Towards the production of sustainable biofuels using cyanobacteria

Daniel Fredy Espiritu Vergara

Esther Karanukaran

MSc Biological & Bioprocess Engineering

Production of bisabolene using cyanobacteria

August 2020

Department of Chemical & Biological Engineering

University of Sheffield

ABSTRACT

Fossil fuels have been used for centuries as a main source of energy, despite the pollutants that fossil fuels generated in the environment. New alternatives, as biofuels, have arisen as alternatives to diminish the environmental impact, but there are some drawbacks that need to be tackled. To date, biofuels from microorganisms, as known as third-generation biofuels, have drawn attention to researchers owing to the capability to generate biofuels by modifying the genetic material of different microorganisms.

Many microorganism consortia to produce biofuels have been studied, but cyanobacteria -a microalgae- have shown to be more sustainable to generate biofuels owing to their similarities to plant chloroplasts to perform photosynthesis. Indeed, cyanobacteria contain the MEP pathway that is a metabolic route to biosynthesise terpenoids. However, the most studied cyanobacterium strain has been *Synechocystis sp. PCC 6803*.

Synechocystis has been studied and engineered to produce terpenoids as squalene, bisabolene, and limonene. To date, bisabolene seems to be more suitable to replace some fossil fuels owing to their physical chemistry properties. However, few studies have been reported to produce bisabolene.

This work has been developed to biosynthesise in *in silico* a construct of a biological device containing both biological elements and a plasmid vector. This biological device is capable of enhancing the yield of bisabolene inside of *Synechocystis* as a host organism.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

Fossil fuels, mainly gas, oil, and coal, are used for human being activities to generate energy and transportation. However, this dependence on fossil fuels has been date back since the first industrial revolution, furthermore, the consumption, nowadays, has barely declined, and fossil fuels remain as the main source of energy. Despite of being the major source of greenhouse gas emissions.

Furthermore, deposit explorations and a decrease of fossil fuel reserves have led to an increase in prices of those sources of energy (Akhtar & Amin, 2011), and, therefore, a search for alternatives to fossil fuels has been explored. Those alternatives should be according to major concerns, such as the reduction of greenhouse gas emissions, their capability to be efficient enough as fossil fuels, and food safety.

Biofuels by using biotechnology have emerged as alternatives to fossil fuels since the first oil crisis in 1970's, but in 2008 the industrialization of biofuels, such as ethanol, from crops emerged as alternative (Marshall, 2010). Despite the development of first-generation biofuels, made from crops, another concern arose such as food security because those biofuels needed extensive farmlands to generate feedstocks to convert them into bioethanol or biodiesel.

The industrialization of crops to produce biofuels generated quantities of agricultural wastes together with municipal wastes were considered as useful biomass, or second-generation feedstocks, to produce ethanol, methane, hydrogen, and biodiesel known as second-generation biofuels. (Saladini et al., 2016). However, drawbacks with agricultural residues as current technologies to get both low production cost and process efficiency led to a search for other alternatives, similar to municipal wastes.

During the search for alternatives to fossil fuels and reduce farmland use, microorganisms emerged as a viable alternative because they do not depend either on farmland or season to grow. Picazo-Espinosa et al. (2011) referred to algae as microorganisms to produce biofuels by taking into account their capability to produce third-generation biofuels. Indeed, the feedstocks for producing third-generation biofuels do not depend on either the season or farmland, but also these feedstocks do not interfere with food safety.

Furthermore, algae can be classified, depending on their size, into macroalgae and microalgae; indeed, microalgae, such as cyanobacteria, have shown to be viable organisms to grow by using carbon dioxide (CO_2), solar light, and nutrients to replicate and secrete products. However, cyanobacteria contain the methylerythritol 4-phosphate (MEP) pathway, a metabolic route to produce terpenoids, that cannot be naturally overexpressed.

Unlike the photosynthesis pathway, the MEP pathway needs to be assembled by using synthetic biology, modifying the sequence of the gene of interest (GOI) together to bioparts to produce terpenoids (Wang et al., 2012). Bioparts such as promoters, ribosomal binding sites (RBS), transcriptional factors, and terminators in cyanobacteria were reported by Heirdorn et al. (2011) and classified by Till et al. (2020). However, Englund et al. (2015) demonstrate that those bioparts can be assembled with the GOI in an engineered plasmid (pEERM). The construction of this plasmid with those parts facilitates the integration of the pEERM plasmid inside of the *Synechocystis sp PCC 6803* (*Synechocystis*) DNA sequence by using homologous recombination.

The pEERM integrative plasmid was created to facilitate the construction of a biological device, which can be constructed with different bioparts, to amplify the expression of the GOI in *Synechocystis*. Indeed, the GOI is referred to a terpenoid gene, obtained from a plant, capable of being replicated in *Synechocystis* genome. Furthermore, the expression of the terpene gene, such as bisabolene gene, has to be optimised to be replicated in *Synechocystis* before being inserted in pEERM plasmid (Zhou et al., 2014).

High-value terpenoids are sesquiterpenoids owing to their capability to be transformed into biofuels that can be used as jet-fuels. Bisabolene is considered an important sesquiterpenoid owing to its capability of being converted into jet-fuel and its ease to blend with other aviation fuels. However, bisabolene production by using *Synechocystis* as a host organism needs to be enhanced with synthetic biology.

The understanding of each biologic part (biopart) and their role inside the construction of a synthetic biological device can lead to an improvement in the metabolic route of the host organism, *Synechocystis*. This enhancement can lead to a high production of bisabolene that can be commercially viable.

1.2. Project overview

The first successful attempt to biosynthesise plant genes to produce terpenoids was performed by Martin et al. (2003). What they found was that microorganisms such as *Escherichia coli* (*E. coli*) could work as a host organism to secrete terpenoids by genetically engineering the metabolic route of this microorganism.

Unlike an engineered metabolic route in *E. coli*, *Synechocystis* contains the MEP pathway, which can be used for terpenoid production. The advantage of *Synechocystis* against *E. coli* is that *Synechocystis* contains a well-defined metabolic pathway to express terpenoid genes as the MEP pathway. However, this pathway cannot be encoded naturally, and some genetic modifications of *Synechocystis* genome need to be performed.

Lindberg et al. (2010) rerouted the MEP pathway of *Synechocystis* to biosynthesise isoprene molecules. Their work showed that synthetic biology could lead to a modification of any gene of the MEP pathway to encode either simple or complex terpenoids. Besides, that isoprene can be converted into either a biosynthetic rubber or a biofuel. A similar approach was performed by Englund et al. (2014), they modified the MEP pathway to express the squalene gene by using the knock-out strategy, demonstrating that complex terpenoids can be produced in Synechocystis.

Unlike knock-out strategy, knock-in strategy enables one to insert any heterologous gene inside the *Synechocystis* genome. The knock-in strategy was performed by Davies et al. (2014) to produce limonene and bisabolene. The yield achieved was considered the highest obtained from a microorganism. Despite the optimisation of the bisabolene gene, the bisabolene yield was lower than they expected. Low yield was owing to the use of a weak promoter, the enzyme kinetics, or an incomplete incorporation of the constructed plasmid inside of cyanobacteria genome. However, Englund et al. (2015) designed a pEERM series plasmids to deal with the plasmid integration inside Synechocystis during the homologous recombination.

Bisabolene production in *Synechocystis* has been performed by Dienst et al. (2020) by using a copper-inducible promoter (P_{petE}) and an optimised codon sequence of *Abies grandis* gene. Their results showed that the highest titre value obtained was 12.2 mg/L; this value was lower than a previous value reported by Sebesta and Peebles (2020) that was 22.2 mg/L by using both an engineered inducible promoter (P_{trc2O}) and the optimised codon sequence of bisabolene synthase.

During the development of this research project, the codon optimisation of *Abies grandis* gene (ag1) is carried out by using the *Synechocystis* codon usage table in order to enhance

the bisabolene expression, when homologous recombination is performed. By following the Standard Assembly Method, the construction of bioparts containing restriction enzymes are constructed in BioBricks. All constructed BioBricks will be designed *in silico* by using Benchling software, which enables a simulation in a computer, *in silico* design, of different bioparts inside of a standard plasmid (pSB1c3) with standard restriction enzymes (EcoRI, XbaI, SpeI, and PstI).

Finally, the BioBrick constructed is assembled in pEERM vector to obtain a plasmid vector with elements capable of being replicated in *Synechocystis* to produce bisabolene terpenoid.

1.3. Aims and objectives

1.3.1. Aim

To biosynthesise, *in silico,* a vector plasmid, with biological elements, capable of being replicated inside of Synechocystis sp PCC 6803 in order to produce bisabolene.

1.3.2. Objectives

- To optimise the bisabolene gene expression from the plant *Abies grandis*, by changing nucleotide sequences, to be encoded inside of *Synechocystis* sp PCC 6803.
- To select strong bioparts capable of enhancing the expression of bisabolene gene.
- To construct BioBricks with restriction enzymes according to the Standard Assembly Method.
- To construct *in silico* a vector plasmid containing strong bioparts, the gene of interest (bisabolene gene), and the plasmid backbone to be integrated inside *Synechocystis* sp PCC 6803.

2. SUMMARY OF LITERATURE REVIEW

2.1. Third-generation biofuels

The depletion of fossil fuel reservoirs, the pollution generated during the combustion process, and an increase in energy demand have encouraged companies and researchers to seek alternatives to fossil fuels. Biofuels are a renewable energy source capable of replacing fossil fuels owing to biofuels are deemed to be greener than fossil fuels.

Biofuels are produced by using different feedstocks, but depending on their origin, biofuels can be as first-, second-, and third-generation. First-generation biofuels are produced from feedstocks obtained from crops such as maize, sugar cane, soybean, beetroot, and among other feedstocks that have a high content of either soluble sugars or fatty acids capable of being converted into bioethanol or biodiesel (Saladini et al., 2016). Indeed, biodiesel and bioethanol are the main products of this group of biofuels.

A biofuel is considered a second-generation when it is produced from wastes, which can be either agricultural, obtained from both residues of first-generation biofuels and feedstock that are not related to human consumption, or municipal wastes. Biofuels produced can vary from biofuels produced in first-generation to biogas (biomethane and biohydrogen).

In the case of third-generation biofuels, they can be produced by microorganisms without the need to use farmlands or competing with food security as first- and second-generation biofuels do. Microorganisms were classified, according to their capability of producing their food, heterotrophic and autotrophic microorganisms (Lackner, 2017). The latter was considered to be more efficient because they could take CO_2 and solar energy to convert them into biofuels, whereas heterotrophic microorganisms depend on organic feedstocks to produce biofuels.

Regarding sustainability, third-generation biofuels are produced by autotrophic microorganisms such as algae (macroalgae and microalgae), which are deemed more efficient and productive than first- and second-generation feedstocks as soybean or municipal wastes (Picazo-Espinosa et al., 2011; Lackner, 2017). Furthermore, algae as feedstock can produce biodiesel, bioethanol, and biogas being more productive without depending on either the season or farmlands.

2.2. Cyanobacteria for biofuel production

Cyanobacteria are prokaryotic organisms, which are found in freshwater, capable of performing the photosynthesis process. This photosynthetic process is similar to plants, but cyanobacteria use internal membranes called thylakoids. Thylakoids work similarly to chloroplasts in plants owing to chloroplast are the evolved form of thylakoids.

Unlike plants, cyanobacteria can grow properly in a harsh environment wherein pH changes, salt concentration, and oxygen starvation are present (Tiwari, 2018). Besides, a characteristic of cyanobacteria is that they have a metabolic route adequate to produce biofuel precursors such as terpenes. Indeed, there are two metabolic routes or pathways to produce terpenes such as mevalonate (MVA) and 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. Plants have both pathways to produce terpenes, one in their cytosol (MVA pathway) and the other in the plastids (MEP pathway), whereas cyanobacteria contain the MEP pathway as the main route. A comparison of both pathways was reported by Wang et al. (2018), they concluded that MEP pathway was more carbon-efficient than MVA pathway to produce an isoprenoid precursor such as isopentenyl diphosphate (IPP).

In the MEP pathway, cyanobacteria take CO_2 as carbon source, the sunlight, and other nutrients to convert them into glyceraldehyde 3-phosphate (G3P) metabolite, which is a precursor of MEP pathway G3P metabolite is regulated to be used in the glycolysis cycle and Calvin cycle, obtaining G3P and pyruvate as precursors of MEP pathway. G3P and pyruvate can be converted into DXS (D-xylulose 5-phosphate synthase) by dxs enzyme.

Metabolites formed after DXS are converted into different metabolites by enzymatic chain reactions until obtaining IPP (isopentenyl diphosphate) and DMAPP (dimethylallyl diphosphate) metabolite (Englund et al., 2015; Pattaniak & Lindberg, 2015). IPP and DMAPP metabolites are capable of being biosynthesised by enzymes in terpenes that can be converted into biodiesel. However, different cyanobacterium strains have been studied in order to improve their metabolic routes to enhance the terpenoids because cyanobacteria are not capable of encoding the MEP pathway naturally.

2.3. Biofuel production in *Synechocystis*

Synechocystis sp. PCC 6803, hereafter *Synechocystis*, is a cyanobacterium strain that can be found in freshwater such as lakes, ponds, or natural water reservoirs. In *Synechocystis*, its complete genome and more than 3000 genes have been sequenced and identified (Yu et al., 2013). Similar to other cyanobacteria, *Synechocystis* contains the MEP pathway as part of its

metabolic route, but due to similarities to plant chloroplasts have made that this strain could be considered an excellent host to manipulate its coding sequence.

In order to produce biofuels in *Synechocystis*, it must be genetically modified or engineered owing to these microalgae are unable to enhance biofuel precursor yields to be transformed in biofuels. Among biofuels that an engineered metabolic route of *Synechocystis* can produce vary from the simplest form of alcohol as ethanol to the most complex as terpenoids.

In the case of ethanol production, Dexter and Fu (2009) isolated and inserted ethanol producer genes from *Zymomonas mobilis* in *Synechocystis*. Their findings were that *Synechocystis* was capable of producing ethanol by using a heterologous gene, despite obtaining a low yield. However, their work unveiled that *Synechocystis* could be a suitable host for heterologous gene expression.

Lindberg et al. (2010) reported that *Synechocystis* could be a host microorganism for isoprene production, an unsaturated hydrocarbon molecule, that can be converted in biofuel. They found that engineering some elements of the MEP pathway of *Synechocystis*, this organism could lead to a high production of isoprene or other terpenes to be used as biofuels. Since this finding, researches have focused on terpenoids production owing to the similarities to biodiesel and jet fuels.

2.4. Synthetic biology of *Synechocystis* for producing bisabolene

Reinsvold et al. (2011) demonstrated that a sesquiterpene could be produced in a photoautotrophic organism such as *Synechocystis*. In their study, they were capable of replicating the β -caryophyllene synthase gene from the plant Artemisia to produce β -caryophyllene terpene, this terpene production lead to a development of studies to perform terpenoid production in different cyanobacteria strains in order to demonstrate which strain could be more suitable to overproduce terpenes. Terpenes as limonene, bisabolene, and squalene have been considered as replacement of some fossil fuels. However, bisabolene has shown to be more useful to be blend with some fossil fuels to enhance cloud point (the temperature to change from liquid state to solid state)

In order to convert *Synechocystis* into a cell factory to produce bisabolene, it is required to recognize every biological element which must be able to encode the MEP pathway in this host organism. Heirdorn et al. (2011) studied some elements that cyanobacteria have in common such as promoter, ribosomal binding sites (RBS), and terminators. What they found was that using synthetic biology, all those parts can be interchangeable to express a gene of interest (GOI) by using cyanobacterium strains.

Furthermore, additional studies were performed in order to determine the influence of each element during the transcription process of a determined gene (Englund et al., 2016; Liu & Pakrasi, 2018; Till et al., 2020). Promoters and RBSs varied depending on the GOI, but terminators such as the BioBrick BBa_B0015 remained similar in all constructs without taking into account the cyanobacterium strain.

To date, Dienst et al. (2020) and Sebesta and Peebles (2020) have designed with the support synthetic biology a metabolic route which enhances bisabolene production as a metabolite, but there are still concerns owing to the dependence of an inducer to enhance such production as copper salts (Dienst et al., 2020) or lactose (Sebesta & Peebles, 2020). Despite the dependence of promoter for an inducer, there are other promoters (Fig. 4.3) that need to be evaluated to determine the most versatile for producing bisabolene in different environmental conditions.

3. RESEARCH METHODS

3.1. Overview

In this section, it is intended to summarize how bibliographical information was found in ProQuest, the use of websites to find parts or elements (iGEM and Addgene), the use of an open software to simulate digestion and ligation for performing a *in silico* design (Benchling).

3.2. Searching bibliographic information in ProQuest

ProQuest is a database wherein millions of electronic journals, scientific textbooks, and scientific newspapers are available for researchers and institutions. ProQuest has been used to find the information required for the outcome of this research project.

Important keywords as promoter, synthetic biology, *Synechocystis*, and bisabolene were used to filter the information from database in order to obtain more reliable information for the purpose of this research. In Fig. 3.1, it is shown how the use of *Synechocystis* and promoter give filtered information of some papers that were used in this project. To note, all filtered information must be scrutinized by the researcher to consider the most relevant for the determined purpose.

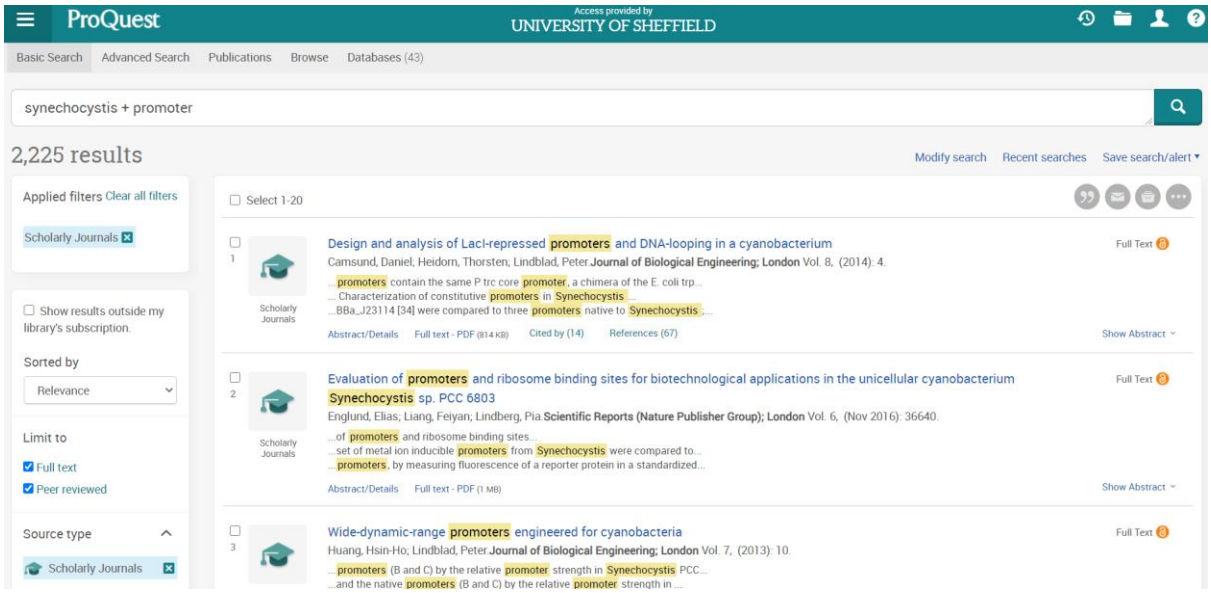


Fig. 3.1 Screenshot of the filtered information by using keyword as *Synechocystis* and promoter. In this study, papers from 1 to 3 were used as technical information.

3.3. Finding plasmid sequences deposited in Addgene

Addgene is a repository library wherein coding sequences of plasmids, which were reported in research papers, are deposited to access freely for the scientific community.

Vector database was used in this research in order to find the coding sequence of the plasmid pEERM which was deposited by Englund et al. (2015). The way the coding sequence was found is depicted in Fig. 3.2. By clicking in the plasmid vectors pEERM3 Km and pEERM4 Cm, their nucleotide sequences are downloaded to manipulate their elements.

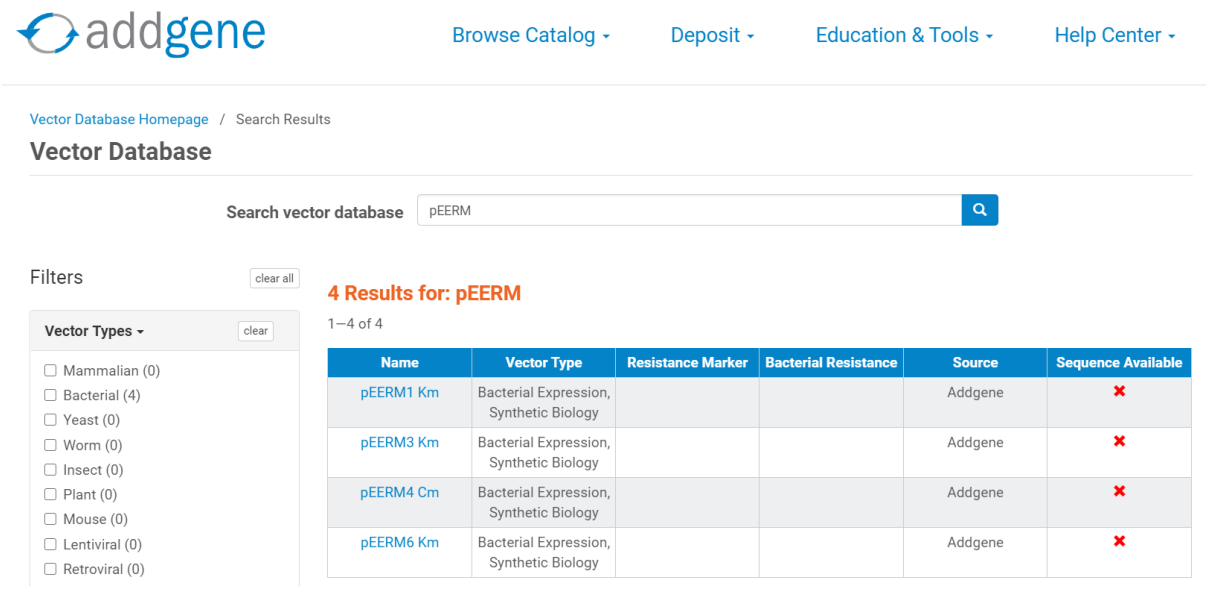


Fig. 3.2 Illustration of how plasmid vectors pEERM3 Km and pEERM4 Cm were found and downloaded from Addgene.

3.4. Searching of standard biological parts in iGEM Parts Registry

iGEM Parts Registry (http://parts.igem.org/Main_Page) is part of the iGEM which is a repository of genetic parts that can be combined to construct synthetic biology devices. Indeed, iGEM Parts Registry contains thousands of genetic parts compatible with each other because they have to follow a standard before being deposited in this repertoire.

Promoter and terminators were found by using iGEM Parts Registry, when any part is searched in this repertoire it is needed to know its BioBrick part which starts with 'BBa' or the full name of the part as Promoter P_{cpcB}. Fig. 3.3 shows how to find the standard part of promoter cpcB.

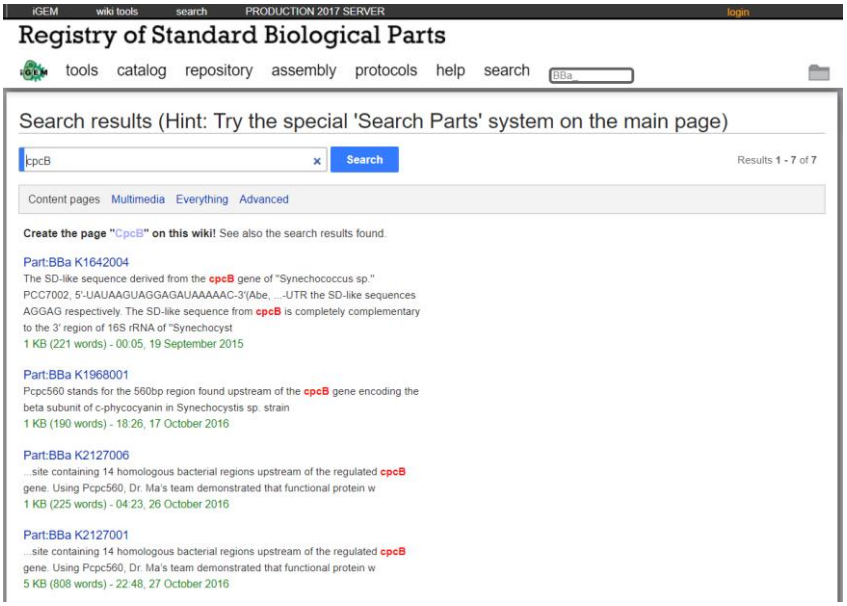


Fig. 3.3 Image of how the searching engine of iGEM Parts Registry could help to find the standard part of the promoter cpcB used in this work.

3.5. Using Benchling for *in silico* design

Benchling is an open platform based on a cloud interface that stores and enables the manipulation of DNA sequences for research purposes. In this work, Benchling was used to manipulate genes and parts to construct the plasmids vectors shown in Fig. 4.40, this software has its tutorial of how to import a DNA sequence from Addgene and iGEM. However, the aim of this section is to give a general idea of how the platform is, instead of explaining details that are found in the free tutorial.

Fig. 3.4 Benchling platform wherein it shows the nucleotide sequence and the linear map of a part (Promoter cpcB, P_{cpcB}).

4. RESULTS

4.1. Selection and optimisation of the gene of interest

Bisabolene is a chemical compound that belongs to a sesquiterpene family and is found as a resin in combination with other terpenes. Indeed, terpenes act as biopesticides to repel some insects and animals, but also those terpenes work as pheromones for other insects. Bisabolene can be extracted together with essential oils, but its purification process is complicated and the yields do not justify its commercialisation (Zheljazkov, 2012).

In nature, there are plants that contain a small bisabolene content such as oregano and lemon. However, Bohlmann et al. (1998) indicate that conifer plants (Abies grandis, Abies menziesii, and Pseudolarix amabilis) have an induced defence mechanism, which is capable of producing more bisabolene when insects attack confiers. Indeed, Abies grandis has been a well-studied plant owing to its capacity to contain many terpene genes to produce bisabolene, limonene, pinene, camphene, and among others. Furthermore, Peralta-Yahya et al. (2011) reported that bisabolene synthase protein coding sequence (ag1 gene) from Abies grandis were capable of being replicated with high values in *E. coli* in comparison with the other conifer gene plants.

The bisabolene gene from the conifer plant Abies grandis, E-α-bisabolene synthase, was sequenced and reported by Bohlmann et al. (1998) in GenBank with an accession number AF006195.1. The ag1 gene was reported in a complementary DNA (cDNA) which a gene sequence capable of being replicated in a prokaryotic organism as *Synechocystis*. During this work, ag1 gene sequence is exported to Benchling software to sequence the gene structure to highlight the restriction enzyme sites inside the ag1 gene. In the Fig. 4.1, restriction enzyme sites are highlighted to stand out the standard restriction enzyme sites (EcoRI and SpeI) that are present inside the ag1 gene sequence.

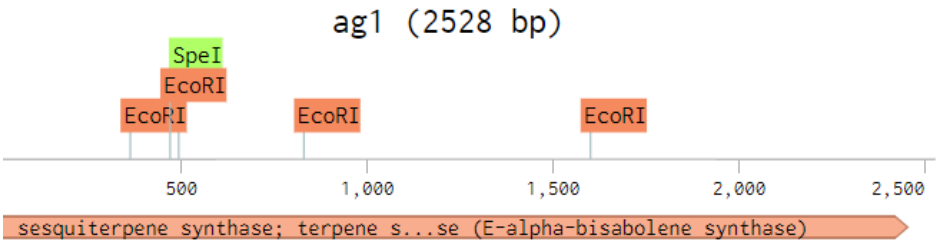


Fig. 4.1 ag1 gene sequence containing restriction enzyme sites (EcoRI and SpeI). Sequenced by the author using Benchling.

The finding of restriction enzyme sites inside ag1 gene can break this gene into pieces when the digestion process is carried out. Those DNA fragments are owing to the presence of restriction enzymes, which lead to a loss of the whole gene sequence.

The removal of restriction enzymes is performed by changing nucleotides of ag1 sequence, and furthermore, this change can optimise the ag1 gene. Indeed, the removal and the optimisation process are made together in order to remove the restriction sites of ag1 and improve gene sequence to be translated by *Synechocistis.* In this study, by using Benchling and the nucleotide sequence highlighted in red from table 4.1, codon optimisation is achieved as it is observed in Fig. 4.2, whereas the optimised nucleotide sequence is observed in Appendix A.2.

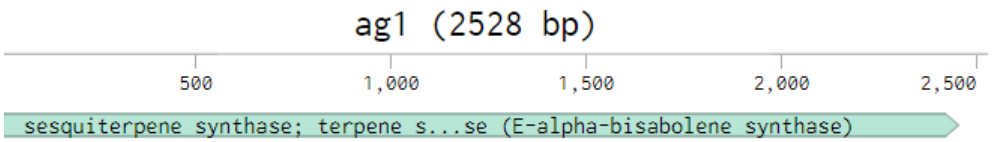


Fig. 4.2 Codon optimisation of ag1, without restriction enzymes. Optimised by the author using Benchling

Table 4.1. Codon usage table of Synechocystis. Adapted from: Kazuza DNA Research Institute (2018)

	T			C			A			G			
T	TTT	29.4	F	TCT	9.0	S	TAT	17.4	Y	TGT	6.3	C	T
	TTC	10.5		TCC	15.8		TAC	11.9		TGC	3.9		C
	TTA	26.4	L	TCA	4.3		TAA	1.4	**	TGA	0.6	**	A
	TTG	28.9		TCG	4.1		TAG	1.1	**	TGG	15.5	W	G
C	CTT	10.2	L	CCT	10.0	P	CAT	11.7	H	CGT	10.3	R	T
	CTC	13.9		CCC	24.5		CAC	7.2		CGC	12.2		C
	CTA	14.0		CCA	8.1		CAA	34.0	Q	CGA	5.4		A
	CTG	20.0		CCG	8.3		CAG	21.2		CGG	13.4		G
A	ATT	40.0	I	ACT	13.9	T	AAT	25.7	N	AGT	15.1	S	T
	ATC	17.8		ACC	26.0		AAC	15.0		AGC	10.3		C
	ATA	4.9		ACA	7.0		AAA	30.1	K	AGA	4.6	R	A
	ATG	19.3	*/M	ACG	7.8		AAG	12.8		AGG	4.8		G
G	GTT	16.8	V	GCT	20.0	A	GAT	32.5	D	GGT	19.8	G	T
	GTC	11.2		GCC	37.5		GAC	17.8		GGC	22.3		C
	GTA	10.5		GCA	10.9		GAA	44.7	E	GGA	12.9		A
	GTG	28.0		GCG	15.2		GAG	16.1		GGG	17.6		G

* Start codon

** Stop codon

4.2. Promoter selection

Microorganisms need to have a promoter capable of initiating the transcription process of the GOI. Generally, those promoters are classified into two groups as native and non-native promoters. In cyanobacteria, promoter classification is similar to other microorganisms; however, those promoters are divided into specific groups.

The first classification of promoters in cyanobacteria is native promoters. They are called natives because they are present in cyanobacteria as part of their nature to encode a specific gene. Native promoters in cyanobacteria were classified in constitutive and inducible promoters.

A classification of constitutive promoters was implemented and divided into strong and weak constitutive promoters. Strong constitutive promoters are P_{psbA2} and P_{cpc}, and they have been used to perform studies in *Synechocystis* to determine the expression of terpenoid genes (Liu & Pakrasi, 2018), a comparison of promoters (Englund et al., 2016), and evaluation of a new plasmid confirming the promoter expression (Englund et al., 2015).

Unlike strong constitutive promoters, weak constitutive promoters (P_{rbc} and P_{rnpB} can work properly, but the level of expression of a GOI is lower (Till et al., 2020). In the case of P_{rbc}, it can be used as promoter of vegetative genes, but the rate of transcription is lower to other native promoters (Liere & Maliga, 2001), while P_{rnpB} expression level is lower than P_{rbc}.

The second group of native promoters is the inducible promoter and, similar to constitutive promoters, are classified into strong and weak inducible promoter. They are considered inducible promoters because they require a physical or chemical factor to initiate the transcription process of the GOI.

Native inducible promoters have more efficient when there are metal ions present (As^{+2}, Fe^{+2}, Cu^{+2}, Ni^{+2}), and some inducible promoters are efficient when the nitrite is present. However, there are inducible promoters that have shown a low level of expression in the presence of metal ions such a As^{+2}, Zn^{+2}, Co^{+2} (Englund et al., 2016). In this group of native inducible promoters, the well-studied promoters are P_{petE}, induced by Cu^{+2} ions (Briggs et al., 1990), P_{nrsB}, induced by Ni^{+2} ions (Englund et al., 2016), and P_{nirA}, induced by nitrites and nitrate (Heirdorn et al., 2011). In Fig. 4.3, all native inducible promoters have been classified to distinguish between strong inducible promoters from which ones are weak promoters.

Non-native promoters are engineered promoters made of a modification of a constitutive promoter or a promoter taken from a different microorganism. When a non-native promoter is engineered from a constitutive promoter, a gene modification of native promoters is carried to enhance the rate of expression. The strong P_{psbA2} promoter was modified by Englund et al.

(2016) in order to obtain strong engineered promoters such as P_{psbA2L}, P_{psbA2M}, and P_{psbA2S}. Their findings unveiled that the longer the promoter sequence was (P_{psbA2L}), the less level of expression was obtained, whereas the shorter the promoter (P_{psbA2S}), the higher-level expression was reached. Conversely, P_{cpc}, another native strong promoter, was engineered by Zhou et al. (2014) to enhance its level of expression.

Indeed, the engineered promoter P_{cpc560} was believed to enhance transcription of heterologous genes inside of cyanobacteria, but its replication inside *Synechocystis* by homologous replication led to a lower level of expression than *E. coli*. However, in further studies carried out by Till et al. (2020), they claimed that the native engineered promoter P_{cpc560} showed inconsistency in its performance despite showing a high strength during the development of this promoter. Unlike native engineered promoters, the inducible engineered promoters were designed to enhance its activity when they were exposed to different light intensity (P_{psbA1} and $P_{psbAIII}$).

The other non-native promoter group is the one engineered from a different organism to cyanobacteria as *E. coli*. Those promoters are based on lac operon and tetracycline-resistance operon from *E. coli*, which are engineered to enhance the transcription in cyanobacteria. lac operon is engineered to biosynthesise strong inducible promoters as P_{tac}, P_{tic}, and P_{trc} which depend on lactose presence. In the case tetracycline- resistance operon, the P_{tet} promoter is induced by anhydrotetracycline (aTc) (Till et al., 2020) but the level of expression inside *Synechocystis* is nondetectable (Heirdorn et al., 2011).

However, Heirdorn et al. (2011) obtained a novel promoter when they modified the downstream region of the P_{tet}, creating a L03 strong promoter that was replicated *Synechocystis*. L03 promoter, or BBa_K1968020, works with the aTc inducer, and its absence with light intensity fluctuations lead to production of basic proteins instead of the GOI, this process is known as leakiness. A flowchart of all promoters which were analysed to determine the one capable of enhancing bisabolene gene is represented in Fig. 4.3.

An analysis of all promoters showed in Fig. 4.3, strong promoters are more suitable to be performed inside *Synechocistys* regardless of their nature. In a study carried out by Liu and Pakrasi (2018), they evaluated 13 promoters in *Synechocystis*, including native constitutive strong promoters as P_{psbA2} and P_{cpcB}, and the weak non-native promoter P_{trc1O}. They concluded that P_{cpcB} promoter showed more activity than the other promoters, in order to confirm this study Till et al. (2020) reported that P_{cpcB} promoter could be used in *Synechocystis* to amplify a heterologous gene. Furthermore, they found that level expression of non-native promoters was lower than native promoters.

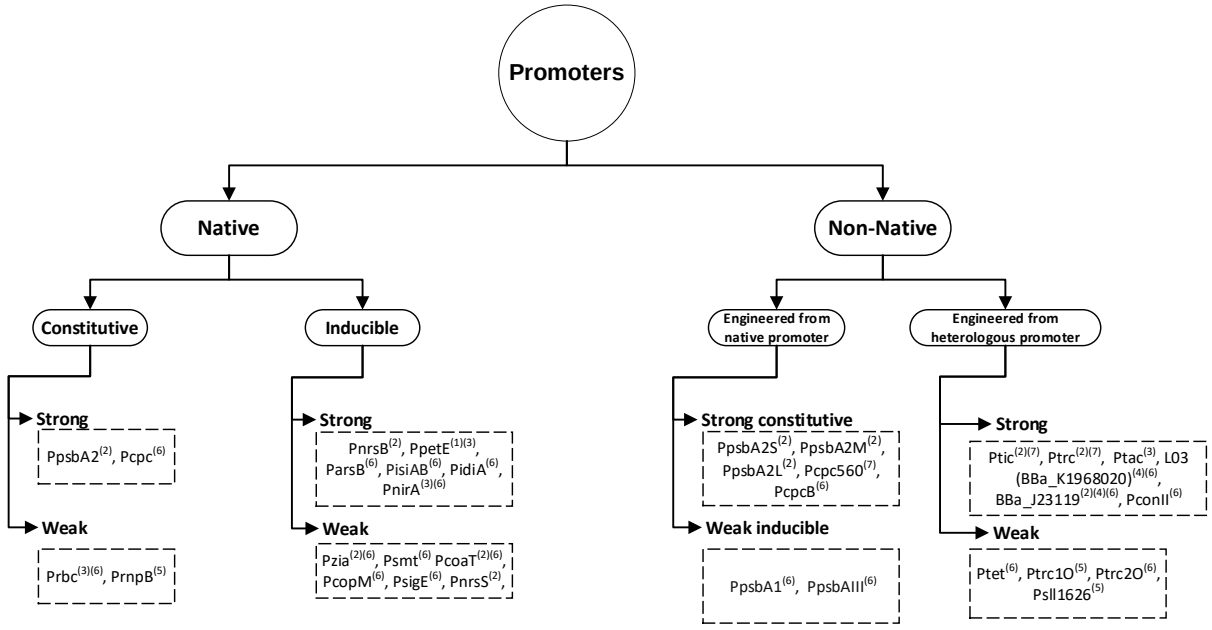


Fig. 4.3 Classification of promoters according to its nature. Elaborated from: (1) Briggs et al. (1990), (2) Englund et al. (2016), (3) Heirdorn et al. (2011), (4) Heirdorn et al. (2011), (5) Liu and Pakrasi (2018), (6) Till et al. (2020), and (7) Zhou et al. (2014).

P_{cpcB} sequence has been registered as biological part (BBa_K2596001), this promoter contains two sections, one labelled as P_{cpcB} promoter 1a (P1a) and the other labelled as P_{cpcB} promoter 1b (P1b) as well as it is depicted in Fig. 4.4.

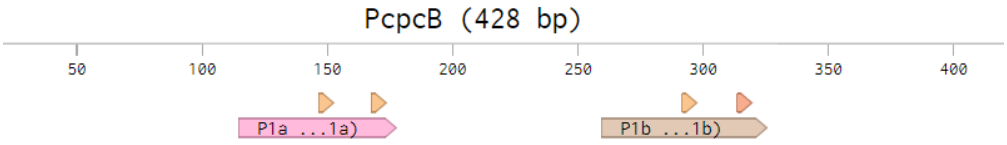


Fig. 4.4 Promoter P_{cpcB} containing P1a and P1b regions, biological part obtained from iGEM and depicted in Benchling.

Sengupta et al. (2020) observed that P_{cpcB} contains two transcriptional start sites (TSS), a site near to the promoter that starts the transcription process. One TSS is placed inside P1a region, in the nucleotide 177 (Fig. 4.5a), whereas the other TSS is in the nucleotide 376 which belongs to the P1b region (Fig. 4.5b). The nucleotide sequence of this promoter is represented in Appendix A.3.

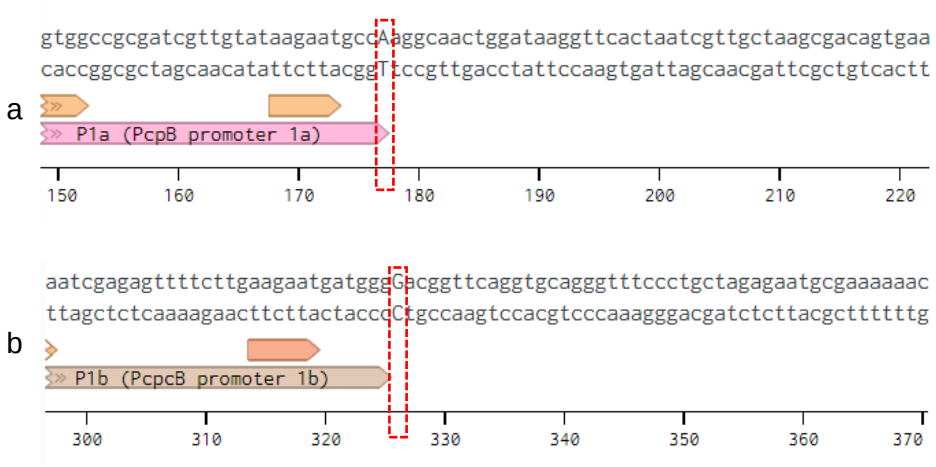


Fig. 4.5 TSS of promoter P_{cpcB} in different regions of the nucleotide sequence. a) TSS in the adenine nucleotide (A), placed in the nucleotide 177 which belongs to P1a region. b) TSS in the guanine nucleotide (G), placed in the nucleotide 326 which belongs to P1a region. Data obtained from iGEM and sequence in Benchling.

4.3. Selection of RBS

Ribosomal Binding Site (RBS) is a small sequence of DNA that can be classified into native RBS which comes from other cyanobacteria and engineered or synthetic RBS.

A RBS is capable of enhancing transcription of a heterologous gene when RBS is linked with a promoter. Heirdorn et al. (2011) indicate that RBS must have the Shine-Dalgrano sequence (SD) which is a nucleotide sequence rich in adenine (A) and guanine (G), as known as “AG region”. They assembled an engineered inducible promoter, P_{trc1O}, with different synthetic RBSs, and they concluded that SD region could lead to an improvement of the level expression of a gene.

Englund et al. (2016) reported a similar study wherein they studied synthetic RBSs by testing the fluorescence of a protein (EYFP). Their findings concluded that a synthetic RBS as BBa_B0035 could reach a fluorescence of 10,000 absorbance units (a.u.) by working together with an engineered promoter as P_{psbA2S}. However, a study performed with native RBSs by Liu and Pakrasi (2018) demonstrated that RBS-psaF sequence (table 4.2) could reach three-fold the fluorescence reported in synthetic RBSs by Englund et al. (2016).

Due to the RBS-psaF sequence is not found as a registered biologic part in iGEM, the nucleotide sequence from table 4.2 is exported to Benchling to be depicted it in a linear map as it is showed in Fig. 4.6.

Table 4.2. Nucleotide sequence of RBS-psaF, the underlined sequence indicated the AG rich region. Data obtained from Liu and Pakrasi (2018)

RBS	Nucleotide sequence
RBS-psaF	TTTAACCAAGGAAACGATTCTT

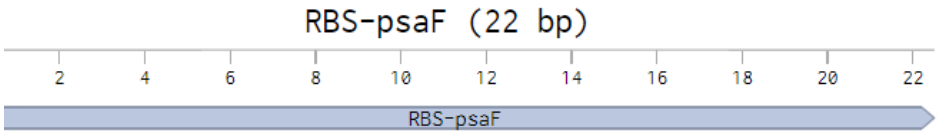


Fig. 4.6 Ribosomal binding site, RBS-psaF, taken from Liu and Pakrasi (2018) and sequenced in Benchling.

4.4. Terminator selection

A terminator is an important element with a small nucleotide sequence capable of regulating the transcription process. During the construction of biological devices, terminators are placed at the end of GOI to drive its expression.

Synechocystis has terminators, however, the library with *Synechocystis* terminators are not available, and the information of cyanobacterium terminators are lower than those from *E. coli*. This finding is similar to Liu and Pakrasi (2018) report, in which they mentioned that *E. coli* terminators are stronger than constitutive terminators from cyanobacteria, and the use of *E. coli* terminators is justified owing to an unavailable library of *Synechocystis* terminators.

In the development of this study, the assembled terminator BBa_B0015 (T7 and rrnB terminators) has been selected because early studies have demonstrated that such a double terminator from *E. coli* can lead to a high-level of expression (Heirdorn et al., 2011). Besides, BBa_B0015 has been used by Englund et al. (2016) to validate that this terminator can both drive the expression of terpenoid gene and replicate inside of *Synechocystis* chromosomal DNA.

Before assembling BBa_B0015 terminator with the GOI, this terminator is represented in a biopart form as it is depicted in Fig. 4.7.

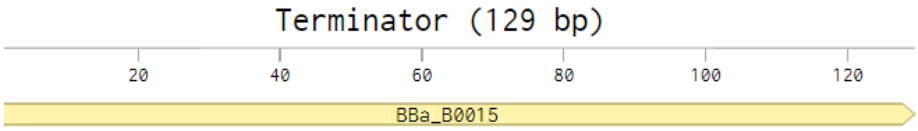


Fig. 4.7 Terminator BBa_B0015 taken from iGEM (2020) and sequenced in Benchling.

Despite having selected the terminator, which is located downstream of the GOI, the likelihood of having non-expression is higher owing to the lack of a transcriptional terminator upstream of the P_{cpcB} promoter. Kelly et al. (2019) reported that elements of a construct needed to be isolated when the homologous recombination inside *Synechocystis* would be carried out to integrate the construct in the chromosomal DNA.

A comparison of Synechocystis terminators with E. coli indicates that ECK120010799 terminator, which belongs to *E. coli*, is capable of enhancing the expression of a construct when that terminator is placed upstream of a promoter. Nucleotide sequence of ECK120010799 was not registered in iGEM, but its sequence was given by Kelly et al. (2019) (Appendix A.4), and hence, this terminator sequence was exported in Benchling as Fig. 4.8 shows. To note, that the terminator in Fig. 4.7 is placed at the end of ag1 gene whereas terminator in Fig. 4.8 is placed before native constitutive promoter P_{cpcB}.

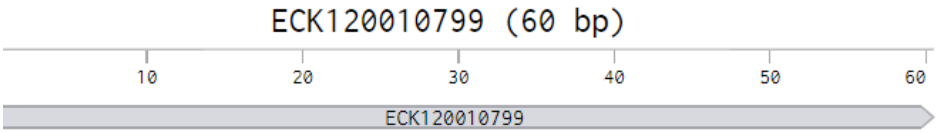


Fig. 4.8 Terminator ECK120010799 taken from Kelly et al. (2019) and sequenced in Benchling.

4.5. Prefix and Suffix addition

Restriction enzymes are in charge of fragmenting a sequence of nucleotides at a specific location capable of creating DNA overhangs. All sequences in Fig. 4.2, Fig. 4.4, Fig. 4.6, Fig. 4.7, and Fig.4.8, are not flanked by any restriction enzymes, which means that those sequences cannot be broken into small DNA fragments.

However, during the construction of BioBricks, there is a need to add restriction enzyme sites to each biopart known as prefix and suffix. Ho-Shing et al. (2012) mention that a

conversion of any DNA sequence can be carried out by adding a BioBrick prefix containing EcoRI, NotI, and XbaI restriction sites, whereas the BioBrick suffix contains SpeI, NotI, and PstI restriction sites.

BioBrick prefix depends on whether the part starts with ATG nucleotide sequence or not as it is illustrated in Fig. 4.9.

Fig. 4.9 BioBrick prefix with its restriction enzyme site sequence. a) BrioBrick prefix sequence when the part starts with a codon sequence different to ATG. b) BioBrick sequence when the part starts with ATG sequence. Data obtained from iGEM (2020).

Unlike BioBrick prefix, BioBrick suffix does not depend on any nucleotide sequence, indeed, BioBrick suffix (Fig. 4.10) is represented by a general sequence that can be used in different parts.

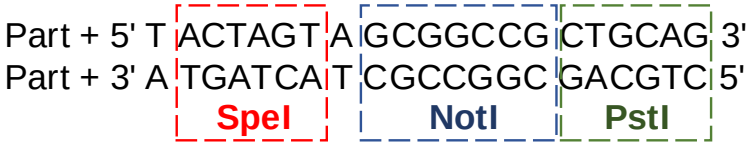


Fig. 4.10 BrioBrick suffix sequence with its restriction enzyme site sequence. Data obtained from iGEM (2020).

First step to construct a BioBrick is to add both prefix and suffix which flank the biopart or element. Prefix must be added upstream of the part as in Fig. 4.9, whereas the suffix must be added in the downstream of the part as it is represented in Fig. 4.10.

In this study, P_{cpcB} promoter, RBS-psaF, BBa_B0015, and ECK120010799 need to contain prefix and suffix restriction enzyme sites. The addition of those restriction enzyme sites is

carried in Benchling by adding the nucleotide sequence shown in Fig. 4.9 and Fig. 4.10 in the flanks of the part.

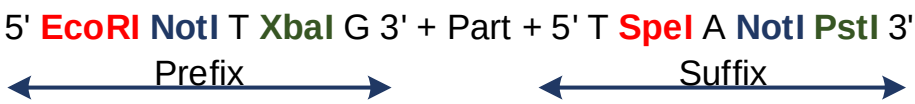


Fig. 4.11 Schematic representation of prefix and suffix restriction enzyme sites flanking a part.

Fig. 4.11 illustrates how a part must be flanked by restriction enzyme sites EcoRI, NotI, XbaI, SpeI, and PstI. P_{cpcB} promoter, RBS-psaF, BBa_B0015, and ECK120010799 are flanked by restriction enzyme sites shown in Fig. 4.9a, whereas ag1 gene shown in Fig. 4.2 is flanked by restriction enzyme sites shown in Fig. 4.9b because this part contains the start codon ATG for methionine.

Restriction enzyme sites for P_{cpcB}, illustrated in Fig. 4.4, are flanked by EcoRI, NotI, and XbaI as prefix, whereas restriction enzyme sites as SpeI, NotI, and PstI are flanked as suffix. The nucleotide sequence of Promoter P_{cpcB} flanked by prefix and suffix is shown, whereas Fig. 4.12 illustrates the final part that is flanking with prefix and suffix by using Benchling.

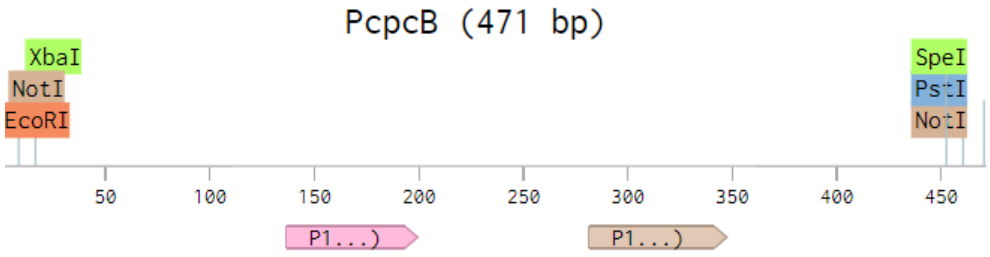


Fig. 4.12 P_{cpcB} promoter flanked by BioBrick prefix and suffix. Nucleotides sequenced in Benchling.

Furthermore, the same procedure was applied to the other parts such as RBS-psaF (Fig. 4.13), BBa_B0015 (Fig. 4.14), and ECK120010799 (Fig. 4.15) which were sequenced in Benchling by adding the BioBrick prefix and suffix shown in Fig. 4.9a and Fig. 4.10.

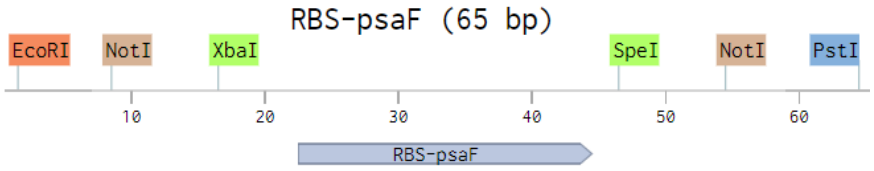


Fig. 4.13 RBS-psaF flanked by BioBrick prefix and suffix. Nucleotides sequenced in Benchling.

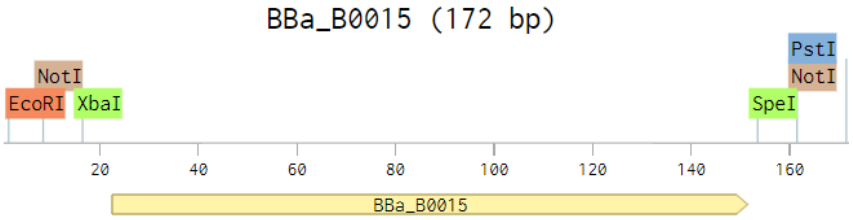


Fig. 4.14 BBa_B0015 terminator flanked by BioBrick prefix and suffix. Nucleotides sequenced in Benchling.

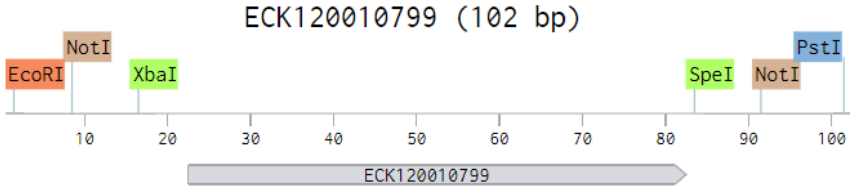


Fig. 4.15 ECK120010799 terminator flanked by BioBrick prefix and suffix. Nucleotides sequenced in Benchling.

ag1 gene has the start codon ATG, the BioBrick prefix to be used is the one represented in Fig. 4.19b. The result of the ag1 gene with the prefix and suffix are similar to the ones represented in Fig. 4.12 to Fig. 4.15 despite the change of the prefix sequence. Fig. 4.16 shows the final result of flanking ag1 gene with those restriction enzyme sites, whereas in Appendix A.5 the sequence of all parts from Fig. 4.12 to Fig 4.16 are shown.

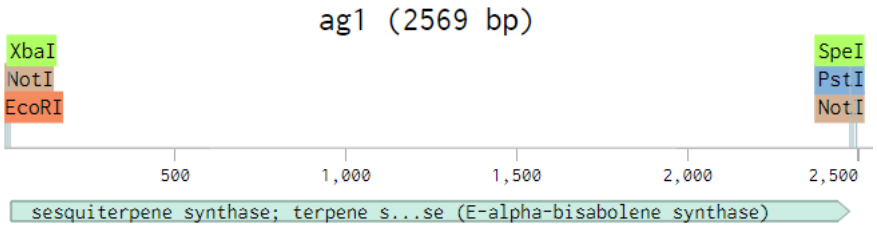


Fig. 4.16 ag1 gene flanked by BioBrick prefix and suffix. Nucleotides sequenced in Benchling.

4.6. BioBrick construction

In order to construct a BioBrick with different parts, each part must be according to BioBirck format. All parts must be flanked by BioBrick prefix (EcoRI and XbaI) and suffix (SpeI and PstI) as it is depicted in Fig. 4.12, Fig. 4.13, Fig. 4.14, Fig. 4.15, and Fig. 4.16. Furthermore, those flanked parts must be inserted inside into a BioBrick plasmid such as pSB1C3.

pSB1C3 plasmid is standard plasmid backbone in which all parts must be inserted to be considered a BioBrick. Furthermore, pSB1C3 plasmid contains the BioBrick prefix and suffix (Fig. 4.17), which facilitate the insertion of bioparts flanked with those restriction enzyme sites.

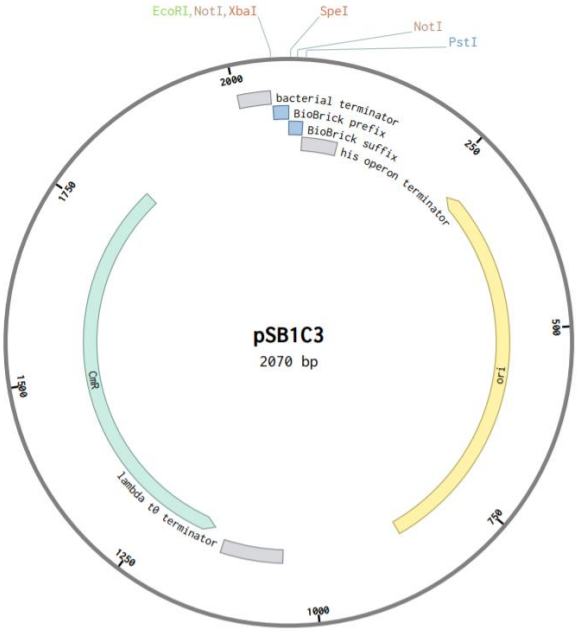


Fig. 4.17 pSB1C3 plasmid with its restriction enzyme sites labelled in the surroundings of the plasmid. Nucleotide sequenced in Benchling.

The insertion of the part inside of the plasmid is carried out by using EcoRI and PstI digestion enzymes to generate DNA overhangs in both the part and plasmid backbone. In the case of parts flanked by BioBrick prefix and suffix, the digestion enzymes cut both ends in specific nucleotide sequences (Fig. 4.18a), generating DNA overhangs (Fig. 4.18b) capable of being inserted inside the plasmid backbone. However, the plasmid backbone needs to be digested by the same restriction enzyme sites (EcoRI and PstI) to make the plasmid compatible with the insert.

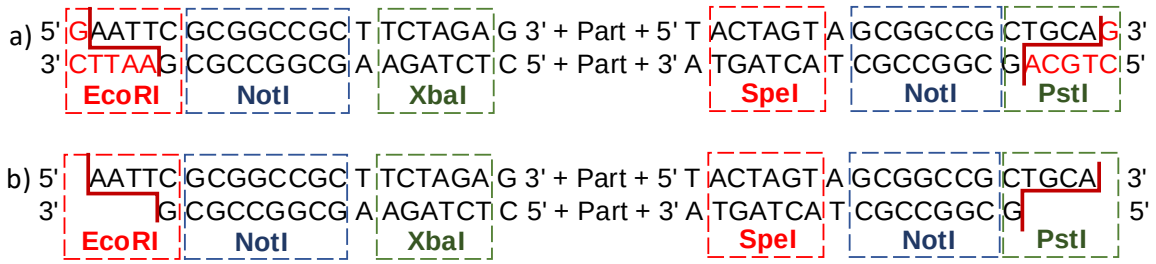


Fig. 4.18 Digestion process of the parts containing BioBirck prefix and suffix. a) Specific site wherein EcoRI and PstI enzymes digest the prefix and suffix. b) Generation of an insert to be placed inside the plasmid backbone.

In the plasmid backbone, the digestion process is performed with the same digestion enzymes (EcoRI and PstI), instead of creating an insert, the circular backbone plasmid is digested to generate a site wherein the insert is placed. The insert and the digested plasmid backbone need to be compatible to get assembled; the created insert (Fig. 4.19a) has sticky ends which are compatible with the digested ends of plasmid backbone (Fig. 4.19b). During the assembly process, both sticky ends are ligated by generating the construction of a BioBrick (Fig. 4.19c).

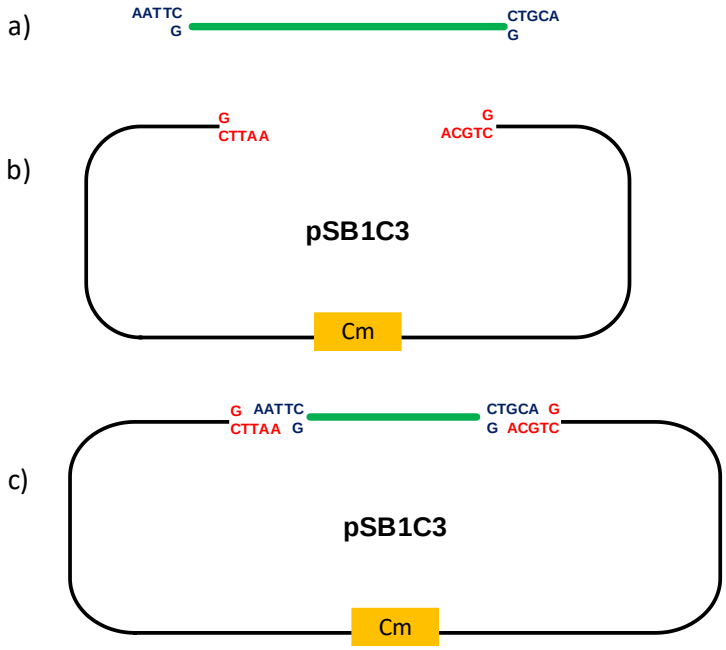


Fig. 4.19 a) Schematic representation of a part after being digested by EcorI and PstI enzyme. b) Plasmid backbone (pSB1C3) after digestion process. c) Insertion of the insert inside the

digested plasmid backbone. To note, in c) the ends of the insert and the plasmid backbone are compatible to regenerate the EcoRI and PstI sites.

In this study, the parts that are flanked by BioBrick prefix and suffix (Fig. 4.12, Fig. 4.13, Fig. 4.14, Fig. 4.15, and Fig. 4.16) need to be digested and assembled inside the backbone plasmid by using digestion process with EcoRI and PstI enzymes for both bioparts and plasmid backbone. Afterwards, the ligation process is carried out to assure the regeneration of the EcoRI and PstI sites inside the construct. To note, digestion and ligation process is performed in *in silico* by using Benchling.

The P_{cpcB} promoter is digested by EcoRI and PstI, and then, it is inserted inside pSB1C3 plasmid to generate a construct that contains the promoter and plasmid backbone.

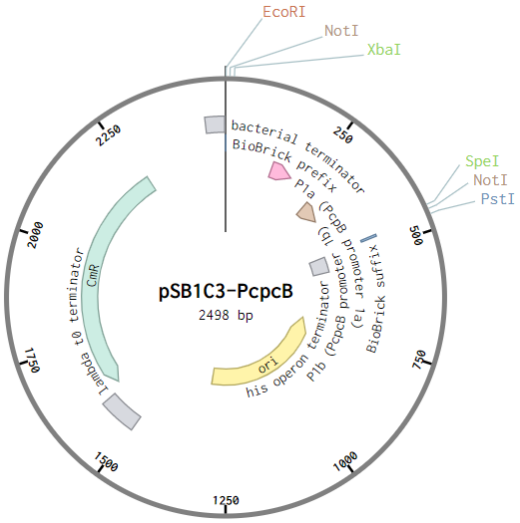


Fig. 4.20 Construct containing the insert (P_{cpcB}) and plasmid backbone (pSB1C3)

Similarly, the GOI (ag1), RBS (RBS-psaF), and terminators (BBa_B0015 and ECK120010799) are assembled inside pSB1C3. Fig. 4.21 shows the final construct of each element after being inserted in the backbone.

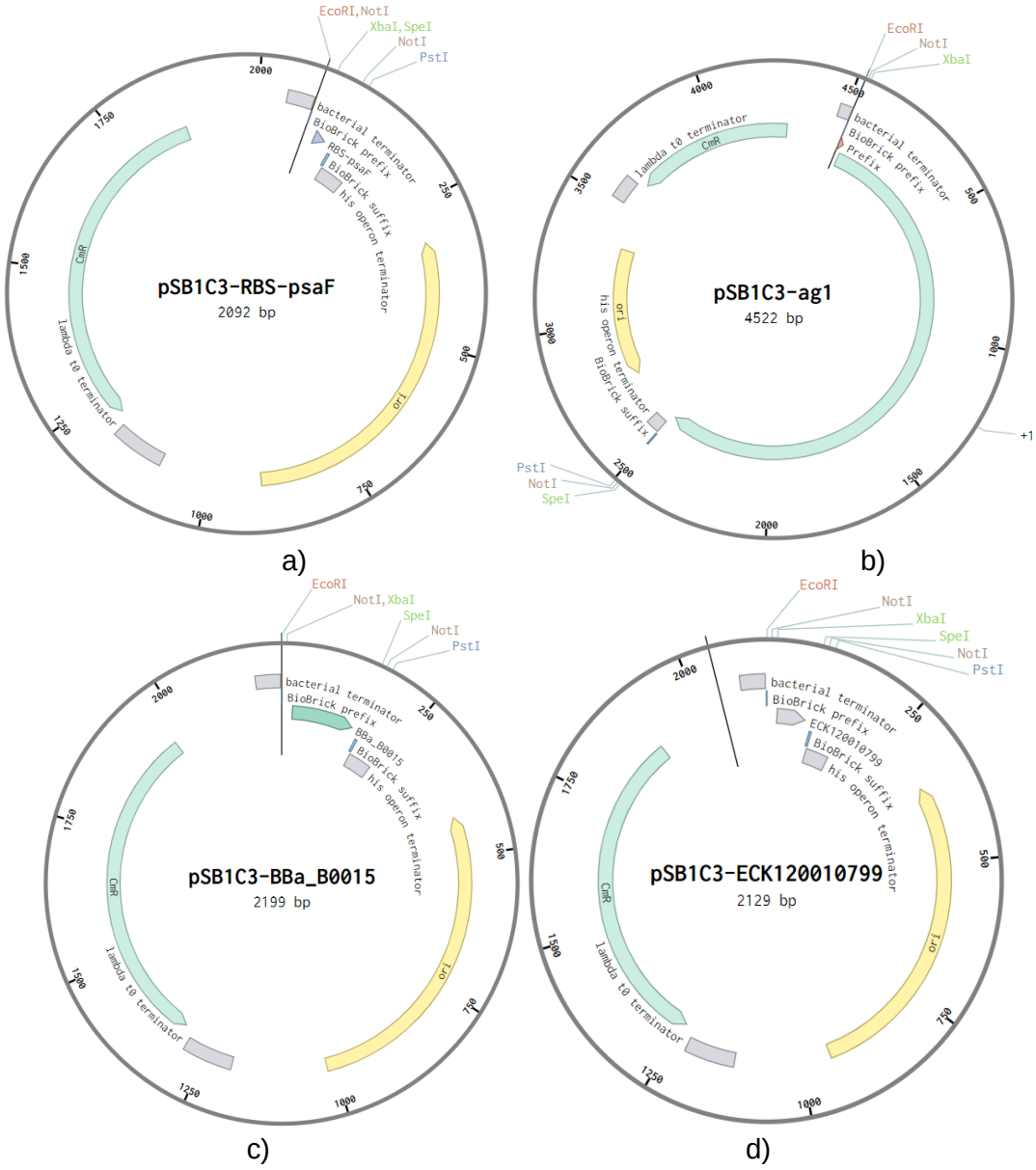


Fig. 4.21 *in silico* design of constructs containing different parts. a) Backbone plasmid containing the RBS sequence. b) Backbone plasmid containing the ag1 gene (bisabolene gene). c) Backbone plasmid containing BBa_B0015 terminator. d) Backbone plasmid containing ECK120010799 terminator. All constructs were designed by using Benchline.

4.7. Setting up digestion and ligation for working in laboratory

Construction of BioBricks shown in Fig. 4.20 and Fig. 4.21 are considered an *in-silico* design because they are accomplished by using a software as Benchling. When BioBrick construction is performed in a laboratory, some parameters need to be taken into account to obtain a similar construct to the one obtained *in silico*. In New England Biolabs (NEB), there are determined protocols to follow in order to digest and ligate the parts and the plasmid backbone properly in laboratory. NEBCloner v1.5.0 from NEB is a workflow simulator in which determined products for digestion and ligation are listed, as well as protocols for each restriction enzymes.

All constructs are according to BioBrick format which means that restriction enzyme sites are standard, those enzymes are EcoRI, XbaI, SpeI, and PstI. In the case of the ligation process, Fig. 4.19 indicates that the restriction enzymes sites to use are EcoRI and PstI to insert the parts inside of pSB1C3 plasmid. NEBCloner v1.5.0 sets up that a double digestion can be implemented.

Restriction Enzyme Single/Double Digestion

EcoRI-HF

PstI-HF ✖ clear 2nd selection

Digest in CutSmart Buffer Show Detailed Protocol

Name	Cat #	Temp °C	Supplied Buffer	Add SAM	% Activity in NEBuffer			
					1.1	2.1	3.1	CutSmart ®
EcoRI-HF®	R3101	37	CutSmart Buffer	No	10	100	10	100
PstI-HF®	R3140	37	CutSmart Buffer	No	10	75	50	100

Name	Time-Saver ™	Heat Inactivation (°C)	Methylation Sensitivity
EcoRI-HF®	Yes	65	cpg (Blocked by Some Combinations of Overlapping)
PstI-HF®	Yes	No	None

Fig. 4.22 Double restriction enzyme to use in digestion, according to NEBCloner v1.5.0.

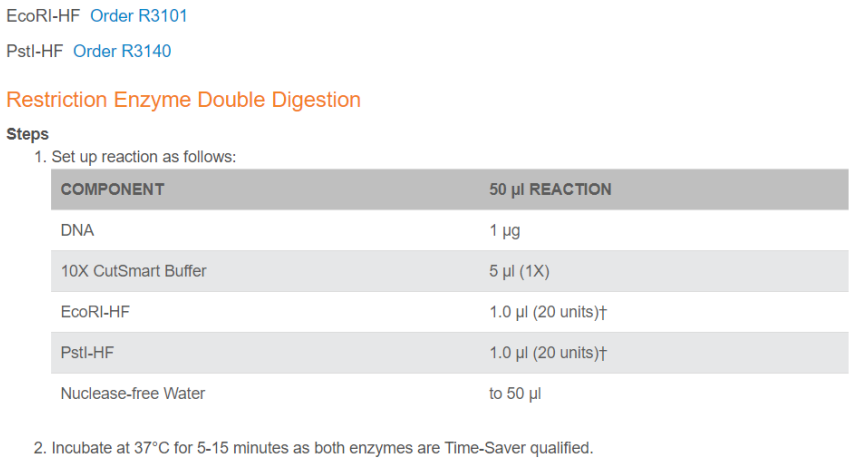
NEBCloner v1.5.0 recommends using those enzymes in their High-Fidelity version (EcoRI-HF and PstI-HF) owing to this version simplify the digestion, according to Fig. 4.22, in one buffer solution (CutSmart Buffer). The digestion temperature for both restriction enzymes is set up in 37°C without affecting enzyme functionalities.

However, during the assembly process of all parts, there is a different combination of enzymes to digest. The more usual enzyme combination is EcoRI/XbaI, EcoRI/SpeI, SpeI/PstI, XbaI/PstI, and EcoRI/PstI (Ho-Shing et al., 2012). In table 4.3, a summary of buffers, temperature, and a combination of restriction enzymes for digestion are summarised.

Table 4.3. Combination of enzymes for digestion, temperature for digestion, and buffer solution to use. Data taken from NEBCloner v1.5.0 and Ho-Shing et al. (2012).

Restriction enzymes		Temperature	Buffer
EcoRI-HF	XbaI	37°	CutSmart Buffer
EcoRI-HF	SpeI-HF	37°	CutSmart Buffer
SpeI-HF	PstI-HF	37°	CutSmart Buffer
XbaI	PstI-HF	37°	CutSmart Buffer
EcoRI-HF	PstI-HF	37°	CutSmart Buffer

HF enzymes ensure that digestion is well-performed due to CutSmart Buffer enhances enzyme activity by reducing incubation time and then a simplification of digestion. According to the protocol of NEBCloner v1.5.0, those HF enzymes are classified as Time-Saver by performing the incubation at 37°C during 5 to 15 minutes (Fig. 4.23). Besides, there is a sequence of reactions and sample quantities that must be followed during digestion in order to obtain a high-quality result (Fig. 4.23).



EcoRI-HF Order R3101

PstI-HF Order R3140

Restriction Enzyme Double Digestion

Steps

1. Set up reaction as follows:

COMPONENT	50 µl REACTION
DNA	1 µg
10X CutSmart Buffer	5 µl (1X)
EcoRI-HF	1.0 µl (20 units)†
PstI-HF	1.0 µl (20 units)†
Nuclease-free Water	to 50 µl

2. Incubate at 37°C for 5-15 minutes as both enzymes are Time-Saver qualified.

Fig. 4.23 Digestion protocol of EcoRI-HF and PstI-HF, protocol sequence obtained from NEBCloner v1.5.0.

Products and protocols for ligation of DNA overhangs created in digestion are different to those obtained in Fig. 4.22. Instead of selecting the enzymes for digestion from NEBCloner v1.5.0, there is a questionnaire that must be filled in order to determine which kind of Ligase enzyme must be used for ligation. Ligation questionnaire must be completed according to Fig. 4.24.

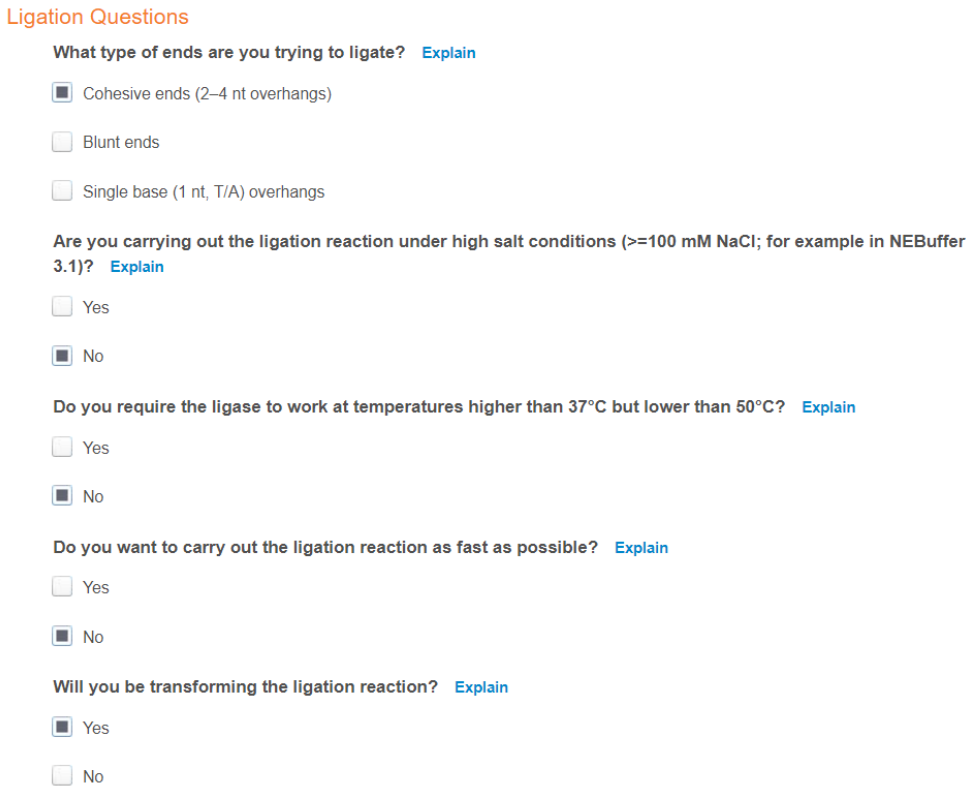
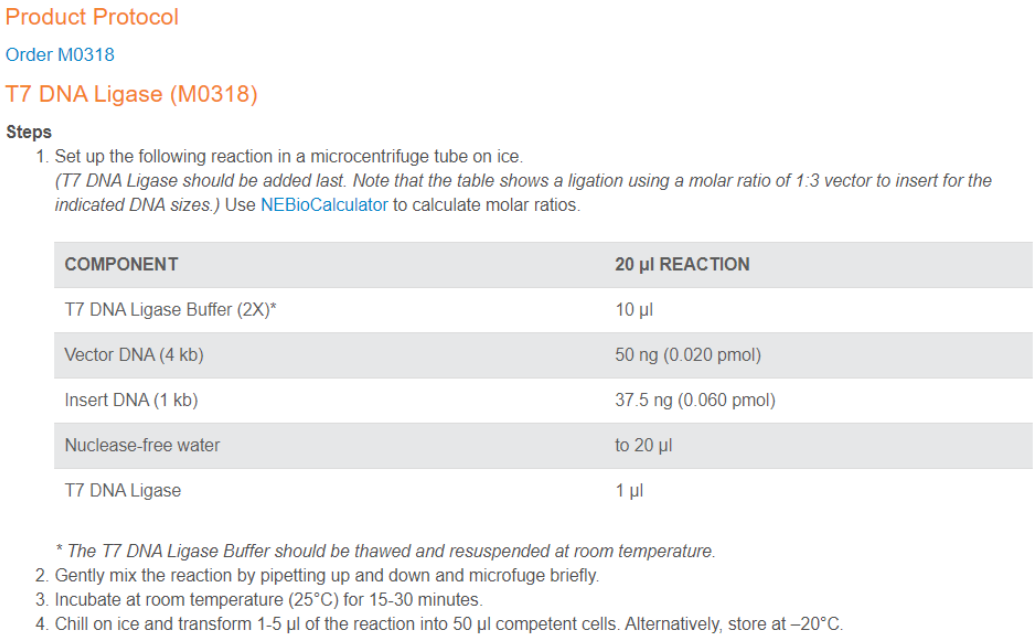


Fig. 4.24 Part of ligation questions filled with the answers to obtain the enzyme for ligation. Questionnaire obtained from NEBCloner v1.5.0.

Once question section is completed, the programme selects the ligation product (Fig. 4.25) and the protocol to be used for ligation. However, quantities of vector and insert DNA must be calculated by using NEBioCalculator from NEBCloner v1.5.0. The aim of the calculator is to estimate the exact mass of DNA of both insert and vector in a molar ratio (Fig. 4.26) because the protocol is designed to be used with a different length of both insert and vector. For example, the ag1 gene has a length of 2569 bp and pSB1C3 plasmid has a length of 2070 bp (Fig. 4.26), whereas the protocol work with 4000 bp for vector and 1000 bp for insert. To note, any insert and vector DNA can be calculated with this ligation calculator.



Product Protocol

Order M0318

T7 DNA Ligase (M0318)

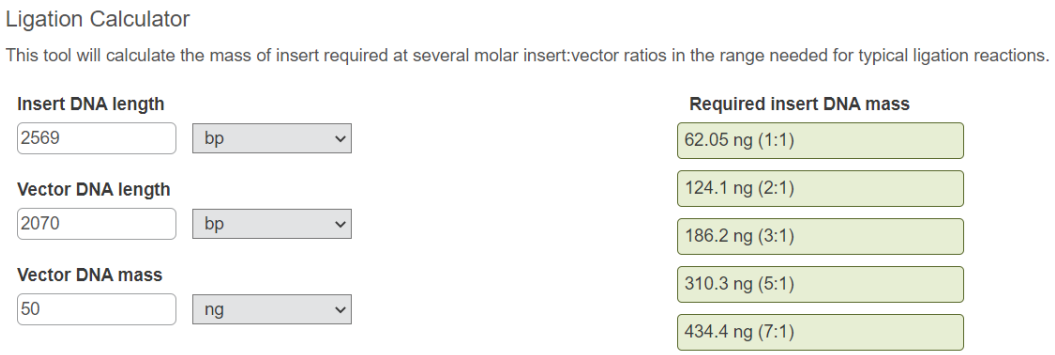
Steps

1. Set up the following reaction in a microcentrifuge tube on ice.
   (T7 DNA Ligase should be added last. Note that the table shows a ligation using a molar ratio of 1:3 vector to insert for the indicated DNA sizes.) Use NEBioCalculator to calculate molar ratios.

COMPONENT	20 µl REACTION
T7 DNA Ligase Buffer (2X)*	10 µl
Vector DNA (4 kb)	50 ng (0.020 pmol)
Insert DNA (1 kb)	37.5 ng (0.060 pmol)
Nuclease-free water	to 20 µl
T7 DNA Ligase	1 µl

   ** The T7 DNA Ligase Buffer should be thawed and resuspended at room temperature.*
2. Gently mix the reaction by pipetting up and down and microfuge briefly.
3. Incubate at room temperature (25°C) for 15-30 minutes.
4. Chill on ice and transform 1-5 µl of the reaction into 50 µl competent cells. Alternatively, store at –20°C.

Fig. 4.25 Protocol for ligation with T7 DNA ligase, data obtained from NEBCloner v1.5.0.



Ligation Calculator

This tool will calculate the mass of insert required at several molar insert:vector ratios in the range needed for typical ligation reactions.

Insert DNA length: 2569 bp

Vector DNA length: 2070 bp

Vector DNA mass: 50 ng

Required insert DNA mass

- 62.05 ng (1:1)
- 124.1 ng (2:1)
- 186.2 ng (3:1)
- 310.3 ng (5:1)
- 434.4 ng (7:1)

Fig. 4.26 Ligation calculation for an insert (ag1 gene) and a vector (pSB1C3) DNA for an *in-silico* design. Data calculated by using NEBCloner v1.5.0.

4.8. Assembly of BioBrick constructs

Fig. 4.20 and Fig. 4.21 contain the pSB1C3 plasmid backbone with a specific gene as promoter, RBS, GOI, and terminators. However, none of those constructs can work individually to drive a gene expression when they are inserted inside of *Synechocystis* to produce bisabolene.

Instead of having several constructs without driving any gene expression, an assembly of those are required in order to have an initiator of transcription (P_{cpcB}) and a terminator (BBa_B0015) of such transcription. In order to accomplish with the assembly of those parts,

the Standard Assembly Method of BioBricks is performed to obtain a construct similar to Fig. 4.20 and Fig. 4.21. When an insert is in a BioBrick form, it means that the insert is inside of the backbone plasmid, the digestion process uses a series of enzyme combinations according to table 4.3.

Before digesting constructs from Fig. 4.20 and Fig. 4.21, a schematic design has to be done in order to know how the parts might be assembled and which ones must be in back and forward. Synthetic Biology Open Language (SBOL) enables a graphic representation of *in silico* biological parts by using a standardized language for each interchangeable biologic part (Fig. 4.27).

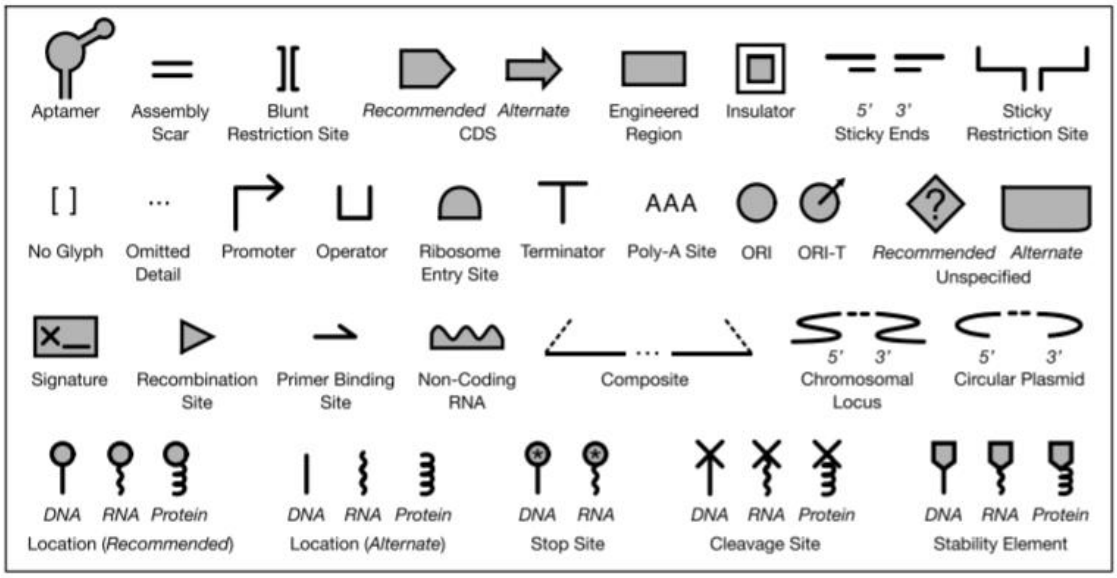


Fig. 4.27 Representation of genetic elements in SBOL Visual Open Language. Taken from (Beal et al., 2019).

One characteristic of SBOL is that interchangeable parts must follow a sequence to be considered standard. The sequence involves placing promoter in the beginning and the terminators at the end. In the middle of those elements, the RBS and a coding sequence have to be placed.

In this study, there are two terminators (BBa_B0015 and ECK120010799), and their location depends on the function they are in charge. ECK120010799 works as a strong terminator to insulate constructs thus, its position in the SBOL is upstream the promoter, whereas BBa_B0015 is a strong promoter to control gene expression during transcription thus, its position is downstream ag1 gene. Fig. 4.28 illustrates how all parts must be assembled to be considered a construct capable of driving the expression of ag1 gene.

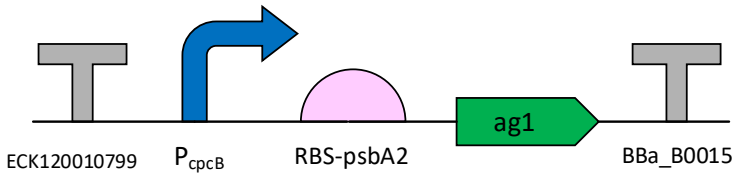


Fig. 4.28 SBOL diagram by assembling parts of Fig. 4.20 and Fig. 4.21.

Due to parts are inside a backbone plasmid (Fig. 4.20 and Fig. 4.21), the digestion process must be completed in order to obtain an insert and a vector with sticky ends, similar to Fig. 4.19a and Fig. 4.19b. The selection of enzymes for digestion is according to table 4.3, and the ligation enzyme is T7 DNA ligase, according to NEBCloner v1.5.0.

Following SBOL representation of elements depicted in Fig. 4.27, Fig. 4.20 and Fig. 4.21 can be represented as it is shown in Fig. 4.29 to set up a standard language in the development of this study. That schematic representation enables to understand which elements are inserted inside the backbone plasmid shown in Fig. 4.20 and 4.21.

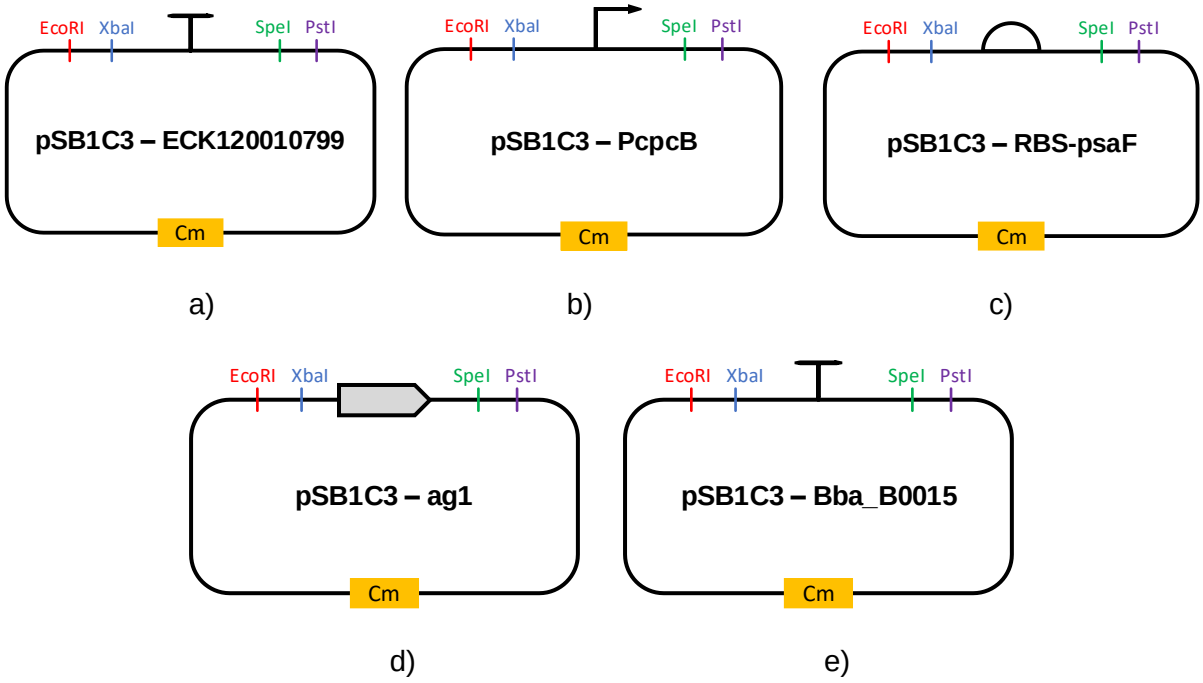


Fig. 4.29 SBOL diagram of Fig. 4.20 and 4.21. a) plasmid backbone containing ECK120010799 terminator. b) plasmid backbone containing P_{cpcB} promoter. c) plasmid backbone containing the RBS, RBS-psaF. d) plasmid backbone containing the gene of interest, ag1. e) plasmid backbone containing BBa_B0015 terminator

In order to obtain the final construct (Fig. 4.28), it is needed to assembly all parts from Fig. 4.29 by using a series of digestion and ligation process, as it is shown in Fig. 4.30. As part of the digestion process, Fig. 4.29a and Fig. 4.29b are digested by EcoRI-SpeI and EcoRI-XbaI enzymes. Those enzymes generate an insert from the BioBrick containing the terminator ECK120010799 and a destination vector from the BioBrick that contains the promoter P_{cpcB}.

Once complementary DNA overhangs are obtained, the insert that contains the ECK120010799 sequence with the prefix and suffix is inserted inside the digested site of the plasmid backbone containing the promoter (Fig. 4.30b). However, those complementary DNA overhangs cannot be joined by themselves, and T7 DNA ligase is used to seal nicks by regenerating both the prefix and suffix and creating a scar site between the terminator and the promoter (Fig. 4.30c).

This new construct (Fig. 4.30c) and the construct with the RBS are digested (Fig. 4.31a) with EcoRI-SpeI and EcoRI-XbaI enzymes to isolate an insert containing both a promoter and a terminator, in the other construct a gap is created to accept the generated insert (Fig. 4.31b). Both DNA overhangs are ligated by T7 DNA ligase to seal and join the insert inside the backbone (Fig. 4.31c). During the ligation, a new scar site is generated between the promoter and the RBS, which is a particular feature of a standard assembly, thus assembling three parts inside one plasmid backbone.

The BioBrick in Fig. 4.31c and the BioBrick, containing ag1 gene, (Fig. 4.32a) are digested. EcoRI-SpeI enzymes isolate an insert from Fig. 4.31c, whereas EcoRI-XbaI creates a gap in the BioBrick that contains the ag1 gene (Fig. 4.32b). The insert and the generated gap are complementary, which means that the insert can be placed inside the gap by ligation with T7 DNA ligase to assemble both complementary DNA overhangs, creating a new BioBrick (Fig. 4.32c).

Similarly, the BioBrick from Fig. 4.32c and the BioBrick with the terminator BBa_B0015 (Fig. 4.33a) are digested respectively by EcoRI-SpeI and EcoRI-XbaI enzymes (Fig. 4.33b). Ligation is performed to assemble both BioBrick in order to create a new BioBrick (Fig. 4.33c) containing all the parts as well as it was depicted in Fig. 4.28.

Indeed, Fig. 4.33c and Fig. 4.28 are equivalents, but the difference is the representation in each figure to illustrate the standard assembly (Fig. 4.33) and the SBOL diagram in a standard language (Fig. 4.28).

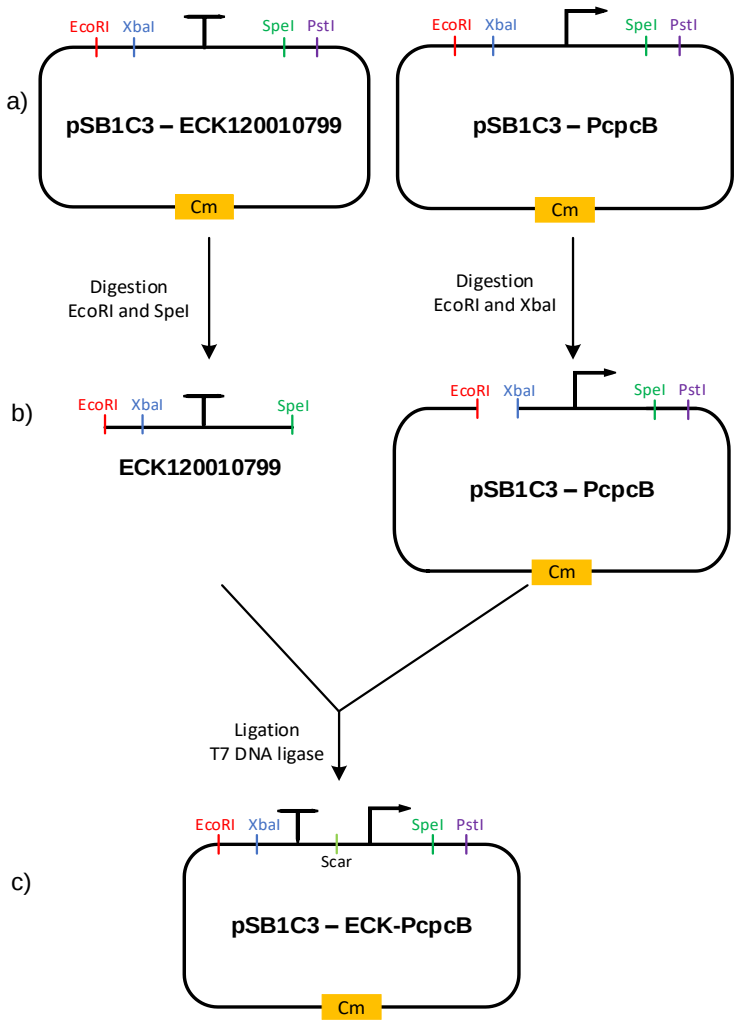


Fig. 4.30 Schematic of the standard assembly. a) BioBircks containing ECK120010799 terminator and P_{cpcB} promoter. b) Digestion process of each BioBrick with digestion enzymes. c) Ligation of complementary DNA overhangs to construct a BioBrick containing both BioBrick parts from a).

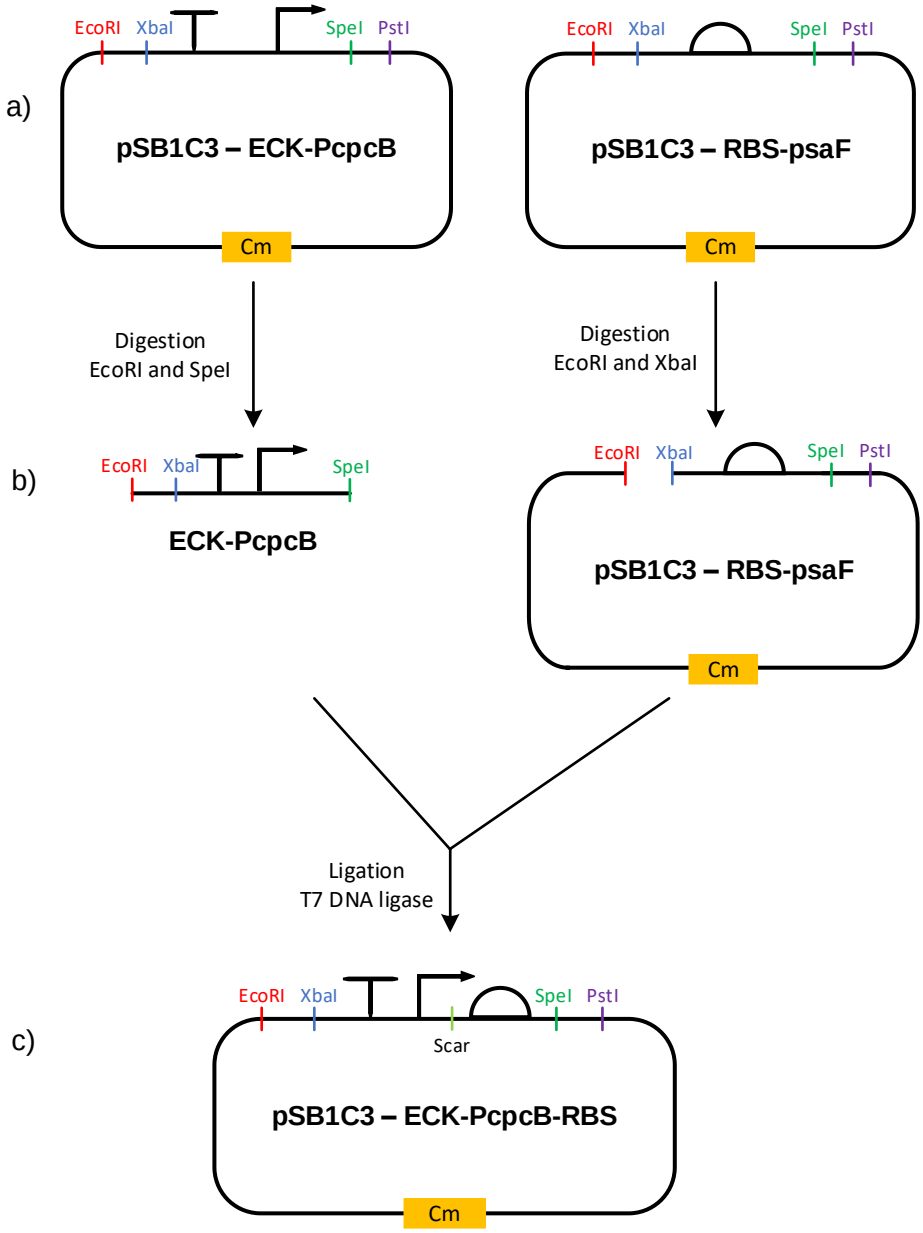


Fig. 4.31 Schematic of the standard assembly. a) BioBrick containing constructed parts from Fig. 4.30c and BioBrick containing RBS-psaF insert. b) Digestion process of each BioBrick with digestion enzymes. c) Ligation of complementary DNA overhangs to construct a BioBrick containing both BioBrick parts from a).

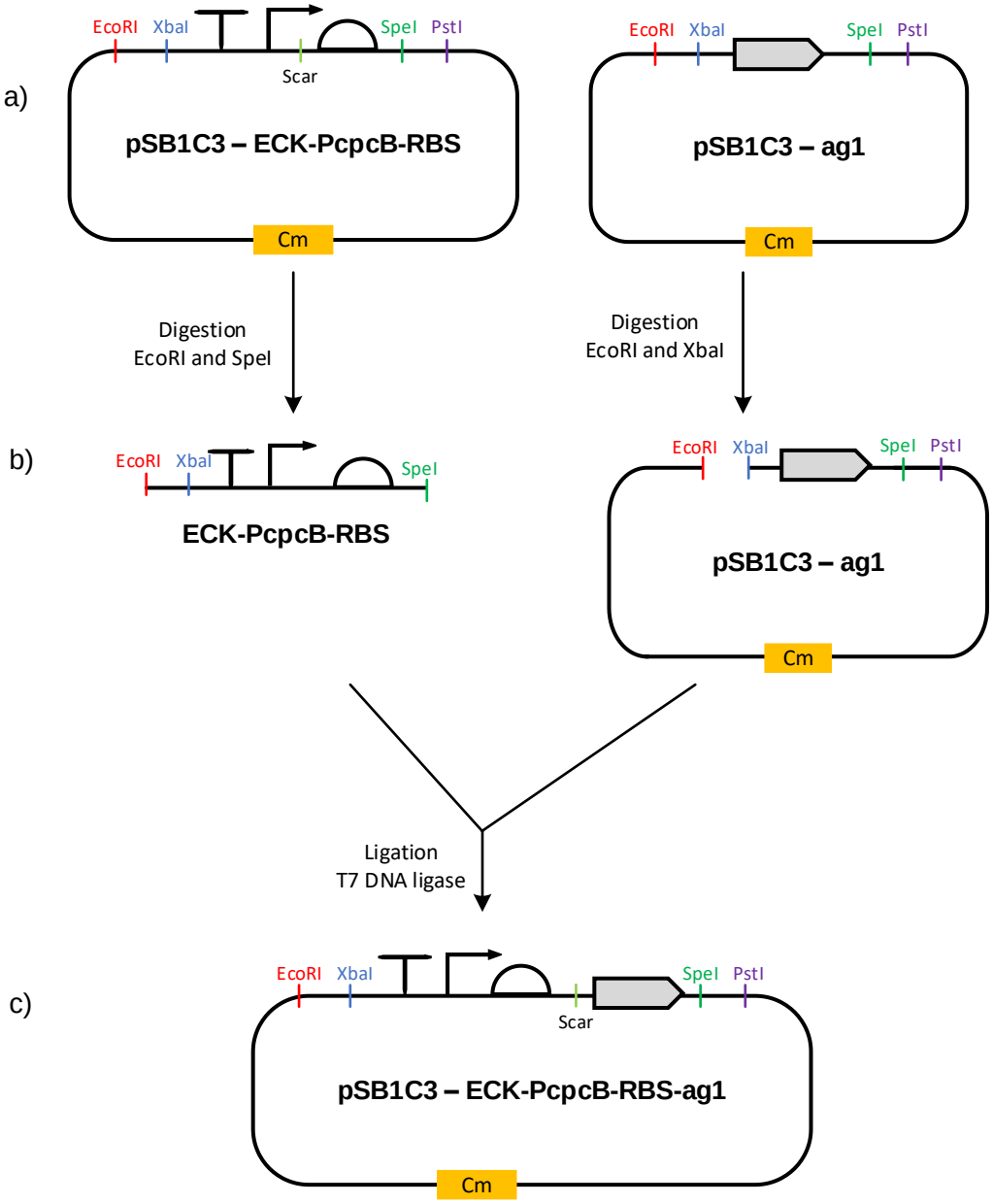


Fig. 4.32 Schematic of the standard assembly. a) BioBrick containing constructed parts from Fig. 4.31c and BioBrick containing ag1 gene insert. b) Digestion process of each BioBrick with digestion enzymes. c) Ligation of complementary DNA overhangs to construct a BioBrick containing both BioBrick parts from a).

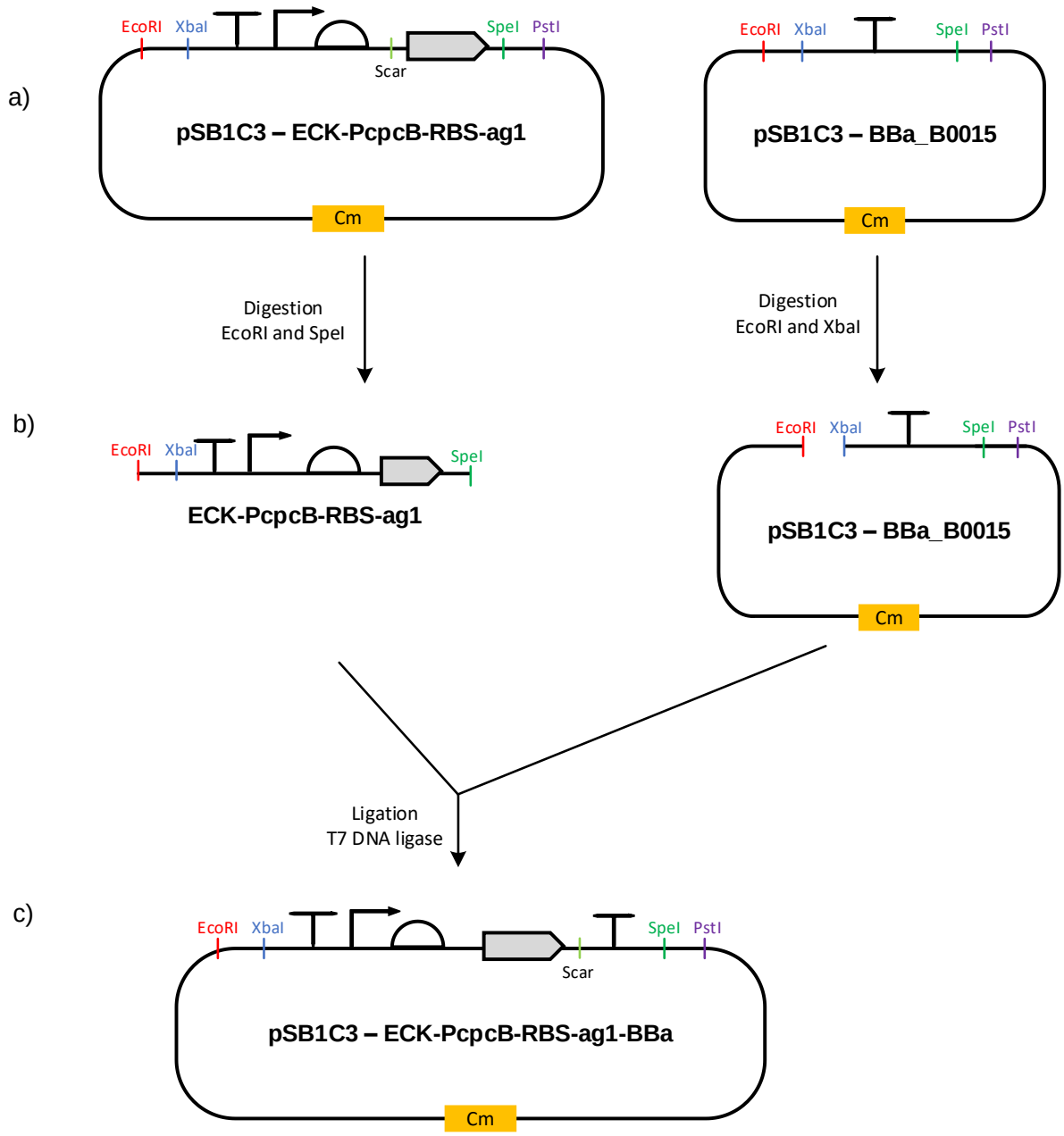


Fig. 4.33 Schematic of the standard assembly. a) BioBrick containing constructed parts from Fig. 4.32c and BioBrick containing BBa_B0015 terminator insert. b) Digestion process of each BioBrick with digestion enzymes. c) Ligation of complementary DNA overhangs to construct a BioBrick containing both BioBrick parts from a).

In Benchling, those constructs can be summarised in Fig. 4.34, in which all inserts are placed inside the pSB1C3 plasmid by following the same digestion and ligation process to obtain Fig. 4.33c.

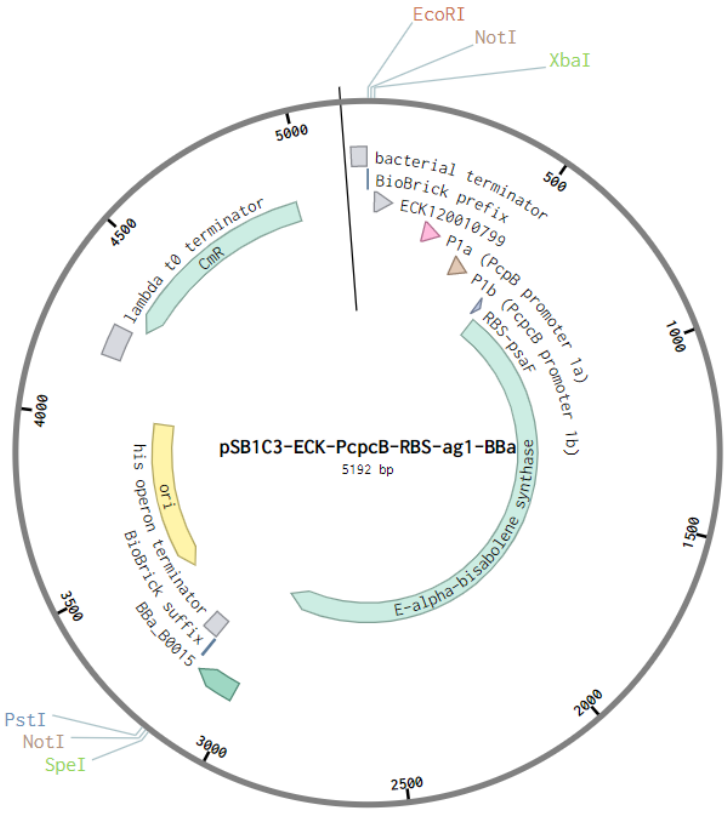
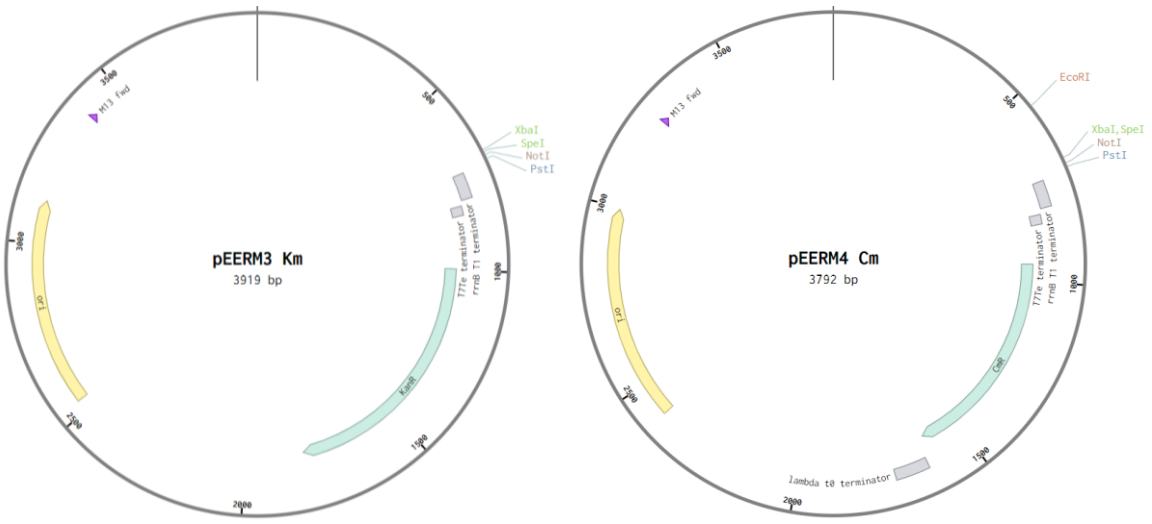


Fig. 4.34 BioBrick obtained after several digestion and ligation processes, as it was described in Fig. 4.30, Fig. 4.31, Fig. 4.32 and Fig. 4.33, by using Benchling.

4.9. Design in *in silico* of BioBrick inside of vector plasmid

Upon sequence analysis of BioBrick with all parts together (Fig. 4.34), this BioBrick must be inserted inside a vector plasmid (pEERM3 or pEERM4). pEERM vector plasmids were created to be replicated inside the chromosomal DNA of *Synechocystis*, locating in the neutral sites of its DNA by using homologous recombination (Englund et al., 2015).

There are some differences between pEERM3 and pEERM4. pEERM4 plasmid has a well-defined prefix and suffix BioBrick, whereas pEERM3 plasmid does not have the EcoRI restriction enzyme site (Fig. 4.35). Furthermore, pEERM3 contains the kanamycin (Km) resistance cassette, whereas pEERM4 has the chloramphenicol (Cm) resistance cassette.



a) b)

Fig. 4.35 a) pEERM3 plasmid containing Km cassette and the lack of EcoRI enzyme restriction site. b) pEERM4 plasmid containing Cm cassette and the prefix and suffix BioBrick. Nucleotide sequence obtained from Addgene and sequenced in Benchling.

Despite these differences, both plasmids can be used as vectors during homologous recombination. However, they need to contain some parts that enable to drive the production of bisabolene. The insertion of parts inside of pEERM plasmids is performed similar to the schematic representation in Fig. 4.30 by using digestion and ligation enzymes, however, as first step a SBOL diagram must be drawn to represent the location of each part from 4.28 inside of pEERM plasmids (Fig. 4.35).

pEERM3 Km has three terminators, one of them is placed downstream of kanamycin gene, and the other two are located upstream of kanamycin promoter (P_{Km}). The RBS and kanamycin gene are located between P_{Km} and the kanamycin terminator. SBOL diagram of the pEERM3 Km plasmid (Fig. 4.36a) shows how its elements are arranged in the plasmid.

pEERM4 Cm plasmid has four terminators, two upstream of the promoter (P_{Cm}) and two downstream of chloramphenicol (Cm) gene. Its RBS is placed between P_{Cm} and the Cm gene. The SBOL diagram of pEERM4 Cm (Fig. 4.36b) illustrates the arrangement of those elements inside the plasmid.

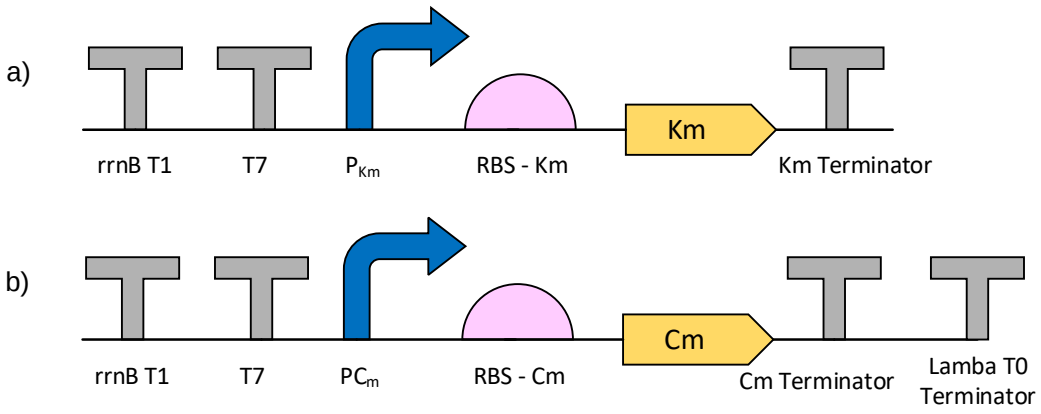


Fig. 4.36 SBOL diagrams**.** a) SBOL diagram of pEERM3 Km elements. b) SBOL diagram of pEERM4 Cm elements.

In order to have a plasmid vector capable of being replicated inside the chromosomal DNA of *Synechocystis*, the constructed BioBrick (Fig. 4.28) must be integrated in both pEERM plasmid vectors, according to Fig. 4.37.

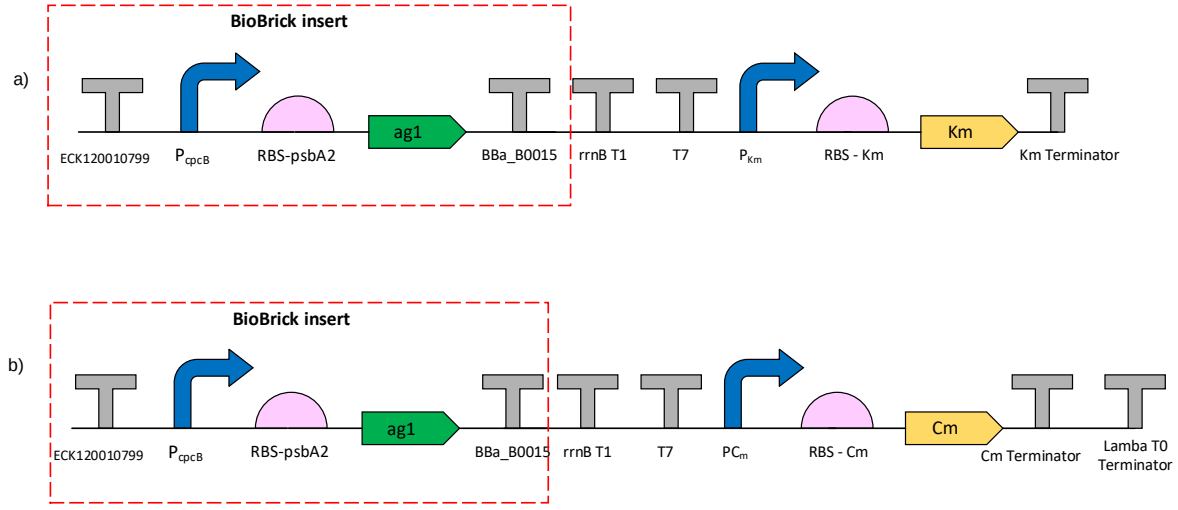


Fig. 4.37 Insertion of a BioBrick inside pEERM plasmid vectors**.** a) SBOL diagram of BioBrick inside of pEERM3 Km. b) SBOL diagram of BioBrick inside of pEERM4 Cm.

The achievement of Fig. 4.37 is reached by using digestion and ligation, similar to the schematic representation of Fig. 4.33. In this case, an insert is generated from BioBrick of Fig. 4.34, and the plasmid backbone is either pEERM3 Km or pEERM4 Cm.

To insert a part of the BioBrick from Fig. 4.34 inside of pEERM3 Km, it is needed that both plasmids have the same restriction enzyme sites (Fig. 4.38a). As it was noted, pEERM3 Km has a lack of EcoRI restriction enzyme site, but it has the other three restriction enzyme sites unaltered (XbaI, SpeI, and PstI). Despite that absence, digestion process is performed by

using different enzymes to create an insert (XbaI-PstI) and a gap in the backbone plasmid (SpeI-PstI) capable of integrating the insert (Fig. 4.38b). Those complementary DNA overhangs are ligated by using T7 DNA ligase enzyme that seals the gaps and joins both DNA overhangs, creating a new construct using pEERM3 Km as plasmid backbone (Fig. 4.38c) which is similar to Fig. 4.37a.

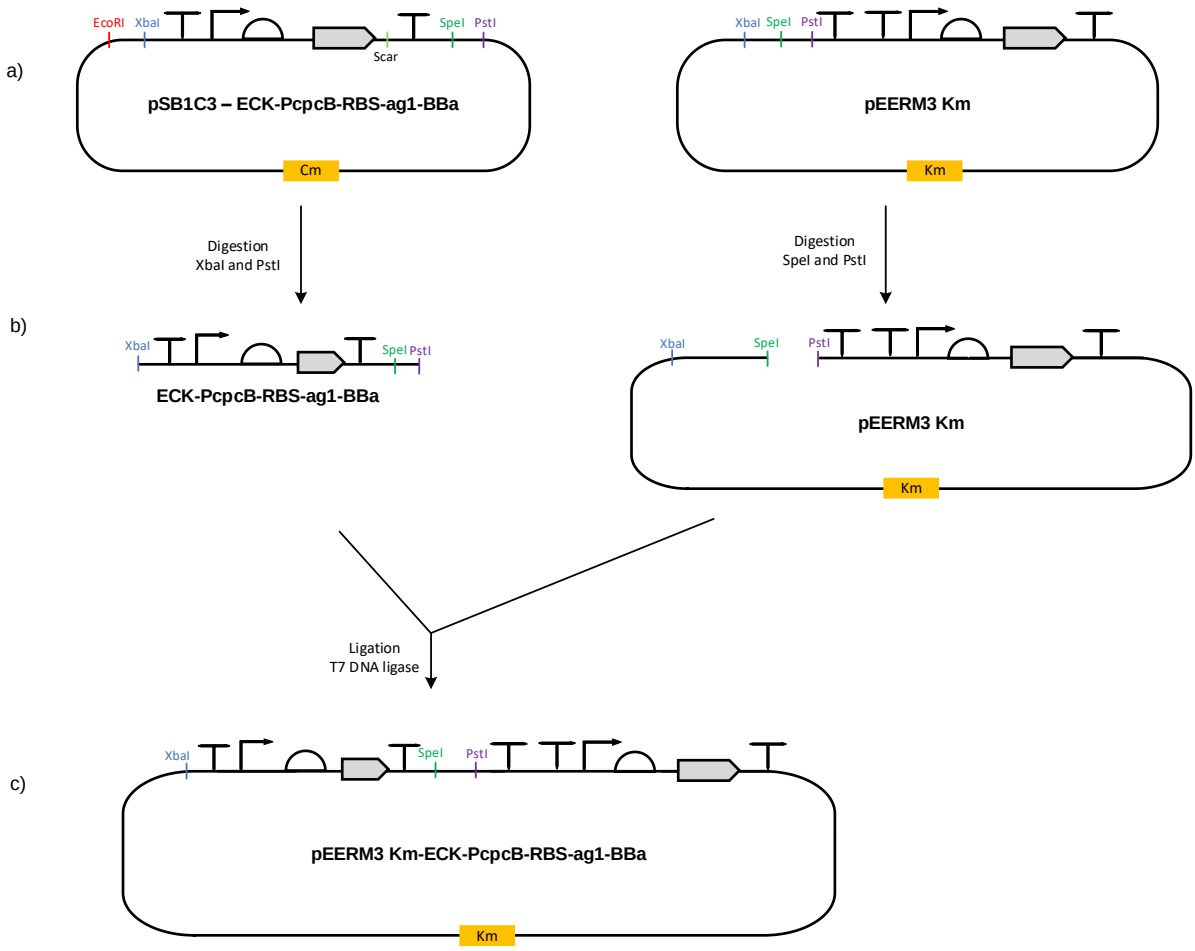


Fig. 4.38 Schematic representation of the assembly of parts. a) BioBrick containing constructed parts from Fig. 4.33c and plasmid backbone (pEERM3 Km). b) Digestion process of each BioBrick with digestion enzymes. c) Ligation of complementary DNA overhangs to construct a BioBrick containing both parts from a).

Unlike pEERM3 Km, pEERM4 Cm plasmid contains all restriction enzyme sites that enable to carry out a digestion and ligation similar to the performed in section 4.7. In pEERM3 Km, the final result is an integration of all parts inside the plasmid backbone (Fig. 4.38c) with similarities to Fig. 4.37b.

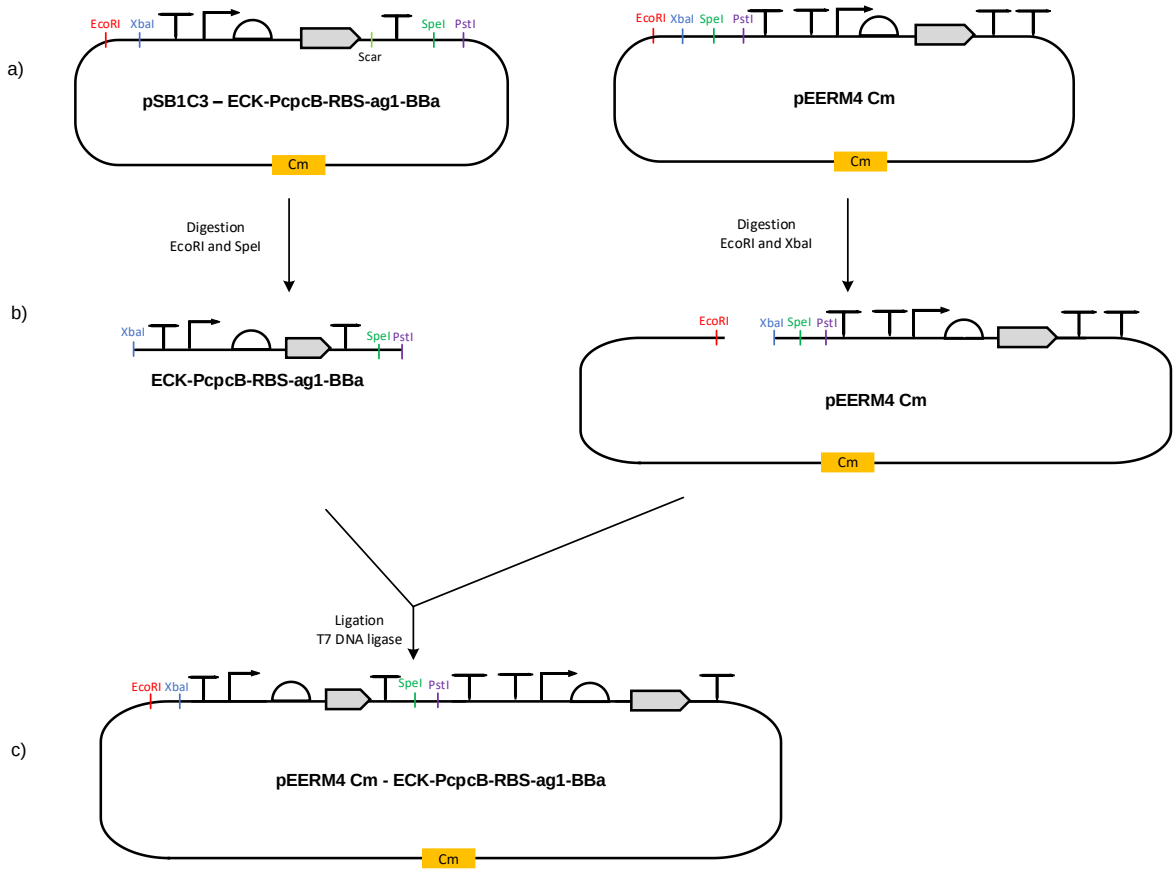
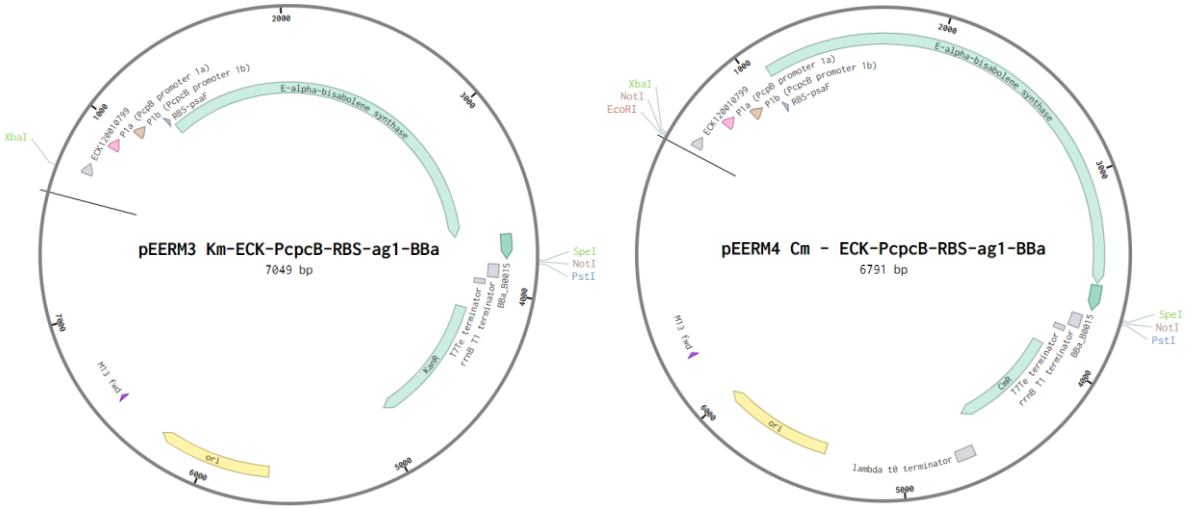


Fig. 4.39 Schematic representation of the assembly of parts. a) BioBrick containing constructed parts from Fig. 4.33c and plasmid backbone (pEERM4 Cm). b) Digestion process of each BioBrick with digestion enzymes. c) Ligation of complementary DNA overhangs to construct a BioBrick containing both parts from a).

Fig. 4.38c and Fig. 4.39c are generated in Benchling to represent both plasmid vectors containing the parts that are capable of driving the expression of bisabolene (Fig. 4.40). Furthermore, the final constructs can be integrated into *Synechocystis* because they are using backbones (pEERM plasmid vectors) that were designed to be integrated into the chromosomal DNA of *Synechocistis*.



a) b)

Fig. 4.40 a) Construct containing elements to produce bisabolene inside a plasmid backbone (pEERM3 Km). b) Construct containing elements to produce bisabolene inside a plasmid backbone (pEERM4 Cm).

Constructs from Fig. 4.40 uses an integrative plasmid (pEERM) as backbone, which is capable of being integrated into the chromosomal DNA of *Synechocystis* during the transformation process. Integration is carried out in the neutral sites, which are considered neutral because no transcription is performed in them, of *Synechocystis*. However, in chromosomal DNA of *Synechocystis*, fifteen neutral sites have been identified wherein heterologous genes can be integrated. (Ng et al., 2015).

Conversely, pEERM plasmids were designed to be integrated in two specific neutral sites in *Synechocystis*. pEERM3 Km plasmid is integrated in the slr0168 gene of *Synechocystis*, whereas pEERM4 Cm plasmid is integrated in the neutral site between slr2030 and slr2031 gene (Englund et al., 2015). The recognition of specific neutral sites for integrative plasmids lead to an integration of the construct, from vector plasmid, to chromosomal DNA of *Synechocystis* by driving bisabolene gene expression.

5. DISCUSSION AND ANALYSIS

In this section, the results obtained are discussed to interpret its functionality and importance during the construction of a BioBrick in order to drive bisabolene expression in a host organism as *Synechocystis*.

5.1. Selection of bioparts

5.1.1. Protein coding sequence

The selection of parts to construct a synthetic biology device is important because those parts are the elements that drive the gene expression to produce bisabolene terpene as main product. In the case of the bisabolene synthase protein coding sequence (ag1 gene), this gene has been widely studied by Bohlmann et al. (1998) to determine its importance in bisabolene production. They sequenced and reported a complementary DNA (cDNA) of ag1 gene in GenBank capable of being replicated in prokaryotic cells. The cDNA has been used in different studies to biosynthesise both limonene and bisabolene in *Synechococus sp. PCC 7002* (Davies et al., 2014), bisabolene in *E. coli* (Peralta-Yahya et al., 2011), and bisabolene in *Synechocystis* (Sebesta & Peebles, 2020).

Based on previous studies, ag1 gene was selected and engineered by using a gene-editing software as Benchling. Upon sequence analysis of ag1 gene, it showed different restriction enzyme sites (Fig. 4.1) that could fragment the gene and lead to a loss of transcription, and a codon optimization was required. In comparison with the codon optimisation performed by Sebesta and Peebles (2020), the optimization, in this study, was performed by using a codon usage shown in table 4.1 and Benchling.

The codon optimisation of ag1 gene differs in its sequence from other studies, and it might be because the method used in this study is manually due to *Synechocystis* codon usage is not included in the codon optimisation algorithm of Benchling. However, the codon optimisation is performed and represented in Fig. 4.2 in which all restriction enzyme sites are deleted, changing their nucleotides as part of the optimisation process.

5.1.2. Promoters, RBS and terminators

In the case of promoters in *Synechocystis*, those promoters were grouped according to their nature (Fig. 4.3). This classification helps to identify which promoters are considered either strong or weak to drive gene expression. Indeed, Heirdorn et al. (2011) attempted to group

promoters in cyanobacteria, but their attempt was not completed. Similar work was reported by Till et al. (2020), who observed that native promoters had more capability to express a gene of interest than engineered ones.

By taking advantage of the flowchart from Fig. 4.3, an evaluation of constitutive and inducible native promoters was achieved to find a promoter capable of overexpressing the ag1 gene. Inducible promoters as P_{nrsB} (Englund et al., 2016), P_{petE} (Briggs et al., 1990), and P_{nirA} (Heirdorn et al., 2011) have been widely studied. P_{nrsB} and P_{petE} are more outstanding because they can overexpress a heterologous gene to produce terpenes although they depend on an inducer to enhance the transcription process. The use of inducers such as nickel ions (P_{nrsB}) or copper ions (P_{petE}) could lead to a dependence on those ions to obtain a high-level of expression of terpenes that could contain those ions in bisabolene product.

Unlike inducible promoters, strong constitutive promoters such as P_{psbA2} (Englund et al., 2016) and P_{cpc} (Davies et al., 2014; Liu & Pakrasi, 2018; Sebesta & Peebles, 2020; Zhou et al., 2014) do not depend upon and inducer to initiate the transcription process but also on the strength of the promoter to bear both light and CO_2 conditions to initiate the transcription. P_{psbA2} and its engineered promoters (P_{psbA2S}, P_{psbA2M}, and P_{psbA2L}) were described by Englund et al. (2016) to demonstrate that the shortest the sequence of P_{sabA2} and its engineered promoters were, the highest expression of a gene was obtained.

However, a comparison of this promoter with P_{cpcB} promoters was made by Liu and Pakrasi (2018). They found that a measurement of fluorescence of both promoters, P_{cpcB} resulted in 80-fold higher than P_{psbA2}. Furthermore, P_{cpcB} promoter was used to design a stronger promoter named P_{cpc560} (Zhou et al., 2014), but Till et al. (2020) reported that despite the strength of P_{cpc560} promoter, its results were inconsistent to be considered as the strongest promoter.

Despite P_{cpcB} promoter was obtained from *Synechococcus elongatus PCC 7942*, this promoter was capable of being initiating gene expression in *Synechococcus sp. PCC 7002* to produce bisabolene (Sebesta & Peebles, 2020), demonstrating its efficacy to be assembled in other microorganisms as *Synechocystis sp. PCC 6803*. Besides, P_{cpcB} promoter contains two different transcription start sites (TSS), as it was depicted in Fig. 4.5, which enable this promoter to be strong enough to be inserted in *Synechocystis* to produce bisabolene. However, the strength needs to be tested in laboratory to determine the activity of the promoter during the production of bisabolene.

Likewise, the selection of RBS was based on its SD sequence because, according to other authors (Heirdorn et al., 2011; Wang et al., 2012), SD sequence is capable of enhancing translation efficiency due to the presence of AG region. Besides, engineered RBSs were

studied (Heirdorn et al., 2011), and their efficacy for translating a gene is lower than natural RBSs, such RBS efficiency was corroborated by Wang et al. (2012) in which they indicated that a SD sequence could improve translation. A natural RBS, RBS-psaF, was selected (Fig. 4.6) owing to its both high-level expression and SD sequence (Liu & Pakrasi, 2018). Furthermore, RBS-psaF has a well-defined AG region (table 4.2) that can be beneficial to encode a heterologous gene in comparison with engineered RBSs.

During the development of this study, two terminators were selected in comparison with other studies in which only one terminator was selected. The most studied terminator has been BBa_B0015 (Fig. 4.7), this terminator was widely reported. Heirdorn et al. (2011) used this terminator in their device design to determine in *Synechocystis* the translation efficiency of a fluorescent protein. Huang and Lindblad (2013) interfaced this terminator with several engineered promoters to evaluate their performance. Englund et al. (2016) used this terminator in their device design to measure the strength of promoters and RBSs, and among other studies.

The majority of those studies, including this study, have used BBa_B0015 as the main terminator for gene expression, however, a second terminator (Fig. 4.8) capable of insulating the final construct must be included the upstream of the promoter. This enhancement was considered by Kelly et al. (2019) in order to insulate and improve construct transcription in *Synechocystis*. Unlike previous works, in this study, both the strong terminator to insulate a construct (Fig. 4.8) and a strong terminator to enhance level expression of a heterologous gene are included.

5.2. Construction of BioBrick

In this study, a BioBrick Standard format was chosen to generate inserts inside a plasmid vector because this BioBrick Standard enables to flank with a well-characterised restriction enzyme sites, as known as BioBrick prefix and suffix, inside a standard BioBrick vector as pSB1C3 plasmid. According to Ho-Shing et al. (2012), advantages of BioBrick standards are:

- The use of standard elements to flank selected parts or DNA fragments, as known as restriction enzyme sites (EcoRI, XbaI, SpeI, and PstI).
- Constructs are registered in an online database as International Genetically Engineered Machines (iGEM).
- BioBricks are interchangeable and compatible with the Standard Assembly Method.

By following this Standard Assembly format, selected parts (promoter, RBS, coding sequence, and terminators) were flanked by restriction enzyme sites. As it was depicted from Fig. 4.12 to Fig. 4.16, the upstream of the parts was flanked by EcoRI and XbaI whereas the downstream of the parts was flanked by SpeI and PstI.

Unlike flanking parts with BioBrick prefix and suffix, the plasmid vector pSB1C3 does not require to be flanked owing to pSB1C3 contains those restriction enzyme sites (Fig. 4.17) wherein flanked parts can be inserted. Indeed, constructing a BioBrick requires to perform both a digestion and a ligation of flanked parts and plasmid vector.

Digestion was performed in *in silico* by using EcoRI and PstI enzymes to generate an insert in the parts and a receptor in the plasmid vector. *In silico* digestion enabled to generate complementary DNA overhangs (Fig. 4.19) capable of sealing the cut sites when ligation was performed. Fig. 4.20 and Fig. 4.21 were constructed by following the digestion and ligation sequence in *in silico* by using Benchling programme. However, those BioBricks needed to be together to execute any gene expression because, in the way they were constructed, they were unable to encode genes.

The results of creating BioBricks, shown in Fig. 4.20 and Fig. 4.21, indicated that those parts needed to be assembled inside a plasmid vector. However, before assembling those BioBricks, a schematic representation of the final construct was made by using SBOL diagram, in which all parts and the final construct were represented in a standard language (Fig. 4.27). This schematic representation enabled to establish a sequence of ligation and digestion to follow in Benchling to obtain a final construct as it was depicted in Fig. 4.34. Indeed, this construct could be capable of encoding a gene sequence, but due to the backbone plasmid was not an integrative vector, it would not work in the chromosomal DNA of *Synechocystis*.

The construct obtained in Fig. 4.34 was digested and ligated inside pEERM plasmid vectors, integrative plasmids that were designed by Englund et al. (2015) to be integrated in *Synechocystis* DNA. A result of that digestion and ligation were two BioBricks containing a sequence to encode bisabolene gene (Fig. 4.40) similar to the construct obtained from Englund et al. (2015). However, the difference between BioBricks from Fig. 4.40 and the construct reported was that Fig. 4.40 contained a terminator in the upstream of promoter to insulate the whole construct that could enhance the encoding sequence of bisabolene gene (ag1 gene), whereas in Englund et al. (2015) construct did not have a terminator to insulate the construct. The position of the terminator to insulate a construct was described by Kelly et al. (2019), who indicated that insulating a construct with a terminator before the promoter could lead to high titre values of a desired product thus in Fig. 4.40 such terminator was included during BioBrick construction.

A high-level of expression can be obtained owing to pEERM plasmid vectors were engineered to improve the MEP pathway. pEERM3 Km is engineered to delete both the native promoter P_{psbA2} and squalene synthase enzyme to avoid having a competition for substrates during the expression of a heterologous gene. Unlike pEERM3 Km, pEERM4 Cm is engineered to enhance dxs enzyme which is the first enzyme of the metabolic route of *Synechocystis*.

Furthermore, Fig. 4.40 was created in order to be integrated inside neutral sites of *Synechocystis*. In the case of pEERM3 Km, it could be integrated in neutral site located inside the *slr0168* gene, whereas pEERM4 Cm could be integrated in neutral sites located between two genes such as *slr2030* and *slr2031*. Those neutral sites, similar to the others studied by Ng et al. (2015), are places in the genomic sequence that do not cause a mutation of the host organism during the integration of a heterologous gene.

The integration of the Fig. 4.40 in those neutral sites could be performed by using the natural transformation which is a well-known technique in this strain of cyanobacteria (Zang et al., 2007; Heirdorn et al., 2011; Wang et al., 2012; Davies et al., 2014; Englund et al., 2015; Sebesta & Peebles, 2020). Indeed, a comparison of natural transformation in *Synechocystis* with electroporation or ultrasonic transformation, natural transformation has shown to be more efficient than the other two techniques (Zang et al., 2007).

5.3. Yield expectation

To date, bisabolene production inside of cyanobacterium strains has been studied by Davies et al. (2014), Dienst et al. (2020), and Sebesta and Peebles (2020). In the case of Davies et al. (2014), they studied bisabolene production by using *Synechococcus sp. PCC 7002*, a cyanobacterium strain, as a host organism. They used both a strong native promoter (P_{cpcB}) and the ag1 gene to biosynthesise limonene and bisabolene as final products. Besides, they reported that after 96 hours of cultivation in a photoreactor, the yield of bisabolene reached 0.6 mg/L.

Unlike the previous study, Sebesta and Peebles (2020) decided to biosynthesise bisabolene with different elements. They used an engineered weak promoter such as P_{trc2O}, an optimization of ag1 gene, and an engineered RBS. When those elements were inserted inside of *Synechocystis*, it was cultivated in a photobioreactor for 36 days. At that time, bisabolene titre values reached 22.2 mg/L. However, the disadvantage of this engineered promoter is that P_{trc2O} promoter has two lac operators that depend on Isopropyl β-D-1-thiogalactopyranoside (IPTG) to relieve repression. Therefore, this promoter can be repressed

by a lac repressor protein (LacI) when IPTG is not present by producing a low-level expression of proteins and, in some cases, a metabolic burden (Till et al., 2020).

A different approach was reported by Peralta-Yahya et al. (2011), who reported bisabolene production by using a different host organisms such as *E. coli* and *Saccharomyces cerevisiae*. They obtained titres values of 900 mg/L by using an engineered promoter (P_{trc}) which was induced by lactose.

In this study, *Synechocystis* was chosen as a host microorganism because it was capable of biosynthesising its glucose molecules by taking light and CO_2. Unlike Synechocystis, *E. coli* and S. cerevisiae are not autotrophic organisms because it depends on organic nutrients to generate terpenoid metabolites. Furthermore, neither E. coli nor S. cerevisiae contain the MEP pathway which can lead to a metabolic burden when this pathway is engineered in those organisms. Conversely, *Synechocystis* has the MEP pathway that is quite similar to plant chloroplasts and, besides that, biochemical synthesis of *Synechocystis* has similarities to plants, this feature makes Synechocystis an ideal host to both express plant genes as terpenes and produce a sustainable biofuel.

Comparing promoters from previous studies, in this study, P_{cpcB} – a strong constitutive promoter - was used to avoid the dependence on IPTG during the growth in a photobioreactor and the metabolic burden of the host. Likewise, ag1 gene was optimised to be inserted in *Synechocystis* by using Benchling. In the case of RBS, constitutive RBS (RBS-psaF) was selected because it was reported by Liu and Pakrasi (2018) that this RBS could lead to a high yield of a heterologous gene. Indeed, the addition of two terminators in the construct could enhance gene expression in order to increase bisabolene production with titre values higher than 0.6 mg/L reported by Davies et al. (2014), but similar to 22.2 mg/L (Sebesta & Peebles, 2020)

6. CONCLUSION AND RECOMMENDATIONS

6.1. Conclusions

This study was aimed at designing in *in silico* a construct capable of being replicated inside a host organism as *Synechocystis* in order to produce bisabolene. The potential of *Synechocystis* as a bisabolene producer and their elements have been evaluated. Conclusion can be summarised as follow:

- Benchling is an open software that enables to optimise, digest, and ligate DNA sequences or elements by saving cost and time when those process are performed in laboratory.
- Bisabolene synthase coding sequence is, to date, the most studied enzyme to produce bisabolene, and it is obtained from the conifer plant *Abies grandies*.
- Native promoter, P_{cpcB}, is, compared with other promoter, capable of enhancing bisabolene synthase expression owing to the strength to translate a heterologous.
- Including a terminator, ECK120010799, in the upstream of the construct could lead to an insulation of the construct of this study. To note, this insulation has never been tested in different studies to produce a terpenoid. However, the inclusion of ECK120010799 terminator could improve the translation in *Synechocystis*.
- The construction *in silico* of a construct inside pEERM plasmids could enhance bisabolene expression because this new plasmid vector contains elements capable of expressing bisabolene gene when these plasmids are integrated inside of chromosomal DNA of *Synechocystis*. For instance, the construct inside of pEERM3 Km can delete genes to avoid a competition for substrates to produce more IPP molecules for producing bisabolene, whereas in the construct in pEERM4 Cm can lead to an overproduction of dxs enzyme which is essential for the MEP pathway.

6.2. Future works

A lab work is needed to be performed in order to evaluate the strength of the selected elements such as promoter, RBS, and terminators. Furthermore, an evaluation of the codon optimization needs to be examined to determine whether there are gene mutations or not.

A recombination of the constructed vector inside *Synechocystis* must be performed. Furthermore, a cultivation of *Synechocystis* in photobioreactor should be considered to determine both the yield and time for producing bisabolene.

REFERENCES

Akhtar, J. & Amin, N., (2011). A review on process conditions for optimum bio-oil yield in hydrothermal liquefaction of biomass. *Renewable and Sustainable Energy Reviews* [online]. **15**(3), 1615-1624. [Viewed 27 March 2020]. Available from: DOI: 10.1016/j.rser.2010.11.054.

Beal, J., Nguyen, T., Gorochowski, E., Goñi-Moreno, A., Scott-Brown, J. & Al, E., (2019). Communicating Structure and Function in Synthetic Biology Diagrams. *ACS Synthetic Biology* [online]. **8**(8), 1818-1825. [Viewed 10 July 2020]. Available from: DOI: 10.1021/acssynbio.9b00139.

Bohlmann, J., Crock, J., Jetter, R. & Croteau, R., (1998). Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-a-bisabolene synthase from grand fir (Abies grandis). *Proc Natl Acad Sci U S A* [online]. **95**(12), 6756-6761. [Viewed 08 July 2020]. Available from: DOI: 10.1073/pnas.95.12.6756.

Briggs, L. M., Pecoraro, V. L. & Mcintosh, L., (1990). Copper-induced expression, cloning, and regulatory studies of the plastocyanin gene from the cyanobacterium Synechocystis sp. PCC 6803. *Plant Molecular Biology* [online]. **15**(1990), 633-642. [Viewed 13 July 2020]. Available from: DOI: 10.1007/BF00017837.

Davies, F. K., Work, V. H., Beliaev, A. S. & Posewitz, M. C., (2014). Engineering limonene and bisabolene production in wild type and a glycogen-deficient mutant of Synechococcus sp. PCC 7002. *Frontiers in Bioengineering and Biotechnology* [online]. **2**(21). [Viewed 01 July 2020]. Available from: DOI: 10.3389/fbioe.2014.00021.

Dexter, J. & Fu, P., (2009). Metabolic engineering of cyanobacteria for ethanol production. *Energy & Environmental Science* [online]. **2**(2009). [Viewed 857-864]. Available from: DOI: 10.1039/B811937F.

Dienst, D., Wichmann, J., Mantovani, O., Rodrigues, J. S. & Linberg, P., (2020). High density cultivation for efficient sesquiterpenoid biosynthesis in Synechocystis sp. PCC 6803. *Scientific Reports* [online]. **10**(5932). [Viewed 01 August 2020]. Available from: DOI: 10.1038/s41598-020-62681-w.

Englund, E., Andersen-Ranber, J., Miao, R., Hamberger, B. & Linberg, P., (2015). Metabolic Engineering of Synechocystis sp. PCC 6803 for Production of the Plant Diterpenoid Manoyl Oxide. *ACS Synthetic Biology* [online]. **4**(12), 1270-1278. [Viewed 27 March 2020]. Available from: DOI: 10.1021/acssynbio.5b00070.

Englund, E., Liang, F. & Lindberg, P., (2016). Evaluation of promoters and ribosome binding sites for biotechnological applications in the unicellular cyanobacterium Synechocystis sp. PCC 6803. *Scientific Reports* [online]. **6**(36640). [Viewed 02 June 2020]. Available from: DOI: 10.1038/srep36640.

Englund, E., Pattaniak, B., Ubhayasekera, S. J. K., Stensjö, K., Bergquist, J. & Linberg, P., (2014). Production of Squalene in Synechocystis sp. PCC 6803. *PLOS ONE* [online]. **9**(3), e90270. [Viewed 27 June 2020]. Available from: DOI: 10.1371/journal.pone.0090270.

Heirdorn, T., Camsund, D., Huang, H., Linberg, P., Oliveira, P., Stensjö, K. & Lindblad, P., (2011). Synthetic Biology in Cyanobacteria: Engineering and Analyzing Novel Functions. *Methods in Enzymology* [online]. **497**(2011), 539-579. [Viewed 27 March 2020]. Available from: DOI: 10.1016/B978-0-12-385075-1.00024-X.

Ho-Shing, O., Lau, K. H., Vernon, W., Eckdahl, T. T. & Campbell, A. M., (2012). Assembly of Standardized DNA Parts Using BioBrick Ends in E. coli. I*n:* Peccoud, J., ed. *Gene Synthesis*. New York: Springer Science+Business Media. pp. 61-76. [Viewed 11 July 2020]. Available from: 10.1007/978-1-61779-564-0_6.

Huang, H. & Lindblad, P., (2013). Wide-dynamic-range promoters engineered for cyanobacteria. *Journal of Biological Engineering* [online]. **7**(10), 1-11. [Viewed 20 March 2020]. Available from: DOI: 10.1186/1754-1611-7-10.

Kelly, C. L., Taylor, G. M., Šatkuté, A., Dekker, L. & Heap, J. T., (2019). Transcriptional Terminators Allow Leak-Free Chromosomal Integration of Genetic Constructs in Cyanobacteria. *Microorganisms* [online]. **7**(8). [Viewed 15 July 2020]. Available from: DOI: 10.3390/microorganisms7080263.

Lackner, M., (2017). Feedstocks for Biofuels. I*n:* Chen, W. Y., Suzuki, T. & Lackner, M., eds. *Handbook of Climate Change Mitigation and Adaptation*. Springer International Publishing. pp. 1-32. [Viewed 06 April 2020]. Available from: 10.1007/978-3-319-14409-2_90.

Liere, K. & Maliga, P., (2001). Plastid RNA Polymerases in Higher Plants. I*n:* Aro, E. M. & Andersson, B., eds. *Regulation of Photosynthesis*. Dordrecht: Springer. pp. 29-49. [Viewed 15 July 2020]. Available from: 10.1007/0-306-48148-0_2.

Lindberg, P., Park, S. & Melis, A., (2010). Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. *Metabolic Engineering* [online]. **12**(1). [Viewed 15 May 2020]. Available from: DOI: 10.1016/j.ymben.2009.10.001.

Liu, D. & Pakrasi, H. B., (2018). Exploring native genetic elements as plug-in tools for synthetic biology in the cyanobacterium Synechocystis sp. PCC 6803. *Microbial Cell Factories* [online]. **17**(48). [Viewed 17 April 2020]. Available from: DOI: 10.1186/s12934-018-0897-8.

Marshall, A. M., (2010). First generation biofuels compete. *New Biotechnology* [online]. **27**(5), 596-608. [Viewed 28 March 2020]. Available from: DOI: 10.1016/j.nbt.2010.06.010.

Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D., (2003). Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology* [online]. **21**(7), 796-802. [Viewed 01 June 2020]. Available from: DOI: 10.1038/nbt833.

Ng, A. H., Berla, B. M. & Pakrasi, H. B., (2015). Fine-Tuning of Photoautotrophic Protein Production by Combining Promoters and Neutral Sites in the Cyanobacterium Synechocystis sp. Strain PCC 6803. *American Society for Microbiology* [online]. **81**(2015), 6857-6863. [Viewed 05 August 2020]. Available from: DOI: 10.1128/AEM.01349-15.

Pattaniak, B. & Lindberg, P., (2015). Terpenoids and Their Biosynthesis in Cyanobacteria. *Life* [online]. **5**(1), 269-293. [Viewed 15 April 2020]. Available from: DOI: 10.3390/life5010269.

Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D. & Lee, T. S., (2011). Identification and microbial production of a terpene-based advanced biofuel. *Nature Communications* [online]. **2**(483). [Viewed 20 May 2020]. Available from: DOI: 10.1038/ncomms1494.

Picazo-Espinosa, R., Gonzalez-Lopez, J. & Manzanera, M., (2011). Bioresources for ThirdGeneration Biofuels. I*n:* Dos Santos, M., ed. *Biofuel's Engineering Process Technology*. InTechOpen. pp. 115-140. [Viewed 28 March 2020]. Available from: 10.5772/17134.

Reinsvold, R. E., Jinkerson, R. E., Radakovits, R. & Posewitz, M. C., (2011). The production of the sesquiterpene -caryophyllene in a transgenic strain of the cyanobacterium Synechocystis. *Journal of Plant Physiology* [online]. **168**(8). [Viewed 27 July 2020]. Available from: DOI: 10.1016/j.jplph.2010.11.006.

Saladini, F., Patrizi, N., Pulselli, F. M., Marchettini, N. & Bastianoni, S., (2016). Guidelines for emergy evaluation of first, second and third generation biofuels. *Renewable and Sustainable Energy Reviews* [online]. **66**(1), 221-227. [Viewed 28 March 2020]. Available from: DOI: 10.1016/j.rser.2016.07.073.

Sebesta, J. & Peebles, C. a. M., (2020). Improving heterologous protein expression in Synechocystis sp. PCC 6803 for alpha-bisabolene production. *Metabolic Engineering Communications* [online]. **10**(e00117). [Viewed 13 April 2020]. Available from: DOI: 10.1016/j.mec.2019.e00117.

Sengupta, A., Madhu, S. & Wangikar, P. P., (2020). A Library of Tunable, Portable, and Inducer-Free Promoters Derived from Cyanobacteria. *ACS Synthetic Biology* [online]. **9**(7). [Viewed 05 August 2020]. Available from: DOI: 10.1021/acssynbio.0c00152.

Till, P., Toepel, J., Bühler, B., Mach, R. L. & Mach-Aigner, A. R., (2020). Regulatory systems for gene expression control in cyanobacteria. *Applied Microbiology and Biotechnology* [online]. **104**(5), 1977-1991. [Viewed 10 Junio 2020]. Available from: DOI: 10.1007/s00253-019-10344-w.

Tiwari, A., (2018). Cyanobacteria: The Wonderful Factories. I*n:* Dos Santos, M., ed. *Cyanobacteria*. InTechOpen. pp. 3-7. [Viewed 08 April 2020]. Available from: 10.5772/intechopen.79751.

Wang, B., Wang, J., Zhang, W. & Meldrum, D. R., (2012). Application of synthetic biology in cyanobacteria and algae. *Frontiers in Microbiology* [online]. **3**(344), 1-15. [Viewed 01 April 2020]. Available from: DOI: 10.3389/fmicb.2012.00344.

Wang, C., Liwei, M., Park, J. B., Jeong, S. H., Wei, G., Wang, Y. & Kim, S. W., (2018). Microbial Platform for Terpenoid Production: Escherichia coli and Yeast. *Frontiers in Microbiology* [online]. **9**(2460). [Viewed Viewed 01 May 2020]. Available from: DOI: doi.org/10.3389/fmicb.2018.02460.

Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y. & Zhang, F., (2013). Development of Synechocystis sp. PCC 6803 as a Phototrophic Cell Factory. *Marine Drugs* [online]. **11**(2013). [Viewed 01 August 2020]. Available from: DOI: 10.3390/md11082894.

Zang, X., Liu, B., Liu, S., Arunakumara, K. & Zhang, X., (2007). Optimum Conditions for Transformation of Synechocystis sp. PCC 6803. *The Journal of Microbiology* [online]. **45**(3). [Viewed 23 July 2020]. Available from: DOI: PMID: 17618230.

Zheljazkov, V. D., (2012). Distillation Time Changes Oregano Essential Oil Yields and Composition but Not the Antioxidant or Antimicrobial Activities. *Hort Science* [online]. **47**(6), 777-784. [Viewed 15 July 2020]. Available from: DOI: 10.21273/HORTSCI.47.6.777.

Zhou, J., Zhang, H., Meng, H., Zhu, Y., Bao, G., Zhang, Y., Li, Y. & Ma, Y., (2014). Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria. *Scientific Reports* [online]. **4**(4500). [Viewed 15 April 2020]. Available from: DOI: doi.org/10.1038/srep04500.

APPENDICES

A.1 Coding sequence of bisabolene synthase (ag1) without optimisation

The following coding sequence was obtained from GenBank with accession number AF006195.1

ATGGCTGGCGTTTCTGCTGTATCAAAGGTTTCCAGCTTGGTTTGTGATTTGTCGAGTACCAGCGGCTTGATT
CGAAGAACTGCCAATCCTCATCCCAATGTCTGGGGTTATGATCTTGTGCATTCTCTTAAATCACCTTATATTG
ATTCTAGTTACAGAGAACGCGCGGAGGTCCTTGTTAGCGAGATTAAAGCGATGCTTAATCCAGCTATTACAG
GAGATGGAGAATCAATGATTACTCCATCTGCTTATGACACAGCATGGGTAGCGAGGGTGCCCGCCATTGAT
GGCTCTGCTCGCCCGCAATTTCCCCAAACAGTTGACTGGATTTTGAAAAACCAGTTAAAAGATGGTTCATGG
GGCATTCAGTCCCACTTTCTGCTGTCCGACCGTCTTCTTGCCACTCTTTCTTGTGTTCTTGTGCTCCTTAAAT
GGAACGTTGGGGATCTGCAAGTAGAGCAGGGAATTGAATTCATAAAGAGCAATCTGGAACTAGTAAAGGAT
GAAACCGATCAAGATAGCTTGGTAACAGACTTTGAGATCATATTTCCTTCTCTGTTAAGAGAAGCTCAATCTC
TGCGCCTCGGACTTCCCTACGACCTGCCTTATATACATCTGTTGCAGACTAAACGGCAGGAAAGATTAGCAA
AACTTTCAAGGGAGGAAATTTATGCGGTTCCGTCGCCATTGTTGTATTCTTTAGAGGGAATACAAGATATAGT
TGAATGGGAACGAATAATGGAAGTTCAAAGTCAGGATGGGTCTTTCTTAAGCTCACCTGCTTCTACTGCCTG
CGTTTTCATGCACACAGGAGACGCGAAATGCCTTGAATTCTTGAACAGTGTGATGATCAAGTTTGGAAATTTT
GTTCCCTGCCTGTATCCTGTGGATCTGCTGGAACGCCTGTTGATCGTAGATAATATTGTACGCCTTGGAATC
TATAGACACTTTGAAAAGGAAATCAAGGAAGCTCTTGATTATGTTTACAGGCATTGGAACGAAAGAGGAATTG
GGTGGGGCAGACTAAATCCCATAGCAGATCTTGAGACCACTGCTTTGGGATTTCGATTGCTTCGGCTGCATA
GGTACAATGTATCTCCAGCCATTTTTGACAACTTCAAAGATGCCAATGGGAAATTCATTTGCTCGACCGGTCA
ATTCAACAAAGATGTAGCAAGCATGCTGAATCTTTATAGAGCTTCCCAGCTCGCATTTCCCGGAGAAAACATT
CTTGATGAAGCTAAAAGCTTCGCTACTAAATATTTGAGAGAAGCTCTTGAGAAAAGTGAGACTTCCAGTGCAT
GGAACAACAAACAAAACCTGAGCCAAGAGATCAAATACGCGCTGAAGACTTCTTGGCATGCCAGTGTTCCGA
GAGTGGAAGCAAAGAGATACTGTCAAGTGTATCGCCCAGATTATGCACGCATAGCAAAATGCGTTTACAAGC
TACCCTACGTGAACAATGAAAAGTTTTTAGAGCTGGGAAAATTAGATTTCAACATTATCCAGTCCATCCACCA
AGAAGAAATGAAGAATGTTACCAGCTGGTTTAGAGATTCGGGGTTGCCACTATTCACCTTCGCTCGGGAGAG
GCCGCTGGAATTCTACTTCTTAGTAGCGGCGGGGACCTATGAACCCCAGTATGCCAAATGCAGGTTCCTCTT
TACAAAAGTGGCATGCTTGCAGACTGTTCTGGACGATATGTATGACACTTATGGAACCCTAGATGAATTGAA
GCTATTCACTGAGGCTGTGAGAAGATGGGACCTCTCCTTTACAGAAAACCTTCCAGACTATATGAAACTATGT
TACCAAATCTATTATGACATAGTTCACGAGGTGGCTTGGGAGGCAGAGAAGGAACAGGGGCGTGAATTGGT
CAGCTTTTTCAGAAAGGGATGGGAGGATTATCTTCTGGGTTATTATGAAGAAGCTGAATGGTTAGCTGCTGA
GTATGTGCCTACCTTGGACGAGTACATAAAGAATGGAATCACATCTATCGGCCAACGTATACTTCTGTTGAGT
GGAGTGTTGATAATGGATGGGCAACTCCTTTCGCAAGAGGCATTAGAGAAAGTAGATTATCCAGGAAGACGT
GTTCTCACAGAGCTGAATAGCCTCATTTCCCGCCTGGCGGATGACACGAAGACATATAAAGCTGAGAAGGC
TCGTGGAGAATTGGCGTCCAGCATTGAATGTTACATGAAAGACCATCCTGAATGTACAGAGGAAGAGGCTCT
CGATCACATCTATAGCATTCTGGAGCCGGCGGTGAAGGAACTGACAAGAGAGTTTCTGAAGCCCGACGACG
TCCCATTCGCCTGCAAGAAGATGCTTTTCGAGGAGACAAGAGTGACGATGGTGATATTCAAGGATGGAGAT
GGATTCGGTGTTTCCAAATTAGAAGTCAAAGATCATATCAAAGAGTGTCTCATTGAACCGCTGCCACTGTAA

A.2 Optimisation of bisabolene synthase (ag1) in Benchling

Red colour indicates that this nucleotide was optimised in Benchling by using the information from table 4.1. The underlined sequence represents methionine amino acid sequence which is not optimised.

ATGGCCGGCGTGTCCGCCGTGTCCAAAGTGTCCTCCTTGGTGTGTGATTTGTCCTCCACCTCCGGCTTGAT
TCGGCGGACCGCCAATCCCCATCCCAATGTGTGGGGCTATGATTTGGTGCATTCCTTGAAATCCCCTATAT
TGATTCCTCCTATCGGGAACGGGCCGAAGTGTTGGTGTCCGAAATTAAAGCCATGTTGAATCCCGCCATTAC
CGGCGATGGCGAATCCATGATTACCCCCTCCGCCTATGATACCGCCTGGGTGGCCCGGGTGCCCGCCATT
GATGGCTCCGCCCGGCCCCAATTTCCCCAAACCGTGGATTGGATTTTGAAAAATCAATTGAAAGATGGCTCC
TGGGGCATTCAATCCCATTTTTTGTTGTCCGATCGGTTGTTGGCCACCTTGTCCTGTGTGTTGGTGTTGTTGA
AATGGAATGTGGGCGATTTGCAAGTGGAACAAGGCATTGAATTTATTAAATCCAATTTGGAATTGGTGAAAGA
TGAAACCGATCAAGATTCCTTGGTGACCGATTTTGAAATTATTTTCCCTCCTTGTTGCGGGAAGCCCAATCC
TTGCGGTTGGGCTTGCCCTATGATTTGCCCTATATTCATTTGTTGCAAACCAAACGGCAAGAACGGTTGGCC
AAATTGTCCCGGGAAGAAATTTATGCCGTGCCCTCCCCCTTGTTGTATTCCTTGGAAGGCATTCAAGATATTG
TGGAATGGGAACGGATTATGGAAGTGCAATCCCAAGATGGCTCCTTTTTGTCCTCCCCCGCCTCCACCGCC
TGTGTGTTTATGCATACCGGCGATGCCAAATGTTTGGAATTTTTGAATAGTGTGATGATTAAATTTGGCAATTT
TGTGCCCTGTTTGTATCCCGTGGATTTGTTGGAACGGTTGTTGATTGTGGATAATATTGTGCGGTTGGGCAT
TTATCGGCATTTTGAAAAAGAAATTAAAGAAGCCTTGGATTATGTGTATCGGCATTGGAATGAACGGGGCATT
GGCTGGGGCCGGTTGAATCCCATTGCCGATTTGGAAACCACCGCCTTGGGCTTTCGGTTGTTGCGGTTGCA
TCGGTATAATGTGTCCCCCGCCATTTTTGATAATTTTAAAGATGCCAATGGCAAATTTATTTGTTCCACCGGC
CAATTTAATAAAGATGTGGCCTCCATGTTGAATTTGTATCGGGCCTCCCAATTGGCCTTTCCCGGCGAAAATA
TTTTGGATGAAGCCAAATCCTTTGCCACCAAATATTTGCGGGAAGCCTTGGAAAAATCCGAAACCTCCTCCG
CCTGGAATAATAAACAAAATTTGTCCCAAGAAATTAAATATGCCTTGAAAACCTCCTGGCATGCCTCCGTGCC
CCGGGTGGAAGCCAAACGGTATTGTCAAGTGTATCGGCCCGATTATGCCCGGATTGCCAAATGTGTGTATA
AATTGCCCTATGTGAATAATGAAAAATTTTTGGAATTGGGCAAATTGGATTTTAATATTATTCAATCCATTCATC
AAGAAGAAATGAAAAATGTGACCTCCTGGTTTCGGGATTCCGGCTTGCCCTTGTTTACCTTTGCCCGGGAAC
GGCCCTTGGAATTTTATTTTTTGGTGGCCGCCGGCACCTATGAACCCCAATATGCCAAATGTCGGTTTTTGTT
TACCAAAGTGGCCTGTTTGCAAACCGTGTTGGATGATATGTATGATACCTATGGCACCTTGGATGAATTGAAA
TTGTTTACCGAAGCCGTGCGGCGGTGGGATTTGTCCTTTACCGAAAATTTGCCCGATTATATGAAATTGTGTT
ATCAAATTTATTATGATATTGTGCATGAAGTGGCCTGGGAAGCCGAAAAAGAACAAGGCCGGAATTGGTGT
CCTTTTTTCGGAAAGGCTGGGAAGATTATTTGTTGGGCTATTATGAAGAAGCCGAATGGTTGGCCGCCGAAT
ATGTGCCCACCTTGGATGAATATATTAAAAATGGCATTACCTCCATTGGCCAACGGATTTTGTTGTTGTCCGG
CGTGTTGATTATGGATGGCCAATTGTTGTCCCAAGAAGCCTTGGAAAAAGTGGATTATCCCGGCCGGCGGG
TGTTGACCGAATTGAATAGTTTGATTTCCCGGTTGGCCGATGATACCAAAACCTATAAAGCCGAAAAAGCCC
GGGGCGAATTGGCCTCCTCCATTGAATGTTATATGAAAGATCATCCCGAATGTACCGAAGAAGAAGCCTTGG
ATCATATTTATTCCATTTTGGAACCCGCCGTGAAAGAATTGACCCGGGAATTTTTGAAACCCGATGATGTGCC
CTTTGCCTGTAAAAAAATGTTGTTTGAAGAAACCCGGGTGACCATGGTGATTTTTAAAGATGGCGATGGCTTT
GGCGTGTCCAAATTGGAAGTGAAAGATCATATTAAAGAATGTTTGATTGAACCCTTGCCCTTGTAA

A.3 Nucleotide sequence of promoter P_{cpcB}

Sequence of promoter, TSSs are highlighted in red colour:

GTCGACCATCAACTTAAAGCATCTTTACAAAGCTAGAGATTGACCCTAGTCGTGGACGACTTACAGATTCAG
GATTGATACGGCCAATCCCGATCGCGATCGCTCTAAATCCCCGTCAGTCAGAGCTTCACAATTTTTAGCGAA
TCTTGTGGCCGCGATCGTTGTATAAGAATGCCAAGGCAACTGGATAAGGTTCACTAATCGTTGCTAAGCGAC
AGTGAACTGCGCCAATTGCCTGACAGGCCCCTCTCGTTTAACAAACGATTTAATGTAAATCATTGTTAAGAGT
CTCTCACAATCGAGAGTTTTCTTGAAGAATGATGGGGACGGTTCAGGTGCAGGGTTTCCCTGCTAGAGAATG
CGAAAAAACCGCGTTCTCGTTTTAGGAATCGAGAGTCAATAAAAGTCGAGAACAGGAGACTGGTTGA

A.4 Nucleotide sequence of terminator ECK120010799

Sequence of ECK120010799 terminator:

GTTATGAGTCAGGAAAAAAGGCGACAGAGTAATCTGTCGCCTTTTTTCTTTGCTTGCTTT

A.5 Nucleotide sequences containing prefix and suffix.

Sequence of PcpcB with prefix and suffix in red colour:

GAATTCGCGGCCGCTTCTAGAGGTCGACCATCAACTTAAAGCATCTTTACAAAGCTAGAGATTGACCCTAGT
CGTGGACGACTTACAGATTCAGGATTGATACGGCCAATCCCGATCGCGATCGCTCTAAATCCCCGTCAGTCA
GAGCTTCACAATTTTTAGCGAATCTTGTGGCCGCGATCGTTGTATAAGAATGCCAAGGCAACTGGATAAGGT
TCACTAATCGTTGCTAAGCGACAGTGAACTGCGCCAATTGCCTGACAGGCCCCTCTCGTTTAACAAACGATT
TAATGTAAATCATTGTTAAGAGTCTCTCACAATCGAGAGTTTTCTTGAAGAATGATGGGGACGGTTCAGGTGC
AGGGTTTCCCTGCTAGAGAATGCGAAAAAACCGCGTTCTCGTTTTAGGAATCGAGAGTCAATAAAAGTCGAG
AACAGGAGACTGGTTGATACTAGTAGCGGCCGCTGCAG

Sequence of RBS-psaF with prefix and suffix in red colour:

GAATTCGCGGCCGCTTCTAGAGTTTAACCAAGGAAACGATTCTTTACTAGTAGCGGCCGCTGCAG

Sequence of BBa_B0015 with prefix and suffix in red colour:

GAATTCGCGGCCGCTTCTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCG
TTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTG
CGTTTATATACTAGTAGCGGCCGCTGCAG

Sequence of ECK120010799 with prefix and suffix in red colour:

GAATTCGCGGCCGCTTCTAGAGGTTATGAGTCAGGAAAAAAGGCGACAGAGTAATCTGTCGCCTTTTTTCTT
TGCTTGCTTTACTAGTAGCGGCCGCTGCAG

Sequence of optimised ag1 gene with prefix and suffix in red colour:

GAATTCGCGGCCGCTTCTAGAGATGGCCGGCGTGTCCGCCGTGTCCAAAGTGTCCTCCTTGGTGTGTGATT
TGTCCTCCACCTCCGGCTTGATTCGGCGGACCGCCAATCCCCATCCCAATGTGTGGGGCTATGATTTGGTG
CATTCCTTGAAATCCCCCTATATTGATTCCTCCTATCGGGAACGGGCCGAAGTGTTGGTGTCCGAAATTAAA
GCCATGTTGAATCCCGCCATTACCGGCGATGGCGAATCCATGATTACCCCCTCCGCCTATGATACCGCCTG
GGTGGCCCGGGTGCCCGCCATTGATGGCTCCGCCCGGCCCCAATTTCCCCAAACCGTGGATTGGATTTTGA
AAAATCAATTGAAAGATGGCTCCTGGGGCATTCAATCCCATTTTTTGTTGTCCGATCGGTTGTTGGCCACCTT
GTCCTGTGTGTTGGTGTTGTTGAAATGGAATGTGGGCGATTTGCAAGTGGAACAAGGCATTGAATTTATTAA
ATCCAATTTGGAATTGGTGAAAGATGAAACCGATCAAGATTCCTTGGTGACCGATTTTGAAATTATTTTTCCCT
CCTTGTTGCGGGAAGCCCAATCCTTGCGGTTGGGCTTGCCCTATGATTTGCCCTATATTCATTTGTTGCAAA
CCAAACGGCAAGAACGGTTGGCCAAATTGTCCCGGGAAGAAATTTATGCCGTGCCCTCCCCCTTGTTGTATT
CCTTGGAAGGCATTCAAGATATTGTGGAATGGGAACGGATTATGGAAGTGCAATCCCAAGATGGCTCCTTTT
TGTCCTCCCCCGCCTCCACCGCCTGTGTGTTTATGCATACCGGCGATGCCAAATGTTTGGAATTTTTGAATA
GTGTGATGATTAAATTTGGCAATTTTGTGCCCTGTTTGTATCCCGTGGATTTGTTGGAACGGTTGTTGATTGT

GGATAATATTGTGCGGTTGGGCATTTATCGGCATTTTGAAAAAGAAATTAAAGAAGCCTTGGATTATGTGTAT
CGGCATTGGAATGAACGGGGCATTGGCTGGGGCCGGTTGAATCCCATTGCCGATTTGGAAACCACCGCCTT
GGGCTTTCGGTTGTTGCGGTTGCATCGGTATAATGTGTCCCCCGCCATTTTTGATAATTTTAAAGATGCCAAT
GGCAAATTTATTTGTTCCACCGGCCAATTTAATAAAGATGTGGCCTCCATGTTGAATTTGTATCGGGCCTCCC
AATTGGCCTTTCCCGGCGAAAATATTTTGGATGAAGCCAAATCCTTTGCCACCAAATATTTGCGGGAAGCCTT
GGAAAAATCCGAAACCTCCTCCGCCTGGAATAATAAACAAAATTTGTCCCAAGAAATTAAATATGCCTTGAAA
ACCTCCTGGCATGCCTCCGTGCCCCGGGTGGAAGCCAAACGGTATTGTCAAGTGTATCGGCCCGATTATGC
CCGGATTGCCAAATGTGTGTATAAATTGCCCTATGTGAATAATGAAAAATTTTTGGAATTGGGCAAATTGGAT
TTTAATATTATTCAATCCATTCATCAAGAAGAAATGAAAAATGTGACCTCCTGGTTTCGGGATTCCGGCTTGC
CCTTGTTTACCTTTGCCCGGGAACGGCCCTTGGAATTTTATTTTTTGGTGGCCGCCGGCACCTATGAACCCC
AATATGCCAAATGTCGGTTTTTGTTTACCAAAGTGGCCTGTTTGCAAACCGTGTTGGATGATATGTATGATAC
CTATGGCACCTTGGATGAATTGAAATTGTTTACCGAAGCCGTGCGGCGGTGGGATTTGTCCTTTACCGAAAA
TTTGCCCGATTATATGAAATTGTGTTATCAAATTTATTATGATATTGTGCATGAAGTGGCCTGGGAAGCCGAA
AAAGAACAAGGCCGGGAATTGGTGTCCTTTTTTCGGAAAGGCTGGGAAGATTATTTGTTGGGCTATTATGAA
GAAGCCGAATGGTTGGCCGCCGAATATGTGCCCACCTTGGATGAATATATTAAAAATGGCATTACCTCCATT
GGCCAACGGATTTTGTTGTTGTCCGGCGTGTTGATTATGGATGGCCAATTGTTGTCCCAAGAAGCCTTGGAA
AAAGTGGATTATCCCGGCCGGCGGGTGTTGACCGAATTGAATAGTTTGATTTCCCGGTTGGCCGATGATAC
CAAAACCTATAAAGCCGAAAAAGCCCGGGGCGAATTGGCCTCCTCCATTGAATGTTATATGAAAGATCATCC
CGAATGTACCGAAGAAGAAGCCTTGGATCATATTTATTCCATTTTGGAACCCGCCGTGAAAGAATTGACCCG
GGAATTTTTGAAACCCGATGATGTGCCCTTTGCCTGTAAAAAAATGTTGTTTGAAGAAACCCGGGTGACCAT
GGTGATTTTTAAAGATGGCGATGGCTTTGGCGTGTCCAAATTGGAAGTGAAAGATCATATTAAAGAATGTTTG
ATTGAACCCTTGCCACTGTAATACTAGTAGCGGCCGCTGCAG