevelein, J.M. and Delvaux, F.R. (2004) The *Saccharomyces cerevisiae* alcohol acetyl transferase Atf1p is localized in lipid particles. *Yeast*, **21**, 367-377.
194. Zahedi, R.P., Sickmann, A., Boehm, A.M., Winkler, C., Zufall, N., Schonfisch, B., Guiard, B., Pfanner, N. and Meisinger, C. (2006) Proteomic analysis of the yeast mitochondrial outer membrane reveals accumulation of a subclass of preproteins. *Molecular biology of the cell*, **17**, 1436-1450.
195. Zhu, J., Schwartz, C. and Wheeldon, I. (2019) Controlled intracellular trafficking alleviates an expression bottleneck in *S. cerevisiae* ester biosynthesis. *Metabolic engineering communications*, **8**, e00085.
196. Galaz, S., Morales-Quintana, L., Moya-León, M.A. and Herrera, R. (2013) Structural analysis of the alcohol acyltransferase protein family from *Cucumis melo* shows that enzyme activity depends on an essential solvent channel. *The FEBS journal*, **280**, 1344-1357.
197. Morales-Quintana, L., Fuentes, L., Gaete-Eastman, C., Herrera, R. and Moya-León, M.A. (2011) Structural characterization and substrate specificity of VpAAT1 protein related to ester biosynthesis in mountain papaya fruit. *Journal of Molecular Graphics and Modelling*, **29**, 635-642.
198. Navarro-Retamal, C., Gaete-Eastman, C., Herrera, R., Caballero, J. and Alzate-Morales, J.H. (2016) Structural and affinity determinants in the interaction between alcohol acyltransferase from *F. x ananassa* and several alcohol substrates: a computational study. *PloS one*, **11**, e0153057.
199. D'Auria, J.C. (2006) Acyltransferases in plants: a good time to be BAHD. *Current opinion in plant biology*, **9**, 331-340.
200. Nancolas, B., Bull, I.D., Stenner, R., Dufour, V. and Curnow, P. (2017) *Saccharomyces cerevisiae* Atf1p is an alcohol acetyltransferase and a thioesterase in vitro. *Yeast*, **34**, 239-251.
201. Morales-Quintana, L., Moya-León, M.A. and Herrera, R. (2015) Computational study enlightens the structural role of the alcohol acyltransferase DFGWG motif. *Journal of molecular modeling*, **21**, 1-10.
202. Morales-Quintana, L., Nuñez-Tobar, M.X., Moya-León, M.A. and Herrera, R. (2013) Molecular dynamics simulation and site-directed mutagenesis of alcohol acyltransferase: a



## References

- proposed mechanism of catalysis. *Journal of chemical information and modeling*, **53**, 2689-2700.
203. Knight, M.J., Bull, I.D. and Curnow, P. (2014) The yeast enzyme Eht1 is an octanoyl-CoA: ethanol acyltransferase that also functions as a thioesterase. *Yeast*, **31**, 463-474.
204. Wang, J., Mahajani, M., Jackson, S.L., Yang, Y., Chen, M., Ferreira, E.M., Lin, Y. and Yan, Y. (2017) Engineering a bacterial platform for total biosynthesis of caffeic acid derived phenethyl esters and amides. *Metabolic engineering*, **44**, 89-99.
205. Lin, J.L., Zhu, J. and Wheeldon, I. (2016) Rapid ester biosynthesis screening reveals a high activity alcohol-O-acyltransferase (AATase) from tomato fruit. *Biotechnology journal*, **11**, 700-707.
206. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S. and Sonnhammer, E.L. (2004) The Pfam protein families database. *Nucleic acids research*, **32**, D138-D141.
207. Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P. and Chatonnet, A. (2012) ESTHER, the database of the  $\alpha/\beta$ -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic acids research*, **41**, D423-D429.
208. Rauwerdink, A. and Kazlauskas, R.J. (2015) How the same core catalytic machinery catalyzes 17 different reactions: the serine-histidine-aspartate catalytic triad of  $\alpha/\beta$ -hydrolase fold enzymes. *ACS catalysis*, **5**, 6153-6176.
209. Jiang, Y., Morley, K.L., Schrag, J.D. and Kazlauskas, R.J. (2011) Different active-site loop orientation in serine hydrolases versus acyltransferases. *ChemBioChem*, **12**, 768-776.
210. Lejon, S., Ellis, J. and Valegård, K. (2008) The last step in cephalosporin C formation revealed: crystal structures of deacetylcephalosporin C acetyltransferase from *Acremonium chrysogenum* in complexes with reaction intermediates. *Journal of molecular biology*, **377**, 935-944.
211. Mirza, I.A., Nazi, I., Korczynska, M., Wright, G.D. and Berghuis, A.M. (2005) Crystal structure of homoserine transacetylase from *Haemophilus influenzae* reveals a new family of  $\alpha/\beta$ -hydrolases. *Biochemistry*, **44**, 15768-15773.
212. Löbs, A.K., Lin, J.L., Cook, M. and Wheeldon, I. (2016) High throughput, colorimetric screening of microbial ester biosynthesis reveals high ethyl acetate production from *Kluyveromyces marxianus* on C5, C6, and C12 carbon sources. *Biotechnology journal*, **11**, 1274-1281.
213. Nuland, V.Y.M., de Vogel, F.A., Scott, E.L., Eggink, G. and Weusthuis, R.A. (2017) Biocatalytic, one-pot diterminal oxidation and esterification of n-alkanes for production of  $\alpha$ ,  $\omega$ -diol and  $\alpha$ ,  $\omega$ -dicarboxylic acid esters. *Metabolic engineering*, **44**, 134-142.
214. Tai, Y.-S., Xiong, M. and Zhang, K. (2015) Engineered biosynthesis of medium-chain esters in *Escherichia coli*. *Metabolic engineering*, **27**, 20-28.
215. Levisson, M., van der Oost, J. and Kengen, S.W. (2009) Carboxylic ester hydrolases from hyperthermophiles. *Extremophiles*, **13**, 567-581.
216. Stergiou, P.-Y., Foukis, A., Filippou, M., Koukouritaki, M., Parapouli, M., Theodorou, L.G., Hatziloukas, E., Afendra, A., Pandey, A. and Papamichael, E.M. (2013) Advances in lipase-catalyzed esterification reactions. *Biotechnology advances*, **31**, 1846-1859.
217. Kumar, A., Dhar, K., Kanwar, S.S. and Arora, P.K. (2016) Lipase catalysis in organic solvents: advantages and applications. *Biological Procedures Online*, **18**, 1-11.

218. Hari, T.K., Yaakob, Z. and Binitha, N.N. (2015) Aviation biofuel from renewable resources: routes, opportunities and challenges. *Renewable and Sustainable Energy Reviews*, **42**, 1234-1244.
219. Ceccoli, R.D., Bianchi, D.A., Fink, M.J., Mihovilovic, M.D. and Rial, D.V. (2017) Cloning and characterization of the Type I Baeyer–Villiger monooxygenase from *Leptospira biflexa*. *AMB Express*, **7**, 1-13.
220. Pereira, J.P., van der Wielen, L.A. and Straathof, A.J. (2018) Perspectives for the microbial production of methyl propionate integrated with product recovery. *Bioresource technology*, **256**, 187-194.
221. Rehdorf, J., Lengar, A., Bornscheuer, U.T. and Mihovilovic, M.D. (2009) Kinetic resolution of aliphatic acyclic  $\beta$ -hydroxyketones by recombinant whole-cell Baeyer–Villiger monooxygenases—formation of enantiocomplementary regioisomeric esters. *Bioorganic & medicinal chemistry letters*, **19**, 3739-3743.
222. Horton, C.E. and Bennett, G.N. (2006) Ester production in *E. coli* and *C. acetobutylicum*. *Enzyme and microbial technology*, **38**, 937-943.
223. Lin, P.P., Jaeger, A.J., Wu, T.-Y., Xu, S.C., Lee, A.S., Gao, F., Chen, P.-W. and Liao, J.C. (2018) Construction and evolution of an *Escherichia coli* strain relying on nonoxidative glycolysis for sugar catabolism. *Proceedings of the National Academy of Sciences*, **115**, 3538-3546.
224. Singh, S., Bhadani, A. and Singh, B. (2007) Synthesis of wax esters from  $\alpha$ -olefins. *Industrial & engineering chemistry research*, **46**, 2672-2676.
225. Layton, D.S. and Trinh, C.T. (2016) Microbial synthesis of a branched-chain ester platform from organic waste carboxylates. *Metabolic Engineering Communications*, **3**, 245-251.
226. Chen, Y. and Nielsen, J. (2016) Biobased organic acids production by metabolically engineered microorganisms. *Current opinion in biotechnology*, **37**, 165-172.
227. Stincone, A., Prigione, A., Cramer, T., Wamelink, M.M., Campbell, K., Cheung, E., Olin-Sandoval, V., Grüning, N.M., Krüger, A. and Tauqeer Alam, M. (2015) The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biological Reviews*, **90**, 927-963.
228. Rutkis, R., Kalnenieks, U., Stalidzans, E. and Fell, D.A. (2013) Kinetic modelling of the *Zymomonas mobilis* Entner–Doudoroff pathway: insights into control and functionality. *Microbiology*, **159**, 2674-2689.
229. Sudarsan, S., Dethlefsen, S., Blank, L.M., Siemann-Herzberg, M. and Schmid, A. (2014) The functional structure of central carbon metabolism in *Pseudomonas putida* KT2440. *Applied and environmental microbiology*, **80**, 5292-5303.
230. Barney, B.M. (2014) The sweet smell of biosynthesis. *Nature chemical biology*, **10**, 246-247.
231. Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y. and Nielsen, J. (2015) Microbial acetyl-CoA metabolism and metabolic engineering. *Metabolic engineering*, **28**, 28-42.
232. Rossum, V.H.M., Kozak, B.U., Pronk, J.T. and van Maris, A.J. (2016) Engineering cytosolic acetyl-coenzyme A supply in *Saccharomyces cerevisiae*: pathway stoichiometry, free-energy conservation and redox-cofactor balancing. *Metabolic engineering*, **36**, 99-115.

## References

233. Knappe, J. and Sawers, G. (1990) A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. *FEMS microbiology reviews*, **6**, 383-398.
234. Bocanegra, J.A., Scrutton, N.S. and Perham, R.N. (1993) Creation of an NADP-dependent pyruvate dehydrogenase multienzyme complex by protein engineering. *Biochemistry*, **32**, 2737-2740.
235. Inui, H., Miyatake, K., Nakano, Y. and Kitaoka, S. (1989) Pyruvate: NADP+ oxidoreductase from *Euglena gracilis*: the kinetic properties of the enzyme. *Archives of biochemistry and biophysics*, **274**, 434-442.
236. Miyagi, H., Kawai, S. and Murata, K. (2009) Two sources of mitochondrial NADPH in the yeast *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, **284**, 7553-7560.
237. Zaldivar, J., Nielsen, J. and Olsson, L. (2001) Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Applied microbiology and biotechnology*, **56**, 17-34.
238. Kozak, B.U., van Rossum, H.M., Benjamin, K.R., Wu, L., Daran, J.-M.G., Pronk, J.T. and van Maris, A.J. (2014) Replacement of the *Saccharomyces cerevisiae* acetyl-CoA synthetases by alternative pathways for cytosolic acetyl-CoA synthesis. *Metabolic engineering*, **21**, 46-59.
239. Rossum, V.H.M., Kozak, B.U., Niemeijer, M.S., Dykstra, J.C., Luttik, M.A., Daran, J.-M.G., Van Maris, A.J. and Pronk, J.T. (2016) Requirements for carnitine shuttle-mediated translocation of mitochondrial acetyl moieties to the yeast cytosol. *MBio*, **7**, e00520-00516.
240. Jong, D.B.W., Shi, S., Siewers, V. and Nielsen, J. (2014) Improved production of fatty acid ethyl esters in *Saccharomyces cerevisiae* through up-regulation of the ethanol degradation pathway and expression of the heterologous phosphoketolase pathway. *Microbial cell factories*, **13**, 39.
241. Bergman, A., Siewers, V., Nielsen, J. and Chen, Y. (2016) Functional expression and evaluation of heterologous phosphoketolases in *Saccharomyces cerevisiae*. *Amb Express*, **6**, 1-13.
242. San, K.-Y., Bennett, G.N., Berrios-Rivera, S.J., Vadali, R.V., Yang, Y.-T., Horton, E., Rudolph, F.B., Sariyar, B. and Blackwood, K. (2002) Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. *Metabolic engineering*, **4**, 182-192.
243. Beld, J., Lee, D.J. and Burkart, M.D. (2015) Fatty acid biosynthesis revisited: structure elucidation and metabolic engineering. *Molecular BioSystems*, **11**, 38-59.
244. Duan, Y., Zhu, Z., Cai, K., Tan, X. and Lu, X. (2011) De novo biosynthesis of biodiesel by *Escherichia coli* in optimized fed-batch cultivation. *PLoS One*, **6**, e20265.
245. Pan, H., Zhang, L., Li, X. and Guo, D. (2017) Biosynthesis of the fatty acid isopropyl esters by engineered *Escherichia coli*. *Enzyme and Microbial Technology*, **102**, 49-52.
246. Rungtaphan, W. and Keasling, J.D. (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metabolic engineering*, **21**, 103-113.
247. Valle-Rodríguez, J.O., Shi, S., Siewers, V. and Nielsen, J. (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid ethyl esters, an advanced biofuel, by eliminating non-essential fatty acid utilization pathways. *Applied Energy*, **115**, 226-232.

248. Xu, P., Qiao, K., Ahn, W.S. and Stephanopoulos, G. (2016) Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proceedings of the National Academy of Sciences*, **113**, 10848-10853.
249. Qiao, K., Wasylenko, T.M., Zhou, K., Xu, P. and Stephanopoulos, G. (2017) Lipid production in *Yarrowia lipolytica* is maximized by engineering cytosolic redox metabolism. *Nature biotechnology*, **35**, 173-177.
250. Dellomonaco, C., Clomburg, J.M., Miller, E.N. and Gonzalez, R. (2011) Engineered reversal of the  $\beta$ -oxidation cycle for the synthesis of fuels and chemicals. *Nature*, **476**, 355-359.
251. Houten, S.M. and Wanders, R.J. (2010) A general introduction to the biochemistry of mitochondrial fatty acid  $\beta$ -oxidation. *Journal of inherited metabolic disease*, **33**, 469-477.
252. Kallscheuer, N., Polen, T., Bott, M. and Marienhagen, J. (2017) Reversal of  $\beta$ -oxidative pathways for the microbial production of chemicals and polymer building blocks. *Metabolic engineering*, **42**, 33-42.
253. Clomburg, J.M., Contreras, S.C., Chou, A., Siegel, J.B. and Gonzalez, R. (2018) Combination of type II fatty acid biosynthesis enzymes and thiolases supports a functional  $\beta$ -oxidation reversal. *Metabolic engineering*, **45**, 11-19.
254. Lian, J. and Zhao, H. (2015) Reversal of the  $\beta$ -oxidation cycle in *Saccharomyces cerevisiae* for production of fuels and chemicals. *ACS synthetic biology*, **4**, 332-341.
255. Clomburg, J.M., Blankschien, M.D., Vick, J.E., Chou, A., Kim, S. and Gonzalez, R. (2015) Integrated engineering of  $\beta$ -oxidation reversal and  $\omega$ -oxidation pathways for the synthesis of medium chain  $\omega$ -functionalized carboxylic acids. *Metabolic engineering*, **28**, 202-212.
256. Kim, S., Cheong, S. and Gonzalez, R. (2016) Engineering *Escherichia coli* for the synthesis of short-and medium-chain  $\alpha$ ,  $\beta$ -unsaturated carboxylic acids. *Metabolic engineering*, **36**, 90-98.
257. Hazelwood, L.A., Daran, J.-M., van Maris, A.J., Pronk, J.T. and Dickinson, J.R. (2008) The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Applied and environmental microbiology*, **74**, 2259-2266.
258. Mooney, B.P., Miernyk, J.A. and Randall, D.D. (2002) The complex fate of  $\alpha$ -ketoacids. *Annual review of plant biology*, **53**, 357-375.
259. Vadali, R.V., Bennett, G.N. and San, K.Y. (2004) Enhanced isoamyl acetate production upon manipulation of the acetyl-CoA node in *Escherichia coli*. *Biotechnology progress*, **20**, 692-697.
260. Lin, J.-L., Zhu, J. and Wheeldon, I. (2017) Synthetic protein scaffolds for biosynthetic pathway colocalization on lipid droplet membranes. *ACS synthetic biology*, **6**, 1534-1544.
261. Layton, D.S. and Trinh, C.T. (2014) Engineering modular ester fermentative pathways in *Escherichia coli*. *Metabolic engineering*, **26**, 77-88.
262. Lee, J.-W. and Trinh, C.T. (2018) De novo microbial biosynthesis of a lactate ester platform. *BioRxiv*, 498576.
263. Nuland, V.Y.M., Eggink, G. and Weusthuis, R.A. (2017) Combination of ester biosynthesis and  $\omega$ -oxidation for production of mono-ethyl dicarboxylic acids and di-ethyl esters in a whole-cell biocatalytic setup with *Escherichia coli*. *Microbial cell factories*, **16**, 1-9.

## References

264. Teo, W.S., Ling, H., Yu, A.-Q. and Chang, M.W. (2015) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid short-and branched-chain alkyl esters biodiesel. *Biotechnology for biofuels*, **8**, 177.
265. Guo, D., Zhu, J., Deng, Z. and Liu, T. (2014) Metabolic engineering of *Escherichia coli* for production of fatty acid short-chain esters through combination of the fatty acid and 2-keto acid pathways. *Metabolic engineering*, **22**, 69-75.
266. Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S.B. and Keasling, J.D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, **463**, 559-562.
267. Lehtinen, T., Efimova, E., Santala, S. and Santala, V. (2018) Improved fatty aldehyde and wax ester production by overexpression of fatty acyl-CoA reductases. *Microbial cell factories*, **17**, 1-10.
268. Santala, S., Efimova, E., Koskinen, P., Karp, M.T. and Santala, V. (2014) Rewiring the wax ester production pathway of *Acinetobacter baylyi* ADP1. *ACS synthetic biology*, **3**, 145-151.
269. Becerra, M., Cerdán, M.E. and González-Siso, M.I. (2015) Biobutanol from cheese whey. *Microbial cell factories*, **14**, 1-15.
270. Akhtar, M.K., Turner, N.J. and Jones, P.R. (2013) Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proceedings of the National Academy of Sciences*, **110**, 87-92.
271. Wang, G., Xiong, X., Ghogare, R., Wang, P., Meng, Y. and Chen, S. (2016) Exploring fatty alcohol-producing capability of *Yarrowia lipolytica*. *Biotechnology for biofuels*, **9**, 1-10.
272. Peralta-Yahya, P.P., Zhang, F., Del Cardayre, S.B. and Keasling, J.D. (2012) Microbial engineering for the production of advanced biofuels. *Nature*, **488**, 320-328.
273. Yuan, J., Mishra, P. and Ching, C.B. (2016) Metabolically engineered *Saccharomyces cerevisiae* for branched-chain ester productions. *Journal of biotechnology*, **239**, 90-97.
274. Atsumi, S., Hanai, T. and Liao, J.C. (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *nature*, **451**, 86.
275. Matsuda, F., Ishii, J., Kondo, T., Ida, K., Tezuka, H. and Kondo, A. (2013) Increased isobutanol production in *Saccharomyces cerevisiae* by eliminating competing pathways and resolving cofactor imbalance. *Microbial cell factories*, **12**, 119.
276. Shen, C.R., Lan, E.I., Dekishima, Y., Baez, A., Cho, K.M. and Liao, J.C. (2011) Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Applied and environmental microbiology*, **77**, 2905-2915.
277. Wu, J., Zhang, X., Xia, X. and Dong, M. (2017) A systematic optimization of medium chain fatty acid biosynthesis via the reverse beta-oxidation cycle in *Escherichia coli*. *Metabolic engineering*, **41**, 115-124.
278. Jong, D.B.W., Shi, S., Valle-Rodríguez, J.O., Siewers, V. and Nielsen, J. (2015) Metabolic pathway engineering for fatty acid ethyl ester production in *Saccharomyces cerevisiae* using stable chromosomal integration. *Journal of industrial microbiology & biotechnology*, **42**, 477-486.
279. Röttig, A., Zurek, P.J. and Steinbüchel, A. (2015) Assessment of bacterial acyltransferases for an efficient lipid production in metabolically engineered strains of *E. coli*. *Metabolic engineering*, **32**, 195-206.

280. Wierzbicki, M., Niraula, N., Yarrabothula, A., Layton, D.S. and Trinh, C.T. (2016) Engineering an *Escherichia coli* platform to synthesize designer biodiesels. *Journal of biotechnology*, **224**, 27-34.
281. Garcia, S. and Trinh, C.T. (2019) Multiobjective strain design: a framework for modular cell engineering. *Metabolic engineering*, **51**, 110-120.
282. Trinh, C.T., Liu, Y. and Conner, D.J. (2015) Rational design of efficient modular cells. *Metabolic engineering*, **32**, 220-231.
283. Ezeji, T.C., Karcher, P.M., Qureshi, N. and Blaschek, H.P. (2005) Improving performance of a gas stripping-based recovery system to remove butanol from *Clostridium beijerinckii* fermentation. *Bioprocess and biosystems engineering*, **27**, 207-214.
284. Pereira, J.P., Verheijen, P.J. and Straathof, A.J. (2016) Growth inhibition of *S. cerevisiae*, *B. subtilis*, and *E. coli* by lignocellulosic and fermentation products. *Applied microbiology and biotechnology*, **100**, 9069-9080.
285. Vázquez, J.A., Durán, A., Rodríguez-Amado, I., Prieto, M.A., Rial, D. and Murado, M.A. (2011) Evaluation of toxic effects of several carboxylic acids on bacterial growth by toxicodynamic modelling. *Microbial Cell Factories*, **10**, 1-11.
286. Casey, G.P. and Ingledew, W.M. (1986) Ethanol tolerance in yeasts. *CRC Critical Reviews in Microbiology*, **13**, 219-280.
287. Ghareib, M., Youssef, K. and Khalil, A. (1988) Ethanol tolerance of *Saccharomyces cerevisiae* and its relationship to lipid content and composition. *Folia microbiologica*, **33**, 447-452.
288. Harnisch, M., Möckel, H. and Schulze, G. (1983) Relationship between log *P*<sub>ow</sub>, shake-flask values and capacity factors derived from reversed-phase high-performance liquid chromatography for *n*-alkylbenzenes and some oecd reference substances. *Journal of Chromatography A*, **282**, 315-332.
289. Heipieper, H.J., Weber, F.J., Sikkema, J., Keweloh, H. and de Bont, J.A. (1994) Mechanisms of resistance of whole cells to toxic organic solvents. *Trends in Biotechnology*, **12**, 409-415.
290. Inoue, A. and Horikoshi, K. (1991) Estimation of solvent-tolerance of bacteria by the solvent parameter log *P*. *Journal of Fermentation and Bioengineering*, **71**, 194-196.
291. Laane, C., Boeren, S., Vos, K. and Veeger, C. (1987) Rules for optimization of biocatalysis in organic solvents. *Biotechnology and Bioengineering*, **30**, 81-87.
292. Straathof, A. (2003) Auxiliary phase guidelines for microbial biotransformations of toxic substrate into toxic product. *Biotechnology Progress*, **19**, 755-762.
293. Bar, R. and Gainer, J.L. (1987) Acid fermentation in water-organic solvent two-liquid phase systems. *Biotechnology progress*, **3**, 109-114.
294. Wilbanks, B. and Trinh, C.T. (2017) Comprehensive characterization of toxicity of fermentative metabolites on microbial growth. *Biotechnology for biofuels*, **10**, 1-11.
295. Urit, T., Löser, C., Wunderlich, M. and Bley, T. (2011) Formation of ethyl acetate by *Kluyveromyces marxianus* on whey: studies of the ester stripping. *Bioprocess and biosystems engineering*, **34**, 547-559.
296. Vitha, M. and Carr, P.W. (2006) The chemical interpretation and practice of linear solvation energy relationships in chromatography. *Journal of chromatography A*, **1126**, 143-194.

## References

297. Labs, A.
298. Stovall, D.M., Acree Jr, W.E. and Abraham, M.H. (2005) Solubility of 9-fluorenone, thianthrene and xanthene in organic solvents. *Fluid phase equilibria*, **232**, 113-121.
299. Qureshi, N., Dien, B., Liu, S., Saha, B., Cotta, M., Hughes, S. and Hector, R. (2012) Genetically engineered *Escherichia coli* FBR5: Part II. Ethanol production from xylose and simultaneous product recovery. *Biotechnology progress*, **28**, 1179-1185.
300. Taylor, F., Kurantz, M.J., Goldberg, N. and Craig, J.C. (1998) Kinetics of continuous fermentation and stripping of ethanol. *Biotechnology letters*, **20**, 67-72.
301. Sander, R. (2015) Compilation of Henry's law constants (version 4.0) for water as solvent. *Atmospheric Chemistry and Physics*, **15**, 4399-4981.
302. Huang, Y. and Yang, S.T. (1998) Acetate production from whey lactose using co-immobilized cells of homolactic and homoacetic bacteria in a fibrous-bed bioreactor. *Biotechnology and bioengineering*, **60**, 498-507.
303. Vidra, A. and Németh, Á. (2018) Bio-produced acetic acid: a review. *Periodica Polytechnica Chemical Engineering*, **62**, 245-256.
304. Groot, W., Van der Lans, R. and Luyben, K.C.A. (1992) Technologies for butanol recovery integrated with fermentations. *Process Biochemistry*, **27**, 61-75.
305. Xue, C., Zhao, J.-B., Chen, L.-J., Bai, F.-W., Yang, S.-T. and Sun, J.-X. (2014) Integrated butanol recovery for an advanced biofuel: current state and prospects. *Applied microbiology and biotechnology*, **98**, 3463-3474.
306. Oudshoorn, A., Van Der Wielen, L.A. and Straathof, A.J. (2009) Assessment of options for selective 1-butanol recovery from aqueous solution. *Industrial & Engineering Chemistry Research*, **48**, 7325-7336.
307. Fujii, T., Narikawa, T., Sumisa, F., Arisawa, A., Takeda, K. and Kato, J. (2006) Production of  $\alpha$ ,  $\omega$ -alkanediols using *Escherichia coli* expressing a cytochrome P450 from *Acinetobacter* sp. OC4. *Bioscience, biotechnology, and biochemistry*, **70**, 1379-1385.
308. Gudiminch, R.K., Randall, C., Opperman, D.J., Olaofe, O.A., Harrison, S.T., Albertyn, J. and Smit, M.S. (2012) Whole-cell hydroxylation of n-octane by *Escherichia coli* strains expressing the CYP153A6 operon. *Applied microbiology and biotechnology*, **96**, 1507-1516.
309. Malca, S.H., Scheps, D., Kühnel, L., Venegas-Venegas, E., Seifert, A., Nestl, B.M. and Hauer, B. (2012) Bacterial CYP153A monooxygenases for the synthesis of omega-hydroxylated fatty acids. *Chemical communications*, **48**, 5115-5117.
310. Scheps, D., Malca, S.H., Hoffmann, H., Nestl, B.M. and Hauer, B. (2011) Regioselective  $\omega$ -hydroxylation of medium-chain n-alkanes and primary alcohols by CYP153 enzymes from *Mycobacterium marinum* and *Polaromonas* sp. strain JS666. *Organic & biomolecular chemistry*, **9**, 6727-6733.
311. Julsing, M.K., Schrewe, M., Cornelissen, S., Hermann, I., Schmid, A. and Bühler, B. (2012) Outer membrane protein AlkL boosts biocatalytic oxyfunctionalization of hydrophobic substrates in *Escherichia coli*. *Applied and environmental microbiology*, **78**, 5724-5733.
312. Schrewe, M., Magnusson, A.O., Willrodt, C., Bühler, B. and Schmid, A. (2011) Kinetic Analysis of Terminal and Unactivated C-H Bond Oxyfunctionalization in Fatty Acid Methyl Esters by Monooxygenase-Based Whole-Cell Biocatalysis. *Advanced Synthesis & Catalysis*, **353**, 3485-3495.

313. Schrewe, M., Julsing, M.K., Lange, K., Czarnotta, E., Schmid, A. and Bühler, B. (2014) Reaction and catalyst engineering to exploit kinetically controlled whole-cell multistep biocatalysis for terminal FAME oxyfunctionalization. *Biotechnology and bioengineering*, **111**, 1820-1830.
314. Jiménez-Díaz, L., Caballero, A., Pérez-Hernández, N. and Segura, A. (2017) Microbial alkane production for jet fuel industry: motivation, state of the art and perspectives. *Microbial biotechnology*, **10**, 103-124.
315. Desai, L.V., Hull, K.L. and Sanford, M.S. (2004) Palladium-catalyzed oxygenation of unactivated sp<sup>3</sup> C–H bonds. *Journal of the American Chemical Society*, **126**, 9542-9543.
316. Hashiguchi, B.G., Konnick, M.M., Bischof, S.M., Gustafson, S.J., Devarajan, D., Gunsalus, N., Ess, D.H. and Periana, R.A. (2014) Main-group compounds selectively oxidize mixtures of methane, ethane, and propane to alcohol esters. *Science*, **343**, 1232-1237.
317. Konnick, M.M., Hashiguchi, B.G., Devarajan, D., Boaz, N.C., Gunnoe, T.B., Groves, J.T., Gunsalus, N., Ess, D.H. and Periana, R.A. (2014) Selective CH functionalization of methane, ethane, and propane by a perfluoroarene iodine (III) complex. *Angewandte Chemie International Edition*, **53**, 10490-10494.
318. Nardini, M. and Dijkstra, B.W. (1999)  $\alpha/\beta$  Hydrolase fold enzymes: the family keeps growing. *Current opinion in structural biology*, **9**, 732-737.
319. Holmquist, M. (2000) Alpha beta-hydrolase fold enzymes structures, functions and mechanisms. *Current Protein and Peptide Science*, **1**, 209-235.
320. Bornscheuer, U.T. and Kazlauskas, R.J. (2006) *Hydrolases in organic synthesis: regio- and stereoselective biotransformations*. John Wiley & Sons.
321. Priyanka, P., Tan, Y., Kinsella, G.K., Henahan, G.T. and Ryan, B.J. (2019) Solvent stable microbial lipases: current understanding and biotechnological applications. *Biotechnology letters*, **41**, 203-220.
322. Lotti, M., Pleiss, J., Valero, F. and Ferrer, P. (2018) Enzymatic production of biodiesel: strategies to overcome methanol inactivation. *Biotechnology journal*, **13**, 1700155.
323. Cavaille-Lefebvre, D. and Combes, D. (1997) Lipase synthesis of short-chain flavour thioesters in solvent-free medium. *Biocatalysis and Biotransformation*, **15**, 265-279.
324. Zaks, A. and Klivanov, A.M. (1985) Enzyme-catalyzed processes in organic solvents. *Proceedings of the National Academy of Sciences*, **82**, 3192-3196.
325. Gapes, J.R., Nimcevic, D. and Friedl, A. (1996) Long-term continuous cultivation of *clostridium beijerinckii* in a two-stage chemostat with on-line solvent removal. *Appl. Environ. Microbiol.*, **62**, 3210-3219.
326. Entian, K.-D. and Kötter, P. (2007) 25 yeast genetic strain and plasmid collections. *Methods in microbiology*, **36**, 629-666.
327. Scotcher, M.C. and Bennett, G.N. (2005) SpoIIE regulates sporulation but does not directly affect solventogenesis in *Clostridium acetobutylicum* ATCC 824. *Journal of bacteriology*, **187**, 1930-1936.
328. Mohanraju, P., van der Oost, J., Jinek, M. and Swarts, D.C. (2018) Heterologous Expression and Purification of CRISPR-Cas12a/Cpf1. *Bio-protocol*, **8**.
329. Cheng, T., Zhao, Y., Li, X., Lin, F., Xu, Y., Zhang, X., Li, Y., Wang, R. and Lai, L. (2007) Computation of octanol–water partition coefficients by guiding an additive model with knowledge. *Journal of chemical information and modeling*, **47**, 2140-2148.



## References

330. Liu, S.-Q., Holland, R. and Crow, V. (2003) Ester synthesis in an aqueous environment by *Streptococcus thermophilus* and other dairy lactic acid bacteria. *Applied microbiology and biotechnology*, **63**, 81-88.
331. Liu, S.-Q., Baker, K., Bennett, M., Holland, R., Norris, G. and Crow, V. (2004) Characterisation of esterases of *Streptococcus thermophilus* ST1 and *Lactococcus lactis* subsp. *cremoris* B1079 as alcohol acyltransferases. *International dairy journal*, **14**, 865-870.
332. Müller, H., Becker, A.-K., Palm, G.J., Berndt, L., Badenhorst, C.P., Godehard, S.P., Reisky, L., Lammers, M. and Bornscheuer, U. (2020) Sequence-Based Prediction of Promiscuous Acyltransferase Activity in Hydrolases. *Angewandte Chemie*.
333. Rizzi, M., Stylos, P., Riek, A. and Reuss, M. (1992) A kinetic study of immobilized lipase catalysing the synthesis of isoamyl acetate by transesterification in n-hexane. *Enzyme and microbial technology*, **14**, 709-714.
334. Chulalaksananukul, W., Condoret, J.-S. and Combes, D. (1992) Kinetics of geranyl acetate synthesis by lipase-catalysed transesterification in n-hexane. *Enzyme and microbial technology*, **14**, 293-298.
335. El-Rassy, H., Perrard, A. and Pierre, A.C. (2004) Application of lipase encapsulated in silica aerogels to a transesterification reaction in hydrophobic and hydrophilic solvents: Bi-Bi Ping-Pong kinetics. *Journal of Molecular Catalysis B: Enzymatic*, **30**, 137-150.
336. Cheng, Y.-C. and Tsai, S.-W. (2003) Effects of water activity and alcohol concentration on the kinetic resolution of lipase-catalyzed acyl transfer in organic solvents. *Enzyme and microbial technology*, **32**, 362-368.
337. Varma, M.N. and Madras, G. (2008) Kinetics of synthesis of butyl butyrate by esterification and transesterification in supercritical carbon dioxide. *Journal of Chemical Technology & Biotechnology: International Research in Process, Environmental & Clean Technology*, **83**, 1135-1144.
338. Yadav, G.D. and Lathi, P.S. (2003) Kinetics and mechanism of synthesis of butyl isobutyrate over immobilised lipases. *Biochemical Engineering Journal*, **16**, 245-252.
339. Bohnenkamp, A.C., Kruis, A.J., Mars, A.E., Wijffels, R.H., van der Oost, J., Kengen, S.W. and Weusthuis, R.A. (2020) Multilevel optimisation of anaerobic ethyl acetate production in engineered *Escherichia coli*. *Biotechnology for biofuels*, **13**, 1-14.
340. Hoydonckx, H.E., De Vos, D.E., Chavan, S.A. and Jacobs, P.A. (2004) Esterification and transesterification of renewable chemicals. *Topics in Catalysis*, **27**, 83-96.
341. Cha, H.-J., Park, J.-B. and Park, S. (2019) Esterification of Secondary Alcohols and Multi-hydroxyl Compounds by *Candida antarctica* Lipase B and Subtilisin. *Biotechnology and bioprocess engineering*, **24**, 41-47.
342. Mathews, I., Soltis, M., Saldajeno, M., Ganshaw, G., Sala, R., Weyler, W., Cervin, M.A., Whited, G. and Bott, R. (2007) Structure of a novel enzyme that catalyzes acyl transfer to alcohols in aqueous conditions. *Biochemistry*, **46**, 8969-8979.
343. de-Leeuw, N., Torrelo, G., Bisterfeld, C., Resch, V., Mestrom, L., Straulino, E., van der Weel, L. and Hanefeld, U. (2018) Ester synthesis in water: *Mycobacterium smegmatis* acyl transferase for kinetic resolutions. *Advanced Synthesis & Catalysis*, **360**, 242-249.
344. Weber, N., Klein, E. and Mukherjee, K. (1999) Long-chain acyl thioesters prepared by solvent-free thioesterification and transthioesterification catalysed by microbial lipases. *Applied microbiology and biotechnology*, **51**, 401-404.

345. Hedfors, C., Hult, K. and Martinelle, M. (2010) Lipase chemoselectivity towards alcohol and thiol acyl acceptors in a transacylation reaction. *Journal of Molecular Catalysis B: Enzymatic*, **66**, 120-123.
346. Zhou, N., Shen, L., Dong, Z., Shen, J., Du, L. and Luo, X. (2018) Enzymatic Synthesis of Thioesters from Thiols and Vinyl Esters in a Continuous-Flow Microreactor. *Catalysts*, **8**, 249.
347. Lodé, T. (2012) For quite a few chromosomes more: the origin of eukaryotes.... *Journal of molecular biology*, **423**, 135-142.
348. Martin, W. and Koonin, E.V. (2006) Introns and the origin of nucleus–cytosol compartmentalization. *Nature*, **440**, 41-45.
349. Hammer, S.K. and Avalos, J.L. (2017) Harnessing yeast organelles for metabolic engineering. *Nature chemical biology*, **13**, 823.
350. Faivre, D. and Schuler, D. (2008) Magnetotactic bacteria and magnetosomes. *Chemical reviews*, **108**, 4875-4898.
351. Blakemore, R. (1975) Magnetotactic bacteria. *Science*, **190**, 377-379.
352. Fernández, L.C., Mesman, R. and van Niftrik, L. (2020), *Bacterial Organelles and Organelle-like Inclusions*. Springer, pp. 107-123.
353. Jetten, M., Cirpus, I., Kartal, B., van Niftrik, L., Van De Pas-Schoonen, K., Sliemers, O., Haaijer, S., Van der Star, W., Schmid, M. and van de Vossenberg, J. (2005) 1994–2004: 10 years of research on the anaerobic oxidation of ammonium. *Biochemical Society Transactions*, **33**, 119-123.
354. Walsby, A.E. (1994) Gas vesicles. *Microbiological reviews*, **58**, 94-144.
355. Pfeifer, F. (2012) Distribution, formation and regulation of gas vesicles. *Nature Reviews Microbiology*, **10**, 705-715.
356. Gabashvili, A.N., Chmelyuk, N.S., Efremova, M.V., Malinovskaya, J.A., Semkina, A.S. and Abakumov, M.A. (2020) Encapsulins—bacterial protein nanocompartments: Structure, properties, and application. *Biomolecules*, **10**, 966.
357. Jones, J.A. and Giessen, T.W. (2021) Advances in encapsulin nanocompartment biology and engineering. *Biotechnology and bioengineering*, **118**, 491-505.
358. Chowdhury, C., Sinha, S., Chun, S., Yeates, T.O. and Bobik, T.A. (2014) Diverse bacterial microcompartment organelles. *Microbiology and Molecular Biology Reviews*, **78**, 438-468.
359. Kerfeld, C.A., Aussignargues, C., Zarzycki, J., Cai, F. and Sutter, M. (2018) Bacterial microcompartments. *Nature Reviews Microbiology*, **16**, 277.
360. Faulkner, M., Szabó, I., Weetman, S.L., Sicard, F., Huber, R.G., Bond, P.J., Rosta, E. and Liu, L.-N. (2020) Molecular simulations unravel the molecular principles that mediate selective permeability of carboxysome shell protein. *Scientific reports*, **10**, 1-14.
361. Park, J., Chun, S., Bobik, T.A., Houk, K.N. and Yeates, T.O. (2017) Molecular dynamics simulations of selective metabolite transport across the propanediol bacterial microcompartment shell. *The journal of physical chemistry B*, **121**, 8149-8154.
362. Herring, T.I., Harris, T.N., Chowdhury, C., Mohanty, S.K. and Bobik, T.A. (2018) A bacterial microcompartment is used for choline fermentation by *Escherichia coli* 536. *Journal of bacteriology*, **200**.

## References

363. Sutter, M., Melnicki, M.R., Schulz, F., Woyke, T. and Kerfeld, C.A. (2021) A Catalog of the Diversity and Ubiquity of Metabolic Organelles in Bacteria. *bioRxiv*.
364. Stewart, A.M., Stewart, K.L., Yeates, T.O. and Bobik, T.A. (2021) Advances in the World of Bacterial Microcompartments. *Trends in Biochemical Sciences*.
365. Lawrence, A.D., Frank, S., Newnham, S., Lee, M.J., Brown, I.R., Xue, W.-F., Rowe, M.L., Mulvihill, D.P., Prentice, M.B. and Howard, M.J. (2014) Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS synthetic biology*, **3**, 454-465.
366. Liang, M., Frank, S., Lünsdorf, H., Warren, M.J. and Prentice, M.B. (2017) Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in *E. coli*. *Biotechnology journal*, **12**, 1600415.
367. Yung, M.C., Bourguet, F.A., Carpenter, T.S. and Coleman, M.A. (2017) Re-directing bacterial microcompartment systems to enhance recombinant expression of lysis protein E from bacteriophage  $\phi$ X174 in *Escherichia coli*. *Microbial cell factories*, **16**, 1-17.
368. Huber, I., Palmer, D.J., Ludwig, K.N., Brown, I.R., Warren, M.J. and Frunzke, J. (2017) Construction of recombinant Pdu metabolosome shells for small molecule production in *Corynebacterium glutamicum*. *ACS synthetic biology*, **6**, 2145-2156.
369. Jakobson, C.M., Kim, E.Y., Slininger, M.F., Chien, A. and Tullman-Ercek, D. (2015) Localization of proteins to the 1, 2-propanediol utilization microcompartment by non-native signal sequences is mediated by a common hydrophobic motif. *Journal of Biological Chemistry*, **290**, 24519-24533.
370. Nichols, T.M., Kennedy, N.W. and Tullman-Ercek, D. (2020) A genomic integration platform for heterologous cargo encapsulation in 1, 2-propanediol utilization bacterial microcompartments. *Biochemical Engineering Journal*, **156**, 107496.
371. Avalos, J.L., Fink, G.R. and Stephanopoulos, G. (2013) Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nature biotechnology*, **31**, 335-341.
372. EFSA. (2012).
373. FAO. (1997).
374. DOW. (2002).
375. Diallo, M., Simons, A.D., van der Wal, H., Collas, F., Houweling-Tan, B., Kengen, S.W. and López-Contreras, A.M. (2019) L-Rhamnose metabolism in *Clostridium beijerinckii* strain DSM 6423. *Applied and environmental microbiology*, **85**.
376. Patinios, C., Lanza, L., Corino, I., Franssen, M.C., Van der Oost, J., Weusthuis, R.A. and Kengen, S.W. (2020) Eat1-Like Alcohol Acyl Transferases From Yeasts Have High Alcoholysis and Thiolysis Activity. *Frontiers in microbiology*, **11**.
377. Batianis, C., Kozaeva, E., Damalas, S.G., Martín-Pascual, M., Volke, D.C., Nikel, P.I. and Martins dos Santos, V.A. (2020) An expanded CRISPRi toolbox for tunable control of gene expression in *Pseudomonas putida*. *Microbial Biotechnology*, **13**, 368-385.
378. Vallenet, D., Calteau, A., Cruveiller, S., Gachet, M., Lajus, A., Josso, A., Mercier, J., Renaux, A., Rollin, J. and Rouy, Z. (2017) MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic acids research*, **45**, D517-D528.

379. Vallenet, D., Engelen, S., Mornico, D., Cruveiller, S., Fleury, L., Lajus, A., Rouy, Z., Roche, D., Salvignol, G. and Scarpelli, C. (2009) MicroScope: a platform for microbial genome annotation and comparative genomics. *Database*, **2009**.
380. Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S. and Gautreau, G. (2020) MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Research*, **48**, D579-D589.
381. Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., Le Fèvre, F., Longin, C., Mornico, D. and Roche, D. (2013) MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic acids research*, **41**, D636-D647.
382. Médigue, C., Calteau, A., Cruveiller, S., Gachet, M., Gautreau, G., Josso, A., Lajus, A., Langlois, J., Pereira, H. and Planel, R. (2019) MicroScope—an integrated resource for community expertise of gene functions and comparative analysis of microbial genomic and metabolic data. *Briefings in bioinformatics*, **20**, 1071-1084.
383. Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., Lajus, A., Pascal, G., Scarpelli, C. and Medigue, C. (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic acids research*, **34**, 53-65.
384. Petit, E., LaTouf, W.G., Coppi, M.V., Warnick, T.A., Currie, D., Romashko, I., Deshpande, S., Haas, K., Alvelo-Maurosa, J.G. and Wardman, C. (2013) Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. *PLoS One*, **8**, e54337.
385. Fan, C., Cheng, S., Sinha, S. and Bobik, T.A. (2012) Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proceedings of the National Academy of Sciences*, **109**, 14995-15000.
386. Kinney, J.N., Salmeen, A., Cai, F. and Kerfeld, C.A. (2012) Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *Journal of Biological Chemistry*, **287**, 17729-17736.
387. Fan, C. and Bobik, T.A. (2011) The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. *Journal of bacteriology*, **193**, 5623-5628.
388. Fan, C., Cheng, S., Liu, Y., Escobar, C.M., Crowley, C.S., Jefferson, R.E., Yeates, T.O. and Bobik, T.A. (2010) Short N-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences*, **107**, 7509-7514.
389. Choudhary, S., Quin, M.B., Sanders, M.A., Johnson, E.T. and Schmidt-Dannert, C. (2012) Engineered protein nano-compartments for targeted enzyme localization. *PloS one*, **7**, e33342.
390. Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P. and Warren, M.J. (2010) Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Molecular cell*, **38**, 305-315.
391. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A. and Potapenko, A. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589.

## References

392. Liu, Y., Jorda, J., Yeates, T.O. and Bobik, T.A. (2015) The PduL phosphotransacylase is used to recycle coenzyme A within the Pdu microcompartment. *Journal of bacteriology*, **197**, 2392-2399.
393. Parsons, J.B., Lawrence, A.D., McLean, K.J., Munro, A.W., Rigby, S.E. and Warren, M.J. (2010) Characterisation of PduS, the pdu metabolosome corrin reductase, and evidence of substructural organisation within the bacterial microcompartment. *PLoS One*, **5**, e14009.
394. Cheng, S. and Bobik, T.A. (2010) Characterization of the PduS cobalamin reductase of *Salmonella enterica* and its role in the Pdu microcompartment. *Journal of bacteriology*, **192**, 5071-5080.
395. Gonzalez-Esquer, C.R., Shubitowski, T.B. and Kerfeld, C.A. (2015) Streamlined construction of the cyanobacterial CO<sub>2</sub>-fixing organelle via protein domain fusions for use in plant synthetic biology. *The Plant Cell*, **27**, 2637-2644.
396. Lee, M.S. and Tullman-Ercek, D. (2017) Practical considerations for the encapsulation of multi-enzyme cargos within the bacterial microcompartment for metabolic engineering. *Current Opinion in Systems Biology*, **5**, 16-22.
397. Farrell, C.M., Grossman, A.D. and Sauer, R.T. (2005) Cytoplasmic degradation of ssrA-tagged proteins. *Molecular microbiology*, **57**, 1750-1761.
398. Flynn, J.M., Levchenko, I., Seidel, M., Wickner, S.H., Sauer, R.T. and Baker, T.A. (2001) Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proceedings of the National Academy of Sciences*, **98**, 10584-10589.
399. Kim, E.Y. and Tullman-Ercek, D. (2014) A rapid flow cytometry assay for the relative quantification of protein encapsulation into bacterial microcompartments. *Biotechnology journal*, **9**, 348-354.
400. Jakobson, C.M., Slininger Lee, M.F. and Tullman-Ercek, D. (2017) De novo design of signal sequences to localize cargo to the 1, 2-propanediol utilization microcompartment. *Protein Science*, **26**, 1086-1092.
401. Nichols, T.M., Kennedy, N.W. and Tullman-Ercek, D. (2019) Cargo encapsulation in bacterial microcompartments: Methods and analysis. *Methods in enzymology*, **617**, 155-186.
402. Lavey, N. (2019) The Caseinolytic Protease P System in *Clostridium difficile*.
403. Lavey, N.P., Shadid, T., Ballard, J.D. and Duerfeldt, A.S. (2018) *Clostridium difficile* ClpP homologues are capable of uncoupled activity and exhibit different levels of susceptibility to acyldepsipeptide modulation. *ACS infectious diseases*, **5**, 79-89.
404. Bohnenkamp, A., Kruis, A., Nap, B., Nielsen, J., Mars, A., Wijffels, R., Jvd, O., Kengen, S. and Weusthuis, R. (2020) From Eat to trEat: Engineering the mitochondrial Eat1 enzyme for enhanced ethyl acetate production in *Escherichia coli*.
405. Ferlez, B., Sutter, M. and Kerfeld, C.A. (2019) Glycyl radical enzyme-associated microcompartments: redox-replete bacterial organelles. *MBio*, **10**, e02327-02318.
406. Wang, Y., Zhang, Z.-T., Seo, S.-O., Lynn, P., Lu, T., Jin, Y.-S. and Blaschek, H.P. (2016) Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable “clean” mutant selection in *Clostridium beijerinckii* as an example. *ACS synthetic biology*, **5**, 721-732.
407. McAllister, K.N. and Sorg, J.A. (2019) CRISPR genome editing systems in the genus *Clostridium*: a timely advancement. *Journal of bacteriology*, **201**, e00219-00219.

408. Li, Q., Chen, J., Minton, N.P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X. and Yang, J. (2016) CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnology journal*, **11**, 961-972.
409. Li, Q., Seys, F.M., Minton, N.P., Yang, J., Jiang, Y., Jiang, W. and Yang, S. (2019) CRISPR–Cas9D10A nickase-assisted base editing in the solvent producer *Clostridium beijerinckii*. *Biotechnology and bioengineering*, **116**, 1475-1483.
410. Wang, Y., Zhang, Z.-T., Seo, S.-O., Choi, K., Lu, T., Jin, Y.-S. and Blaschek, H.P. (2015) Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *Journal of biotechnology*, **200**, 1-5.
411. Patinios, C., Creutzburg, S., Arifah, A., Perez, B., Ingham, C., Kengen, S., van der Oost, J. and Staals, R. (2021) SIBR-Cas enables host-independent and universal CRISPR genome engineering in bacteria. *bioRxiv*.
412. Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P. and Minton, N.P. (2007) The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *Journal of microbiological methods*, **70**, 452-464.
413. Heap, J.T., Cartman, S.T., Kuehne, S.A., Cooksley, C. and Minton, N.P. (2010), *Clostridium difficile*. Springer, pp. 165-182.
414. Wang, Y., Zhang, Z.T., Seo, S.O., Lynn, P., Lu, T., Jin, Y.S. and Blaschek, H.P. (2016) Gene transcription repression in *Clostridium beijerinckii* using CRISPR-dCas9. *Biotechnology and bioengineering*, **113**, 2739-2743.
415. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *science*, **337**, 816-821.
416. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J. and Regev, A. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759-771.
417. Oultram, J., Loughlin, M., Swinfield, T., Brehm, J., Thompson, D. and Minton, N. (1988) Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. *FEMS microbiology letters*, **56**, 83-88.
418. Diallo, M., Hocq, R., Collas, F., Chartier, G., Wasels, F., Wijaya, H.S., Werten, M.W., Wolbert, E.J., Kengen, S.W. and van der Oost, J. (2019) Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*. *Methods*.
419. Claassens, N.J., Siliakus, M.F., Spaans, S.K., Creutzburg, S.C., Nijse, B., Schaap, P.J., Quax, T.E. and Van Der Oost, J. (2017) Improving heterologous membrane protein production in *Escherichia coli* by combining transcriptional tuning and codon usage algorithms. *PLoS One*, **12**, e0184355.
420. Clarke, T.F. and Clark, P.L. (2008) Rare codons cluster. *PloS one*, **3**, e3412.
421. McAllister, K.N., Bouillaut, L., Kahn, J.N., Self, W.T. and Sorg, J.A. (2017) Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Scientific reports*, **7**, 1-12.
422. Wang, S., Hong, W., Dong, S., Zhang, Z.-T., Zhang, J., Wang, L. and Wang, Y. (2018) Genome engineering of *Clostridium difficile* using the CRISPR-Cas9 system. *Clinical Microbiology and Infection*, **24**, 1095-1099.

## References

423. Wang, S., Dong, S., Wang, P., Tao, Y. and Wang, Y. (2017) Genome editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 system. *Applied and environmental microbiology*, **83**, e00233-00217.
424. Nariya, H., Miyata, S., Kuwahara, T. and Okabe, A. (2011) Development and characterization of a xylose-inducible gene expression system for *Clostridium perfringens*. *Applied and environmental microbiology*, **77**, 8439-8441.
425. Müh, U., Pannullo, A.G., Weiss, D.S. and Ellermeier, C.D. (2019) A xylose-inducible expression system and a CRISPR interference plasmid for targeted knockdown of gene expression in *Clostridioides difficile*. *Journal of bacteriology*, **201**, e00711-00718.
426. Ravagnani, A., Jennert, K.C., Steiner, E., Grünberg, R., Jefferies, J.R., Wilkinson, S.R., Young, D.I., Tidswell, E.C., Brown, D.P. and Youngman, P. (2000) Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Molecular microbiology*, **37**, 1172-1185.
427. Creutzburg, S.C., Wu, W.Y., Mohanraju, P., Swartjes, T., Alkan, F., Gorodkin, J., Staals, R.H. and van der Oost, J. (2020) Good guide, bad guide: spacer sequence-dependent cleavage efficiency of Cas12a. *Nucleic acids research*, **48**, 3228-3243.
428. Kim, H.K., Song, M., Lee, J., Menon, A.V., Jung, S., Kang, Y.-M., Choi, J.W., Woo, E., Koh, H.C. and Nam, J.-W. (2017) In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nature methods*, **14**, 153-159.
429. Zhu, H. and Liang, C. (2019) CRISPR-DT: designing gRNAs for the CRISPR-Cpf1 system with improved target efficiency and specificity. *Bioinformatics*, **35**, 2783-2789.
430. Liao, C., Ttofali, F., Slotkowski, R.A., Denny, S.R., Cecil, T.D., Leenay, R.T., Keung, A.J. and Beisel, C.L. (2019) Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nature communications*, **10**, 1-14.
431. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: analysis and design of nucleic acid systems. *Journal of computational chemistry*, **32**, 170-173.
432. Dillingham, M.S. and Kowalczykowski, S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and Molecular Biology Reviews*, **72**, 642-671.
433. Cui, L. and Bikard, D. (2016) Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic acids research*, **44**, 4243-4251.
434. Wannier, T.M., Ciaccia, P.N., Ellington, A.D., Filsinger, G.T., Isaacs, F.J., Javanmardi, K., Jones, M.A., Kunjapur, A.M., Nyerges, A. and Pal, C. (2021) Recombineering and MAGE. *Nature Reviews Methods Primers*, **1**, 1-24.
435. Breaker, R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology*, **4**, a003566.
436. Park, S.V., Yang, J.-S., Jo, H., Kang, B., Oh, S.S. and Jung, G.Y. (2019) Catalytic RNA, ribozyme, and its applications in synthetic biology. *Biotechnology advances*, **37**, 107452.
437. Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, **152**, 17-24.
438. Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Reviews Genetics*, **8**, 776-790.
439. Weinberg, C.E., Weinberg, Z. and Hammann, C. (2019) Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic acids research*, **47**, 9480-9494.

440. Chen, S., Bagdasarian, M., Kaufman, M. and Walker, E. (2007) Characterization of strong promoters from an environmental *Flavobacterium hibernum* strain by using a green fluorescent protein-based reporter system. *Appl. Environ. Microbiol.*, **73**, 1089-1100.
441. Gómez, E., Álvarez, B., Duchaud, E. and Guijarro, J.A. (2015) Development of a markerless deletion system for the fish-pathogenic bacterium *Flavobacterium psychrophilum*. *PLoS One*, **10**, e0117969.
442. Accetto, T. and Avguštin, G. (2011) Inability of *Prevotella bryantii* to form a functional Shine-Dalgarno interaction reflects unique evolution of ribosome binding sites in Bacteroidetes. *PloS one*, **6**.
443. Hausner, G., Hafez, M. and Edgell, D.R. (2014) Bacterial group I introns: mobile RNA catalysts. *Mobile DNA*, **5**, 1-12.
444. Edgell, D.R., Belfort, M. and Shub, D.A. (2000) Barriers to intron promiscuity in bacteria. *Journal of Bacteriology*, **182**, 5281-5289.
445. Nielsen, H. and Johansen, S.D. (2009) Group I introns: moving in new directions. *RNA biology*, **6**, 375-383.
446. Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N. and Breaker, R.R. (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *science*, **329**, 845-848.
447. Chen, A.G., Sudarsan, N. and Breaker, R.R. (2011) Mechanism for gene control by a natural allosteric group I ribozyme. *Rna*, **17**, 1967-1972.
448. Thompson, K.M., Syrett, H.A., Knudsen, S.M. and Ellington, A.D. (2002) Group I aptazymes as genetic regulatory switches. *BMC biotechnology*, **2**, 21.
449. Carrión, V.J., Perez-Jaramillo, J., Cordovez, V., Tracanna, V., De Hollander, M., Ruiz-Buck, D., Mendes, L.W., Van Ijcken, W.F., Gomez-Exposito, R. and Elsayed, S.S. (2019) Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science*, **366**, 606-612.
450. Chu, F.K., Maley, G.F., Maley, F. and Belfort, M. (1984) Intervening sequence in the thymidylate synthase gene of bacteriophage T4. *Proceedings of the National Academy of Sciences*, **81**, 3049-3053.
451. Chu, F.K., Maley, G.F., West, D.K., Belfort, M. and Maley, F. (1986) Characterization of the intron in the phage T4 thymidylate synthase gene and evidence for its self-excision from the primary transcript. *Cell*, **45**, 157-166.
452. Cech, T.R., Damberger, S.H. and Gutell, R.R. (1994) Representation of the secondary and tertiary structure of group I introns. *Nature structural biology*, **1**, 273-280.
453. Pichler, A. and Schroeder, R. (2002) Folding problems of the 5' splice site containing the P1 stem of the group I thymidylate synthase intron: substrate binding inhibition in vitro and mis-splicing in vivo. *Journal of Biological Chemistry*, **277**, 17987-17993.
454. Sandegren, L. and Sjöberg, B.-M. (2007) Self-splicing of the bacteriophage T4 group I introns requires efficient translation of the pre-mRNA in vivo and correlates with the growth state of the infected bacterium. *Journal of bacteriology*, **189**, 980-990.
455. Hamidjaja, R., Capoulade, J., Catón, L. and Ingham, C.J. (2020) The cell organization underlying structural colour is involved in *Flavobacterium* IR1 predation. *The ISME Journal*, **14**, 2890-2900.
456. Johansen, V.E., Catón, L., Hamidjaja, R., Oosterink, E., Wilts, B.D., Rasmussen, T.S., Sherlock, M.M., Ingham, C.J. and Vignolini, S. (2018) Genetic manipulation of structural



## References

- color in bacterial colonies. *Proceedings of the National Academy of Sciences*, **115**, 2652-2657.
457. Schertel, L., van de Kerkhof, G.T., Jacucci, G., Catón, L., Ogawa, Y., Wilts, B.D., Ingham, C.J., Vignolini, S. and Johansen, V.E. (2020) Complex photonic response reveals three-dimensional self-organization of structural coloured bacterial colonies. *Journal of the Royal Society Interface*, **17**, 20200196.
458. Pyne, M.E., Moo-Young, M., Chung, D.A. and Chou, C.P. (2015) Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Applied and environmental microbiology*, **81**, 5103-5114.
459. Bassalo, M.C., Garst, A.D., Halweg-Edwards, A.L., Grau, W.C., Domaille, D.W., Mutalik, V.K., Arkin, A.P. and Gill, R.T. (2016) Rapid and efficient one-step metabolic pathway integration in *E. coli*. *ACS synthetic biology*, **5**, 561-568.
460. Ao, X., Yao, Y., Li, T., Yang, T.-T., Dong, X., Zheng, Z.-T., Chen, G.-Q., Wu, Q. and Guo, Y. (2018) A multiplex genome editing method for *Escherichia coli* based on CRISPR-Cas12a. *Frontiers in microbiology*, **9**, 2307.
461. Aparicio, T., de Lorenzo, V. and Martínez-García, E. (2018) CRISPR/Cas9-based counterselection boosts recombineering efficiency in *Pseudomonas putida*. *Biotechnology journal*, **13**, 1700161.
462. Wu, Z., Chen, Z., Gao, X., Li, J. and Shang, G. (2019) Combination of ssDNA recombineering and CRISPR-Cas9 for *Pseudomonas putida* KT2440 genome editing. *Applied microbiology and biotechnology*, **103**, 2783-2795.
463. Wirth, N.T., Kozaeva, E. and Nikel, P.I. (2020) Accelerated genome engineering of *Pseudomonas putida* by I-SceI—mediated recombination and CRISPR-Cas9 counterselection. *Microbial biotechnology*, **13**, 233-249.
464. Mougiakos, I., Mohanraju, P., Bosma, E.F., Vrouwe, V., Bou, M.F., Naduthodi, M.I., Gussak, A., Brinkman, R.B., Van Kranenburg, R. and Van Der Oost, J. (2017) Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature communications*, **8**, 1-11.
465. Ahn, S.J. and Park, I.K. (2003) The coenzyme thiamine pyrophosphate inhibits the self-splicing of the group I intron. *The international journal of biochemistry & cell biology*, **35**, 157-167.
466. Jung, C., Shin, S. and Park, I.K. (2005) Pyridoxal phosphate inhibits the group I intron splicing. *Molecular and cellular biochemistry*, **280**, 17-23.
467. Jung, W.S., Shin, S. and Park, I.K. (2008) Novobiocin inhibits the self-splicing of the primary transcripts of T4 phage thymidylate synthase gene. *Molecular and cellular biochemistry*, **314**, 143-149.
468. Kim, J.H. and Park, I.K. (2003) Inhibition of the group I ribozyme splicing by NADP<sup>+</sup>. *Molecular and cellular biochemistry*, **252**, 285-293.
469. LIU, Y., TIDWELL, R.R. and Leibowitz, M.J. (1994) Inhibition of in vitro splicing of a group I intron of *Pneumocystis carinii*. *Journal of Eukaryotic Microbiology*, **41**, 31-38.
470. Park, I.K. (2000) Effects of deamido-NAD<sup>+</sup> on self-splicing of primary transcripts of phage T4 thymidylate synthase gene. *Korean Journal of Biological Sciences*, **4**, 141-144.

471. Park, I.K., Kim, J.Y., Lim, E.H. and Shin, S. (2000) Spectinomycin inhibits the self-splicing of the group I intron RNA. *Biochemical and biophysical research communications*, **269**, 574-579.
472. Park, I.K. and Kim, J.Y. (2001) NAD<sup>+</sup> Inhibits the self-splicing of the group I intron. *Biochemical and biophysical research communications*, **281**, 206-211.
473. von Ahsen, U., Davies, J. and Schroeder, R. (1991) Antibiotic inhibition of group I ribozyme function. *Nature*, **353**, 368-370.
474. von Ahsen, U., Davies, J. and Schroeder, R. (1992) Non-competitive inhibition of group I intron RNA self-splicing by aminoglycoside antibiotics. *Journal of molecular biology*, **226**, 935-941.
475. Waldsich, C., Semrad, K. and Schroeder, R. (1998) Neomycin B inhibits splicing of the td intron indirectly by interfering with translation and enhances missplicing in vivo. *Rna*, **4**, 1653-1663.
476. Groher, F. and Suess, B. (2014) Synthetic riboswitches—a tool comes of age. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, **1839**, 964-973.
477. Paulino, B.N., Sales, A., Felipe, L., Pastore, G.M., Molina, G. and Bicas, J.L. (2021) Recent advances in the microbial and enzymatic production of aroma compounds. *Current Opinion in Food Science*, **37**, 98-106.
478. SÁ, A.G.A., de Meneses, A.C., de Araújo, P.H.H. and de Oliveira, D. (2017) A review on enzymatic synthesis of aromatic esters used as flavor ingredients for food, cosmetics and pharmaceuticals industries. *Trends in Food Science & Technology*, **69**, 95-105.
479. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N. and Schaeffer, R.D. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, **373**, 871-876.
480. Gronnier, J., Crowet, J.-M., Habenstein, B., Nasir, M.N., Bayle, V., Hosy, E., Platre, M.P., Gouguet, P., Raffaele, S. and Martinez, D. (2017) Structural basis for plant plasma membrane protein dynamics and organization into functional nanodomains. *Elife*, **6**, e26404.
481. Perraki, A., Cacas, J.-L., Crowet, J.-M., Lins, L., Castroviejo, M., German-Retana, S., Mongrand, S. and Raffaele, S. (2012) Plasma membrane localization of *Solanum tuberosum* remorin from group 1, homolog 3 is mediated by conformational changes in a novel C-terminal anchor and required for the restriction of potato virus X movement. *Plant physiology*, **160**, 624-637.
482. Cauet, G., Degryse, E., Ledoux, C., Spagnoli, R. and Achstetter, T. (1999) Pregnenolone esterification in *Saccharomyces cerevisiae*: A potential detoxification mechanism. *European journal of biochemistry*, **261**, 317-324.
483. Howard, D. and Anderson, R. (1976) Cell-free synthesis of ethyl acetate by extracts from *Saccharomyces cerevisiae*. *Journal of the Institute of Brewing*, **82**, 70-71.
484. MALCORPS, P. and DUFOUR, J.P. (1992) Short-chain and medium-chain aliphatic-ester synthesis in *Saccharomyces cerevisiae*. *European Journal of Biochemistry*, **210**, 1015-1022.
485. Wiedemann, N. and Pfanner, N. (2017) Mitochondrial machineries for protein import and assembly. *Annual review of biochemistry*, **86**, 685-714.
486. Casas-Godoy, L., Duquesne, S., Bordes, F., Sandoval, G. and Marty, A. (2012) Lipases: an overview. *Lipases and phospholipases*, 3-30.

## References

487. Pleiss, J., Fischer, M., Peiker, M., Thiele, C. and Schmid, R.D. (2000) Lipase engineering database: understanding and exploiting sequence–structure–function relationships. *Journal of Molecular Catalysis B: Enzymatic*, **10**, 491-508.
488. Fülöp, V., Böcskei, Z. and Polgár, L. (1998) Prolyl oligopeptidase: an unusual  $\beta$ -propeller domain regulates proteolysis. *Cell*, **94**, 161-170.
489. Dunn, G., Montgomery, M., Mohammed, F., Coker, A., Cooper, J., Robertson, T., Garcia, J.L., Bugg, T. and Wood, S. (2005) The structure of the C–C bond hydrolase MhpC provides insights into its catalytic mechanism. *Journal of molecular biology*, **346**, 253-265.
490. Schleberger, C., Sachelaru, P., Brandsch, R. and Schulz, G.E. (2007) Structure and Action of a CC Bond Cleaving  $\alpha/\beta$ -Hydrolase Involved in Nicotine Degradation. *Journal of molecular biology*, **367**, 409-418.
491. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F. and Marraffini, L.A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research*, **41**, 7429-7437.
492. Yang, L., Briggs, A.W., Chew, W.L., Mali, P., Guell, M., Aach, J., Goodman, D.B., Cox, D., Kan, Y. and Lesha, E. (2016) Engineering and optimising deaminase fusions for genome editing. *Nature communications*, **7**, 1-12.
493. Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J. and Huang, X. (2018) Base editing with a Cpf1–cytidine deaminase fusion. *Nature biotechnology*, **36**, 324-327.
494. Kleinstiver, B.P., Sousa, A.A., Walton, R.T., Tak, Y.E., Hsu, J.Y., Clement, K., Welch, M.M., Horng, J.E., Malagon-Lopez, J. and Scarfò, I. (2019) Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nature biotechnology*, **37**, 276-282.
495. Patinios, C., Creutzburg, S.C., Arifah, A.Q., Adiego-Pérez, B., Gyimah, E.A., Ingham, C.J., Kengen, S.W., van der Oost, J. and Staals, R.H. (2021) Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas. *Nucleic Acids Research*.
496. Chu, F.K., Maley, G., Belfort, M. and Maley, F. (1985) In vitro expression of the intron-containing gene for T4 phage thymidylate synthase. *Journal of Biological Chemistry*, **260**, 10680-10688.
497. Gibb, E.A. and Edgell, D.R. (2010) Better late than early: delayed translation of intron-encoded endonuclease I-TevI is required for efficient splicing of its host group I intron. *Molecular microbiology*, **78**, 35-46.
498. Gott, J.M., Shub, D.A. and Belfort, M. (1986) Multiple self-splicing introns in bacteriophage T4: evidence from autocatalytic GTP labeling of RNA in vitro. *Cell*, **47**, 81-87.
499. Brogna, S., McLeod, T. and Petric, M. (2016) The meaning of NMD: translate or perish. *Trends in Genetics*, **32**, 395-407.
500. Culbertson, M.R. and Neeno-Eckwall, E. (2005) Transcript selection and the recruitment of mRNA decay factors for NMD in *Saccharomyces cerevisiae*. *Rna*, **11**, 1333-1339.
501. Ono, B.-i., Yoshida, R., Kamiya, K. and Sugimoto, T. (2005) Suppression of termination mutations caused by defects of the NMD machinery in *Saccharomyces cerevisiae*. *Genes & genetic systems*, **80**, 311-316.

502. Chabelskaya, S., Gryzina, V., Moskalenko, S., Le Goff, C. and Zhouravleva, G. (2007) Inactivation of NMD increases viability of sup45 nonsense mutants in *Saccharomyces cerevisiae*. *BMC Molecular Biology*, **8**, 1-12.
503. Wen, J. and Brogna, S. (2010) Splicing-dependent NMD does not require the EJC in *Schizosaccharomyces pombe*. *The EMBO journal*, **29**, 1537-1551.
504. Wen, J., He, M., Petric, M., Marzi, L., Wang, J., Piechocki, K., McLeod, T., Singh, A.K., Dwivedi, V. and Brogna, S. (2020) An intron proximal to a PTC enhances NMD in *Saccharomyces cerevisiae*. *Biorxiv*, 149245.
505. Gietz, R.D. and Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols*, **2**, 31-34.
506. Kertsburg, A. and Soukup, G.A. (2002) A versatile communication module for controlling RNA folding and catalysis. *Nucleic acids research*, **30**, 4599-4606.
507. Saldanha, R., Mohr, G., Belfort, M. and Lambowitz, A.M. (1993) Group I and group II introns. *The FASEB Journal*, **7**, 15-24.
508. Ferat, J.-L. and Michel, F. (1993) Group II self-splicing introns in bacteria. *Nature*, **364**, 358-361.
509. Shearman, C., Godon, J.J. and Gasson, M. (1996) Splicing of a group II intron in a functional transfer gene of *Lactococcus lactis*. *Molecular microbiology*, **21**, 45-53.
510. Jarrell, K., Peebles, C., Dietrich, R., Romiti, S. and Perlman, P. (1988) Group II intron self-splicing. Alternative reaction conditions yield novel products. *Journal of Biological Chemistry*, **263**, 3432-3439.
511. Daniels, D.L., Michels Jr, W.J. and Pyle, A.M. (1996) Two Competing Pathways for Self-splicing by Group II Introns: A Quantitative Analysis of *In Vitro* Reaction Rates and Products. *Journal of molecular biology*, **256**, 31-49.
512. Hiller, R., Hetzer, M., Schweyen, R.J. and Mueller, M.W. (2000) Transposition and exon shuffling by group II intron RNA molecules in pieces. *Journal of molecular biology*, **297**, 301-308.
513. Lambowitz, A.M. and Zimmerly, S. (2004) Mobile group II introns. *Annu. Rev. Genet.*, **38**, 1-35.
514. Lambowitz, A.M. and Zimmerly, S. (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harbor perspectives in biology*, **3**, a003616.
515. Asija, K., Sutter, M. and Kerfeld, C.A. (2021) A survey of bacterial microcompartment distribution in the human microbiome. *Frontiers in Microbiology*, **12**, 1090.
516. Kerfeld, C.A. and Sutter, M. (2020) Engineered bacterial microcompartments: apps for programming metabolism. *Current opinion in biotechnology*, **65**, 225-232.
517. Planamente, S. and Frank, S. (2019) Bio-engineering of bacterial microcompartments: a mini review. *Biochemical Society Transactions*, **47**, 765-777.
518. Kerfeld, C.A. and Erbilgin, O. (2015) Bacterial microcompartments and the modular construction of microbial metabolism. *Trends in microbiology*, **23**, 22-34.
519. Crowley, C.S., Cascio, D., Sawaya, M.R., Kopstein, J.S., Bobik, T.A. and Yeates, T.O. (2010) Structural insight into the mechanisms of transport across the *Salmonella enterica* Pdu microcompartment shell. *Journal of Biological Chemistry*, **285**, 37838-37846.

## References

- 520. Mahinthichaichan, P., Morris, D.M., Wang, Y., Jensen, G.J. and Tajkhorshid, E. (2018) Selective permeability of carboxysome shell pores to anionic molecules. *The Journal of Physical Chemistry B*, **122**, 9110-9118.
- 521. Chowdhury, C., Chun, S., Pang, A., Sawaya, M.R., Sinha, S., Yeates, T.O. and Bobik, T.A. (2015) Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proceedings of the National Academy of Sciences*, **112**, 2990-2995.
- 522. Chowdhury, C. and Bobik, T.A. (2019) Engineering the PduT shell protein to modify the permeability of the 1, 2-propanediol microcompartment of Salmonella. *Microbiology*, **165**, 1355.
- 523. Hagen, A.R., Plegaria, J.S., Sloan, N., Ferlez, B., Aussignargues, C., Burton, R. and Kerfeld, C.A. (2018) In vitro assembly of diverse bacterial microcompartment shell architectures. *Nano letters*, **18**, 7030-7037.
- 524. Tan, Y.Q., Ali, S., Xue, B., Teo, W.Z., Ling, L.H., Go, M.K., Lv, H., Robinson, R.C., Narita, A. and Yew, W.S. (2021) Structure of a Minimal  $\alpha$ -Carboxysome-Derived Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules*, **22**, 4095-4109.



## About the author



Constantinos (Costas) Patinios was born on the 24<sup>th</sup> of February, 1992, in Nicosia, Cyprus. After finishing his high school in 2009 and completing his compulsory 1 year military service in 2010, he moved to UK at the University of Bath to pursue a BSc degree in Biology. During his BSc, he had a 1 year professional placement at Campden BRI where he performed heat resistance and decontamination studies in the microbiology department of the company. For his BSc thesis, he focused on the tickborne Lyme disease and the associated *Borrelia burgdorferi* pathogen, where he examined the prevalence of ticks carrying the disease in the fields surrounding the city of Bath, UK. After obtaining his BSc degree in 2014, he moved to the Netherlands to pursue an MSc in Plant Biotechnology (specializing in molecular plant breeding and pathology) at Wageningen University & Research. For his MSc major thesis he used host-induced gene silencing to confer resistance of tomato plants to the notorious fungal plant pathogen, *Verticillium dahliae*. Fascinated by research and entrepreneurship, Costas dedicated almost another year (2016-2017) on his MSc minor thesis, to create yeast strains that produce the valuable anthocyanin color/antioxidant compounds. Based on the outcome of his MSc minor thesis, he founded a startup company called N-CHROMA, which (to date) still operates at a university/startup scale.

In October 2017, Costas started his PhD research at the Laboratory of Microbiology and the Laboratory of Bioprocess Engineering at Wageningen University & Research, under the supervision of Prof. Dr John van der Oost, Prof. Dr Ruud A. Weusthuis and Dr Servé W. M. Kengen. Costas' work mainly focused on engineering microbial ester production in *Clostridium beijerinckii* with focus on the EatI family of enzymes. Moreover, he focused on developing novel and universal genome engineering tools for prokaryotes. Most of this work has been described in this thesis.

Currently, Costas is employed as a Postdoctoral researcher at the Laboratory of Microbiology at Wageningen University & Research. His work focuses on deciphering protein-protein interactions in (hyper)thermophilic prokaryotes, on engineering and re-purposing bacterial microcompartments and on developing novel genome engineering tools in prokaryotic and eukaryotic microorganisms.

## List of publications

Levisson, M., **Patinios, C.**, Hein, S., de Groot, P.A., Daran, J.M., Hall, R.D., Martens, S. and Beekwilder, J., 2018. Engineering de novo anthocyanin production in *Saccharomyces cerevisiae*. *Microbial cell factories*, 17(1), pp.1-16.

Kruis, A.J., Bohnenkamp, A.C., **Patinios, C.**, van Nuland, Y.M., Levisson, M., Mars, A.E., van den Berg, C., Kengen, S.W. and Weusthuis, R.A., 2019. Microbial production of short and medium chain esters: enzymes, pathways, and applications. *Biotechnology advances*, 37(7), p.107407.

**Patinios, C.**, Lanza, L., Corino, I., Franssen, M.C., Van der Oost, J., Weusthuis, R.A. and Kengen, S.W., 2020. Eat1-Like Alcohol Acyl Transferases From Yeasts Have High Alcoholysis and Thiolysis Activity. *Frontiers in microbiology*, 11.

**Patinios, C.**, Creutzburg, S.C., Arifah, A.Q., Adiego-Pérez, B., Gyimah, E.A., Ingham, C.J., Kengen, S.W., van der Oost, J. and Staals, R.H., 2021. Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas. *Nucleic acids research*, 49(19), pp.11392-11404.

**Patinios, C.**, Yildiz, C., López-Contreras, A.M., Van der Oost, J., Weusthuis, R.A., Kengen, S.W.. Enzymes in a box: Repurposing clostridial microcompartments for the production of esters. *Manuscript in preparation*.

**Patinios, C.**, de Vries, S.T., Diallo, M., Lanza L., Verbrugge, P.L.J.V.Q., López-Contreras, A.M., Van der Oost, J., Weusthuis, R.A., Kengen, S.W.. Multiplex genome engineering in *Clostridium beijerinckii* using CRISPR-Cas12a. *Manuscript in preparation*.

## Patent applications

Patinios, C., Creutzburg, S.C., Van der Oost, J., Staals, R.H.. Universal riboswitch for inducible gene expression. PCT/EP2021/077682 (filed on October 8, 2020).



# List of completed training activities

## Discipline specific activities

### *Meetings and conferences*

- Brazilian Bioenergy Science and Technology Conference (2017)
- Netherlands Biotechnology Conference (2018)
- Clostridium XV (2018)
- Netherlands Biotechnology Conference (2019) \*
- Synthetic Biology: Engineering, Evolution & Design (2019) \*
- Netherlands Biotechnology Conference (2021) \*
- CRISPR (2021) \*
- 4<sup>th</sup> International Conference on Microbiome Engineering (2021) \*\*
- Chemistry as Innovating Science (2021) \*\*

\* poster presentation; \*\* oral presentation

### *Courses*

- Metabolic engineering and Systems Biology, Chalmers University of Technology, Gothenburg, Sweden

## General courses

- VLAG PhD week, the Netherlands (2018)
- Bioprocess design, the Netherlands (2018)
- Scientific writing skills, the Netherlands (2020)
- Introduction to R, the Netherlands (2020)
- Career perspectives, the Netherlands (2020)

## Optionals

- Preparation of research proposal (2017)
- Applied molecular microbiology course, Wageningen, the Netherlands (2017)
- PhD study trip and organization, Boston-New York, US (2019)
- Monthly microbiology department PhD meetings, Wageningen, the Netherlands
- Bacterial genetics group weekly meetings, Wageningen, the Netherlands

# Acknowledgements

What a journey it has been... 4 years, full of ups and many downs. 4 years of endless hours, early mornings, late nights, countless caffeine boosts and continuous psychological rollercoasters. If I could, I would do everything all over again, cause all of you, made this a heck of a ride! The stress was high but, the reward was high as well. You, played an important role in making this journey unforgettable. I cannot express in words my gratitude to you but you need to know that you have made me a better person and a better scientist. For this, I forever thank you!

**Servé.** I cannot not start with you. You are a man to admire really! Smart, elegant and humble. Your door was always open for me. You gave me the freedom to do whatever my heart desired, something that allowed me to enjoy my PhD but also to develop my skills as an independent scientist. You were always open to new ideas and projects, always giving me the space to create and experiment, and always making sure to be nice and supportive to me. I feel that our personalities match perfectly, although I will never forget how you tricked me with the anaerobes! We now have another project together which feels like you tricked me again with a hyperthermophilic anaerobe. It seems I always fall for you (just kidding). I am forever grateful for everything you taught me, for your patience and the freedom you gave to me. Thank you! (Oh, you still owe us a BBQ at your place).

**John.** You are one of a kind. An unbelievably good scientist with a great sense of humor (I know you have a good sense of humor since you always laugh at my jokes!). You are the mentor every young scientist needs in their life. You are always open to suggestions and crazy project ideas, you help the entrepreneurs in our group and you put our comfort above yours. Not only you build great scientists in your group but you also secure a bright future for all of us. You have built a great group, with great scientists and a great environment. This is rare in our field and I am thankful for everything that you offered to me and our group. Thank you!

**Ruud.** I will never forget when you challenged my skills, knowledge and independence as a scientist. I have to say, that was a turning point during my PhD, where it turned me into a better scientist. You are always very friendly which makes me feel very comfortable around you. Especially at parties, we can talk virtually about everything. You are a great scientist with an inspiring passion in what you do and what you teach. I will take the enthusiasm you have in your eyes and apply it in my future endeavors. We did not get the chance to collaborate closely but you were a big influence in fueling my curiosity in science. Thank you!

**Raymond.** What a unique person you are... Your knowledge on basically everything is impressive; let it be biology, the stock market and cryptocurrencies, and coding for fun. It is always fun to talk to you as I always learn something new. You take very good care of your PhD students which I have also experienced during our SIBR-Cas collaboration which was very smooth and enjoyable. SIBR-Cas is still at its beginning and I am sure we will have many more collaborations coming up. I always enjoy working with you because novelty and high risk projects are always on the menu. I admire especially your attitude and I will apply this to the future me. Thank you!

**Ana, Anna, Alex, Yuri, Mamou, Astrid, Mark, Sjoerd, Belen, Adini, Lucrezia, Inge, Maurice, Colin, Caglar, Stijn and Pepijn,** thank you for the very smooth and good collaboration. This thesis is also the result of your efforts and I am forever grateful for your contribution and support.

**Gianni mou.** If altruism exists, you are the closest I have ever seen to altruism. You are an amazing human being, an incredible friend, definitely a great husband to Eleftheria and absolutely the best dad for your son Iasonas. You are the most important pillar during my Wageningen life and my PhD. You always supported me and believed in me; let it be science, basketball, crossfit or life decisions. We have spent countless hours and days together and we share pretty much each other's life. You are a brother to me and it has been an honor knowing you and even a greater honor to be the godfather of your boy. I admire your patience, your love for science and your persistence in everything you do. Life has not been easy for you but you have taken every opportunity life has given you and turned it into a beautiful outcome. I truly and honestly admire and love you. Thank you for everything file!

**Prokopi mou.** I know we will never agree for the *in vitro* / *in vivo* terminology but we can both live with that. I always look up to you as the more mature and more knowledgeable friend. You are the person that will bring me down to earth and make me realize that many things are not as ideal as I would imagine them to be. I admire your critical thinking on different scientific subjects but also on political subjects. You are patient with many things but also very explosive with many other things. This is a characteristic that we share and this is why we are very good friends. We share an equal passion for science and we both dream how great it would be if we return to Cyprus and have a good life, applying our knowledge in Cyprus. Maybe, one day we will achieve this and we will be having lots of fun saying stupid jokes all the time. Thank you for your support during all these years file!

**Adini mou.** Agapi mou. Sweetie mou. I told you many times that you have helped me in ways that is not easy to comprehend. You have basically filled-in a huge gap that I had in me. You are the person I want to share every experience in my life. You have a great personality and you are a great

scientist too. You have contributed to my life but also to my thesis as well. You fill me up with happiness, ambitions and dreams. You are my boost when I am down, my relief when I am tired and the reason why I am happy. You are the best thing that has happened in my life the last few years. I love you and thank you.

**Cas, Pepijn, Pierre, Stijn, Lanza, Adini, Inge and Caglar**, thank you very much for your help during my PhD. You helped me massively to complete my projects and essentially my PhD. You are great people and very good scientists. I enjoyed working with you and it has been an honor being your supervisor. Some of you proceeded for a career in academia and some of you in industry. I am confident you will all do great in your life and careers. Thank you for all your help.

**Dear BacGeners, Wen, Sjoerd, Prat, Joy, Jeroen, Mougiakos, Mihris, James, Eric, Belen, Lorenzo, Thomas, Isabelle, Ricardo, Vittorio, Olufemi, Despoina, Eugene, Nico, Max, Joep, Maartje, Carina, Thijs, Miguel, Stijn, Jurre, Catarina, Suzan and Mamou**. You made my stay in our group very pleasant. You are incredible people with amazing personalities and with great scientific skills. You have contributed in making me a better scientist and for that I thank you a lot. It has been an honor working with you.

**Wen**. I can write a whole page for you really... A vibrant personality, a great scientist and very good athlete too. You are welcoming everyone in the group and you make us feel comfortable from the first moment we step in MIB. You are the reason why I started crossfit and for 2 years we have been there, every single day (except weekends for me), at 7.00 in the morning, giving everything that we got. It is really admirable how you can do all the things that you do. I honestly feel honor to be your friend, to have been your paranymp and to be your colleague. Your love and passion for science really inspires me and I know I can always come to you for good advice and novel ideas. Thank you for everything that you have done for me!

**Thijs** I am very impressed with your multiple skills. You really inspire me in the approaches you use and with how creative you are. I believe you are the perfect business partner and I hope I can be the perfect business partner for you. Hopefully we can create the startup we both dream about and to start enjoying science and life in our own creation. **Sjoerd** I am always amazed by your knowledge and the way you are thinking about- and perform science. I enjoyed working with you on the SIBR technology and of course, without your work, SIBR-Cas would never be there! Thank you a lot Sjoerd. **Prarthana** thank you for being a great lab mate. Your work inspired me to achieve my best. Also, thank you for teaching Adini so well. I received a student ready to do amazing work and this is mainly because of you. **Joy**, thank you for all the talks in our office and for being always fair and humble. **Jeroen**, my nice office mate and partner in the Munich crime. You know, I will

never forget the Munich trip we had together. That was a blast man. Thanks for all the good times and the brainstorm. **Mougiako**, I really admire you as a scientist and I am curious to see what you will discover in the near future. Thank you for trusting me with some personal things and I hope I helped you as much as I could. **Mihros**, you will always stay in my heart as the most hard working and humble person I know. You are very disciplined and you work always with a target in mind. I enjoyed our trip to Sweden very much and I learned a lot from you. Thank you my friend. **James**, my man crush. Always looking pretty and being smiley. Thank you for all the talks, the brainstorm and also the nights out. **Eric**, thank you for being a nice colleague and also thank you for letting me borrow your car when I needed it. I really appreciate it! **Belen**, thank you for trusting in the SIBR-yeast project. I believe we will get something beautiful out of it with big potential to be applied in the industry. Also, I believe your choice for staying at BacGen is the best thing our lab could have and I am very happy about it. Now your choice for being with **Lorenzo** is questionable...Just kidding! **Lorenzo** dude, you are a science geek and you know it and you should be proud about it. Really love talking science with you. Also, thank you for the duck joke. I use it often and people hate me. **Thomas**, you are a silent force. Thank you for keeping calm at any situation. I should copy you. **Isabelle** and **Stephan**, thank you for letting me borrow your car so many times. Also, thank you for being such great people and so full of energy. Looking forward working with you in the coming years. **Ricardo**, I love your passion for science and I always like brainstorming with you. Would be nice to have you in the office for the coming years. **Vittorio and Olufemi** you are new in the office so I am looking forward working with you. **Despoina**, your laugh and energy can make everyone happier. Thank you for the talks and laughs. **Eugene**, you get a separate text below. **Nico, Max, Joep, Maartje, Carina, Miguel, Catarina and Suzan**, thank you for being amazing colleagues. **Stijn and Jurre**, you live the dream that I had a few years back. I admire your efforts, persistency and ambitions. You are a true inspiration to me! **Mamou**, thank you for teaching me everything about clostridia. I really appreciate the time you have dedicated to teach me. Thank you!

**Other friends from MIB, Menia, Sudi, Chen, Nancy, Anastasia, Caifang, Burak, Patricia, Felix, Stephan, Lukas, Lot, Diana, Ivette, Irene, Ran, Emmy, Janneke, Martha, Sharon, Ruth, Jannie, Kate, Marie-Luise, Hugo, Taojun and Carrie**, thank you for being nice people and making my time in MIB enjoyable. Also, to the people who joined the PhD trip in 2019, thank you for the unforgettable moments. To the people who we danced together, thank you for the great dance moves and vibes. To the people that we got drunk together, thank you for not remembering it. Kidding!

**Menia mou**, a big thank you for being a part of my PhD journey. We had great times in Wageningen but also in Greece and the US. Thank you for the talks, the laughs and the stupid moments. Thank

you for the nice talks on the Lombardi balcony and thank you also for the chit-chatting in our offices.

**To my friends in SSB, Maria, Enrique, Lyon and Christos**, thank you for all the good times, especially in the US. You all made that trip unforgettable. Thank you!

**Friends from BPE**, we did not get the chance to collaborate a lot but the interaction I had with you was always pleasant. Thank you for the good times.

**To the best PhD trip committee ever, Giannis, Catarina, Ivette, Caifang, Ran, Lot, Thanaporn and Enrique**, we can all say that we worked perfectly well together, we had an amazing trip and we also enjoyed organizing it. Probably the best collaboration I had with a big group of people!! Thank you!

**To the technician team of MIB, Rob, Steven, Tom, Guus, Laura, Ton, Iame, Hans, Merlijn and Ineke**, thank you for all your help. **Rob**, thank you for the assistance in BacGen. **Ton**, thank you for all the help with GC and HPLC. Your contribution to the analytics labs is invaluable. **Steven**, thank you for overseeing the good functioning of the labs.

**To the best and most friendly secretaries Anja and Heidi**, I want to say a big thank you. You always helped me with my weird requests and you always find a way out of complex bureaucratic situations. Also, thank you for helping me with my company.

**To my dearest friends from the corridor above the clouds in Bennekom, Tanvir, Nacho, Hannah, Fefo, Anna, Andres and Maria**, you have made my life in Wageningen so meaningful. You are a family to me. You complete me as a person. The memories I have from you are simply the best. I love you and I wish we remain close for the rest of our lives. **Hannah**, a special thanks belongs to you. You changed me in so many ways, making me a much better person. You will always remain close to my heart. Thank you!

**My entrepreneurship fellows, Thomas, Mark and Jules**, thank you for believing in me in the N-CHROMA endeavors. **Thomas**, from day 1 you were there. You coached me, you nurtured me and you dedicated lots of your personal time on N-CHROMA, without ever asking anything back. You are a real friend, a true entrepreneur and generally a great person. I will never forget what you have taught me and what you have done for me. Also, that trip to Paris... Man that was quite something!! **Mark**, you taught me everything I know about biotech, microbiology and metabolic engineering. Your teaching was the fuel to make me what I am right now. Thank you for everything Mark! **Jules**,

thank you for having me as your student during my crazy phase with N-CHROMA. You taught me a lot and you believed in our project. It has been unforgettable working with you. Thank you!

**Charlotte, Tjamke, Xanthe, Nienke, Lisette, Ivy, Esther, Eva, Eva Koopman, Renee, Simona, Marta, Elisa, Nuria, Sonja, Manon, Lisette, Wiebke, Maro, Espejita, Ana, Federica, Victoria, Pauline and Femke**, thank you for all the intense moments we had in the basketball courts, thank you for all the hours you dedicated in trainings and games, thank you for all the wins but also the loses, thank you for the championships, thank you for trusting me as your coach. Being your coach was an honor.

**Coach Pieter, Bram, Jose, Thijs, Daniel, Nicolo, Richard, Jan, Tom, Julian, Jerry, Matthias, Steven and Miguel**, thank you for being good and fun basketball teammates. I had the best time in Sphinx H1 team. Thank you for everything.

**Wen, Rutger, Judith, Manos, George, Keon, Erik, Kutay and Claudia**. Thank you very much for all your support during my crossfit adventure. I enjoyed training with you a lot. You always motivated me and made me a better athlete. I have very good and intense moments from crossfit that will stay close to my heart.

**My lovely Lombardi housemates Philipp, Belinda and Eugene**, thank you for sustaining me all these years with all my weirdness. **Philipp**, you are a person I always talk about because you are simply amazing. You are the best housemate someone can ever have. It was so smooth and so easy living with you that I always recall those times with a smile on my face. I learned a lot from you and for that I thank you deeply! **Belinda**, you stayed with me and Philipp during the pandemic lockdown and I have to say it was very smooth. Thank you for all the delicious food you shared with me, for keeping company to Philipp and for having all the deep and intellectual conversations during dinners. **Eugene**, you are one of a kind man. You taught me how to compromise for many things that, essentially, are not really important. You taught me how to escape from predispositions that we carry from our own countries/families. We share a love for science and music and I hope you pursue your dreams in both.

**To my other friends in Wageningen, Paul, Thomas, Demetris, Stephan, Jules and Beril**, thank you for all the good times, the nice discussions and the laughs.

**Eleni Papadatou**, you are the reason why I am a biologist. Your passion for biology and for teaching made me love this subject. If it was not for you I do not know whether I would have studied biology. I remember the experiments we did back in school, the fun we had in our class and of

course your quality teaching. I cannot thank you enough for the positive impact you had on my life. I never regret choosing biology as my career. Thank you!

**My amazing friends from Cyprus, Demetri, Giorgio, Ilia, Ktisti, Kotsio, Foto, Foti, Elena, Maria, Georgia, Anastasia, Antigoni and Marigo**, your support and help over the years is what shaped me to who I am. **Demetri mou**, you are a very strong pillar in my life and you are supporting me no matter what. You are always there to take the hits and you always bring me back to reality. We share everything together and we are bonded very tightly like brothers. You are the best friend to me. I cannot thank you enough file! **Giorgio mou**, we are connected with something special. We have genuine love and appreciation for each other. Thank you for all your support. **Marigo mou**, you have been there for me since my mom left us. We share something special the two of us that can never split us apart. We move on with our lives but we still can communicate and support each other only by talking for a few minutes. Your support and genuine love has been invaluable for me. Thank you!

**Eleftheria mou**, your support to me throughout the years in Wageningen has been invaluable. Countless days spent together and countless meals consumed together. Thank you for keeping an eye on me on difficult times, thank you for providing to me everything that I ever needed and thank you for being a great wife to Giannis and a perfect mother to Iasonas. It is an honor to be the godfather of your son. Also a big thank you to your family for their love and support. Thank you!

**Iasona mou**, you have brought a new meaning to my life. I love you to the moon and back. You are the sweetest little boy and I hope to serve rightfully my duties as your godfather. I am looking forward to the moment when you say “Tata” and to all of your achievements in life. Love you aggele mou!

**Nonna mou**, thank you for all the support and love throughout the years. You have been very supportive to me and my family. Thank you!

**Yiayia Christina, yiayia Argyri and yiayia Kallou**, thank you for all the love, support and amazing food you provided to me and my siblings. You are big contributors to who I am today. Thank you!

**Thia Demetra**, thank you for all the support you provided to my family. You are one of the strongest person I have ever met and you sacrifice yourself for the best of the others. I love you a lot. Thank you!



**Therese mou**, from the moment you entered our lives, you have made our family a better family. Thank you for all the support you provided to me during the army, during my studies and later during my PhD. A huge thank you should go for taking care of our dad. I cannot imagine a better person being next to our dad! Also thank you for taking care of the house and our cat Freddy. You contribute massively to the wellbeing of our family and you have been very important for me. I love you a lot. Thank you!

**Brother mou Mario**, you are my role model. I always copy your every move and try to become better than you. Only when I become better than you is considered a success to me. You always set the bar very high so I always have to excel, under any conditions and circumstances. I always envy how successful, smart and how organized you are. If I did not have you to push me and drive me to become better every day, then I would not be where I am right now. Thank you for making me who I am today. Also, thank you for maintaining the balance in our family. I love you and thank you! Also, a big thanks goes to **Bea** for completing your life. Thank you **Bea** mou.

**Sister mou Christina**, you are the person who unravels the emotional part in me. You have been very supportive and encouraging to me and I owe a lot to you. I admire your creativity and your talent which is reflected in everything you touch. In everything that involves creativity, I always think about you. I try to copy your creativity but I cannot even touch your quality. You have been carrying a lot of our family weight for all these years and you have been doing a great job. I love how outspoken and rebellious you are. I share these characteristics with you. I love you a lot. Thank you!

**Mama mou**, I miss you terribly much. I wish you could see what we have achieved as a family. You made us stronger as individuals and you would have been very proud of us. Thank you mom!

**Papa mou**, words cannot explain the respect, admiration and love that I have for you. As a widow man with three kids, you did an amazing job getting all three of us to study and fulfil our dreams. We had everything that we ever needed and all of this is because of your hard working attitude. You have an immense amount of power to keep doing what you are doing. My working attitude is directly reflecting your working attitude which also directly reflects the results of my profession. You put me and my siblings above anything in your life. We see this and we appreciate it massively. You support us in anything that we want to do and you are always, always, always there for us. I hope we will continue to make you proud and happy in the future. I love you more than anything. Thank you for everything!



The research described in this thesis was financially supported by the VLAG graduate school.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Thesis cover design, chapters cover design and invitation design by: Christina Patiniou

Printed by: Proefschriftmaken; [www.proefschriftmaken.nl](http://www.proefschriftmaken.nl)



