Master of Science by Research in Chemical Biology Thesis

Sugars sensing and amplification using bacterial operons



Artem Shesterikov

Supervised by Professor Benjamin G. Davis and Professor Andrew Baldwin

University of Oxford

St Cross College 2019

Abstract

Carbohydrates, especially oligosaccharides are essential molecules present in all the living organisms. To analyse and biosynthesize saccharides of high interest new techniques are required. In this project, a novel biosynthetic approach for sugar sensing and amplification is proposed. This approach is based on the use of sugar operon regulated recombinant gene expression as a detector for the presence of a specific sugar with subsequent activation of amplification circuits for the given sugar

Initially, bacterial lactose metabolism operon and β-1,4-galactosyltransferase 1 were used in the design of a self-inducing lactose amplification gene circuit. This proof-of-principle model showed that sugars can be sensed using an autoinduction loop system. However, an inevitable effect of promoter leakiness led to constant basal amplification of sugar in the autoinduction system. Consequently, this simple approach was not deemed to be practically useful for sugar sensing and additional complexity was added to the design.

Subsequently, a sucrose-lactose synthetic oscillator was designed; that relies on separating the sensing and biosynthetic parts into distinct compartments to bypass the promoter leakiness. A chromosomally encoded sucrose metabolism operon (Csc) was studied as a part of the design because it is homological to the Lac operon, it has been poorly studied, and sucrose is the cheapest commercially available sugar [1].

Finally, the first sucrose regulated recombinant plasmid for *E. coli* was constructed and tested, with sucrose sensing and following lactose synthesis as part of the designed sucrose-lactose oscillator allowing an initial step towards applicable sugar sensing and amplification system. Using the generated biosynthetic construct, the potential of such approaches was illustrated by studying sucrose transport into *E. coli* cells. The results suggest diffusion of sucrose without any active transport.

Acknowledgments

I thank my supervisor Professor Benjamin G. Davis for the opportunity he gave me to work in his avant-garde group and to conduct independent research. I became both a better researcher and person due to Professor Davis' patient and supportive supervision. I thank my supervisor Professor Andrew Baldwin for his help in NMR studies and scientific writing.

I thank my internal examiner Professor Chris Schofield and my external examiner Professor Lars Nielsen for their careful evaluation of my work and productive scientific discussion.

I thank Dr Nan Yang for constant day to day supervision of my project. He taught me most of the experimental methods and research approaches I used in this project. I also thank Dr Jitka Dadova for her supervision over me in the lab during the first year of my study.

I thank Dr Kim Wals for the support and scientific advice she gave me. I thank Dr Ben Bower, Yuliya Dubianok and Simon Nadal for the proofreading of my thesis and useful scientific advice. I thank Aga Borkowska for useful academic advice. I thank Professor Dan Olteanu my college advisor for the academic advice.

I also thank Dr Charlie Fehl, Dr Sylavin Royer, Dr Abirami Lakshminaryanan, Dr Sebastien Galan, Dr Takuya Machida, Dr David Dias, Dr Francisco Perez-Balderas, Dr Matthew Bilyard, Dr Mylène Richard, Dr Ravinash Krishna Kumar, Gogulan Karunanithy, Gustaf Hemberg, Satoshi Kishigami, Sadra Hamedzadeh, Hayley Bastable, Takvor Takvorian, Virginia Casablancas Antras, Frances Armytage-Green, Sarah Faulkner, Mateusz Imiolek, Alexandra Maria Gafitescu, Yue-Jin Oh, Ed Toma, Wouter Remmerswaal, Aziz Khan Frances Armytage-Green, all the LG 10 members, and all the members of Davis and Baldwin groups.

Finally, I thank my family for funding of my program and their support. I thank my mother Irina for her patience and my brother Egor for his kind support.

Table of contents

Abstract **2**

Acknowledgments **3**

Abbreviations **6**

1. CHAPER I. Introduction **8**

1.1. Sugar-analysis methods **8**

1.2. Sugar sensing using operons **10**

1.3. Synthetic biology circuits **13**

1.4. β-1, 4-Galactosyltransferase 1 ..... **15**

1.5. Sucrose as an unnatural sugar in the *E. coli* Csc operon **16**

1.6. Sucrose-inducible recombinant plasmids **19**

2. CHAPTER II. The sugar PCR concept **20**

2.1. Aims of this project **20**

2.2 Motivation and relevance **20**

2.3. Sugar PCR **22**

2.4. Practical implications of sugar PCR **23**

2.5. Initial design **26**

3. CHAPTER III. Results and Discussion **27**

3.1. Lactose synthesis autoinduction plasmids **27**

3.2. Sucrose based sugar amplification system and its benefits **32**

3.3. Sucrose transport into *E. coli* cells **43**

4. CHAPTER IV. Conclusions 46

4.1. Future perspectives and applications 47

5. Experimental details 49

6. Appendices 74

7. References 88

Abbreviations

^{1}H NMR	Proton nuclear magnetic resonance
Allolactose	6-*O*-β-D-Galactopyranosyl-D-glucose
Ara	Arabinose operon
B4GALT1	β-1,4-galactosyltransferase 1
BSA	*N*-Trimethylsilyl-1-trimethylsilyloxyethanimine
cAMP	3',5'-Cyclic adenosine monophosphate
CAP	3',5'-Cyclic adenosine monophosphate activator protein
CRP	3',5'-Cyclic adenosine monophosphate repressor protein
Csc	Chromosomally encoded sucrose metabolic pathway
CscA	Sucrose-6-phosphate hydrolase gene
CscB	Sucrose permease gene
CscK	D-fructokinase gene
CscO	Csc operator site
CscR	Csc repressor gene
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
Gal	Galactose operon
GalNAc	*N*-Acetylgalactosamine
GALT	Galactosyltransferase gene
GFP	Green fluorescent protein
GC-MS	Gas chromatography mass spectrometry
GlcNAc	*N*-Acetylglucosamine
HPTLC	High-performance thin layer chromatography
IM-MS	Ion-mobility mass spectrometry
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVTT	*In vitro* transcription/translation
Lac	Lactose operon
LacA	Galactoside *O*-acetyltransferase gene
LacI	Lac repressor protein gene
LacY	Lactose permease gene
LacZ	β-galactosidase gene

LB	Lysogeny broth
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
MW	Molecular weight
OD	Optical density
PCR	Polymerase chain reaction
RFP	Red fluorescent protein
Rha	Rhamnose operon
RNAP	Ribonucleic acid polymerase
SDM	Site-directed mutagenesis
SPS	Sucrose-phosphate synthase
SOC	Super optimal broth with catabolite repression
Tac	Tryptophan operon + lactose operon hybrid promoter
TAE	Tris-acetate-ethylenediaminetetraacetic acid
TMG	Methyl-1-thio-β-D-galactopyranoside
TRIS	2-Amino-2-(hydroxymethyl)propane-1,3-diol
TXR	Texas Red
UDP	Uridine diphosphate
UDP-gal	Uridine diphosphate galactose
UV	Ultraviolet
Xyl	Xylose operon

1. CHAPER I. Introduction

1.1. Sugar-analysis methods

Sugars are involved in many cellular processes, including as a carbon source in the case of oligosaccharides, and as signalling molecules in glycoproteins. Sugar sensing can be defined as the detection of a specific signal related only to a particular sugar of interest. Many organisms sense sugars, and the expression of specific genes can be regulated in the presence of a given sugar. This biological sensing depends on sugar chemistry because enzymes recognise specifically chemical structures. Observation of sugar interactions with different enzymes and regulatory proteins could give an insight about the molecular nature of this sugar. As a result, enzymatic sugar sensing could be used to identify a given sugar. Such a technique would add to the current pool of available analytical techniques for carbohydrates.

Sugars can be directly analysed by various contemporary methods[2] such as gas chromatography mass spectrometry (GC-MS). However, there are several limitations of this method. Carbohydrates require additional functionalisation for GC-MS analysis. As a result, functionalised sugars are analysed, with consequent potential loss in accuracy, particularly in the case of large and complex sugars. Moreover, GC-MS is not a very precise method for identifying individual sugars: MS cannot distinguish epimeric monosaccharides[3] neither provide information about the type of glycosidic linkages. To distinguish linkages in oligosaccharides, NMR can be used, but it is challenging to distinguish different monosaccharide blocks due to their similar chemical shifts. The ion-mobility mass spectrometry (IM-MS) method is probably the best way to sequence carbohydrates, because it is able to distinguish ions with identical masses with a high accuracy. However, this method is expensive and not well developed[4]. When applied together with NMR, these methods can provide a good picture of the composition of the polysaccharide of interest[5]. However, chromatographic methods depend on expensive equipment

and cannot determine the intracellular concentration[6]. Moreover, comprehensive information about the stereochemistry of sugars requires NMR or X-ray crystallography analysis[7]. Thus, both MS and NMR methods are often needed to characterise a specific sugar, which is expensive, time-consuming, and requires expertise. Biosynthetic methods that take advantage of the natural specificity between enzymes and their substrates are a possible alternative for NMR and MS (Figure 1).

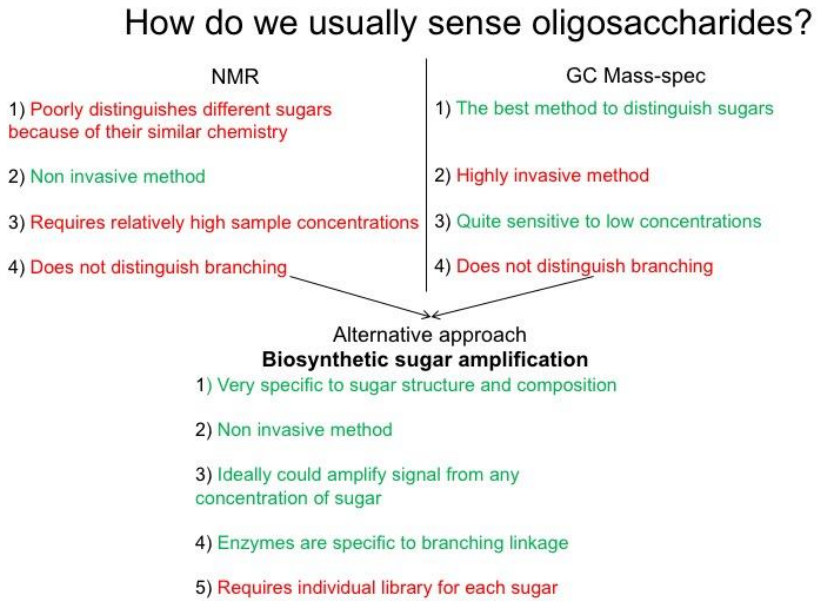


Figure 1. Comparison of NMR and GC-MS, the two main methods for sugar characterisation, with potential new biosynthetic methods employing sugar amplification.

The main aim of this project was to construct a biosynthetic system for the detection of simple carbohydrates. This model system could serve as a proof-of-principle for creating a library for sensing and amplification of specific sugars. Such an analytical method would be rewarding in medicine as sugar structures in glycoproteins are associated with diseases. Specific libraries for disease-related oligosaccharides could be used for brief, simple, and non-invasive diagnostic purposes. The low detection limit of biosynthetic sensing and amplification, relying

on high-affinity riboswitches for detection and operons for sugar-specific amplification, would be a powerful tool for sensitive and specific detection of sugars in biological system.

1.2. Sugar sensing using operons

The main advantage of operon-based methods, which are discussed in this project, over MS and NMR is simplicity, because they only require libraries of sugar operon circuits and bacterial cultures. Moreover, sugar-sensing circuits could potentially give concentration-dependent feedback based on oscillations of fluorescence signals. Like many other human 'inventions', sugar sensors are already present in nature. These natural genetic circuits allow organisms to manipulate metabolism and to conserve resources so that genes for metabolic and transport enzymes are expressed only in the presence of their specific sugar substrates and/or silenced in their absence[8]. For example, plants express the gene of hexokinase enzyme involved in the sucrose and glucose sensing pathway[9]. Similarly, bacterial sugar operons are simple, but robust natural sugar sensors, controlling groups of genes via negative or positive feedback[10]. This project mainly focused on the prokaryotic operons responsible for sugar metabolism because they are simple and representative of natural sugar sensors. Of course, eukaryotes also possess systems for sugar-sensing, as reported for plants and the yeast *Saccharomyces*. However, they are much more complicated and not very suitable for biosynthetic constructs.

Operons involve polycistronic genome organisation in which several genes (generally one gene = one protein) are collectively transcribed from DNA to mRNA, because they are involved in common metabolic pathways or other cellular processes. Operons are mainly present in prokaryotic organisms (bacteria and archaea)[11], possibly because unlike eukaryotic cells, prokaryotic cells must support themselves without organelles, and eukaryotic cells contain more introns and require specific regulation for each gene. As a result, there is a large time difference

between transcription and translation in eukaryotes in comparison with prokaryotes; it is considerably longer in eukaryotic cells. Interestingly, an analogue of an operon (a dicistronic transcript) has been found in *Drosophila melanogaster*[12]. Polycistronic transcripts have also been found in nematodes and trypanosomes[13], but operons are still considered an attribute of prokaryotic organisms.

Operons have a specific DNA structure and can have one promoter for all genes within the operon, to which RNA polymerase (RNAP) binds and begins transcription of the whole polycistron. The most significant part of an operon is the operator, which binds specific protein repressors that regulate transcription and gene expression. This system is well researched in the case of the *Escherichia coli* lactose (Lac) operon, which contains three sequential genes (LacZ, LacY and LacA). Analyses have provided insights on the three-dimensional structure of the repressor protein (LacI) and its binding to the operator sequence of the Lac operon and the operon inducer isopropyl-β-delta-1-thiogalactoside (IPTG)[14]. The *E. coli* rhamnose (Rha) operon is also well-studied, and has a more complex structure than the Lac operon, with five genes coding five proteins (RhaEW, RhaR, RhaB, RhaM and RhaA)[15]. However, only the function of RhaEW is well-defined; this gene encodes a bifunctional L-rhamnulose-1-phosphate aldolase / L-lactaldehyde dehydrogenase enzyme responsible for two steps of rhamnose metabolism – cleavage of the aldose group, and subsequent hydrogenation[15]. Other proteins are also important in the regulation of rhamnose metabolism[15]. While the Lac operon is negatively regulated, the Rha operon is regulated both negatively and positively[16]. In addition to Lac and Rha operons, there are operons for most mono- and oligosaccharides, including Ara (L-arabinose)[17], Xyl (Xylose)[18], Gal (galactose)[19], D-Allose[20] and Sac/Scr/Csc (sucrose) [21,22,23,24].

The sucrose operon is poorly studied in *E. coli* because most industrial strains do not contain a sucrose operon and cannot metabolise sucrose[25]. A search of the existing literature indicates that the sucrose operon has not been used in recombinant plasmids for *E. coli*, and most

studies on the sucrose operon focus on sucrose metabolism, but not the regulatory role of sucrose. Sucrose is also an important signalling molecule for plants. Moreover, only oxygenic phototrophs such as plants can synthesise sucrose[26,27]. As decomposers, some bacteria are able to metabolise sucrose using specific Csc or Sac operons that have a very similar structure.

Nonetheless, operon structure can vary between genera. For example, natural sucrose metabolism regulated by the Csc operon is present in less than half of wild *E. coli* strains[28], whereas *Lactococcus* can catabolise saccharose and members possess specific operons for this (SacBK and SacAR) that are regulated by the SacR repressor protein, which blocks transcription from the Sac promoter in the absence of sucrose[29]. Despite the differences, the organisation and products of these operons are quite similar. For example, the *E. coli* Csc operon contains analogues of all genes in the Sac operon of *Lactococcus*[25]. Additionally, the Scr repressor protein belongs to the LacI-GalR repressor family and shares 60% sequence similarity[30] with the Csc repressor that is from the same family, but is not well-studied[31].

Many operons are positively controlled by cyclic adenosine monophosphate (cAMP), an important molecule in chemotaxis and cellular signalling[32]. The cAMP repressor protein (CRP) changes its conformation when binding cAMP, resulting in binding to the promoter at the transcription start site of a single gene or operon[33]. The main impact of CRP might be on the transcription initiation process. For example, in the presence of cAMP, CRP binds to the P1 promoter of the Lac operon, activating its transcription because RNA polymerase is more likely to bind a promoter to which CRP is bound[33].

The natural modularity of bacterial operons has been utilised to make synthetic biology circuits, and the Lac operon has been widely applied in the design of different synthetic biology circuits in bacterial cells[34]. Similar to electronic circuits, synthetic biology circuit components can be modified, replaced and removed. This important feature allows choosing specific genes as model objects and using them in a model circuit design which is expected to work similarly for

any homologous genes. For example, β-1,4-galactosyltransferase acts as a model for all glycosyltransferases, and LacI is a model for all sugar transcription regulators, as used in this project to design a model circuit for sugar sensing and amplification.

1.3. Synthetic biology circuits

Synthetic biology is a rapidly developing area, including in the field of circuit engineering. Specific strong promoters are required to construct gene circuits as they are tunable and do not require large amounts of repressor protein for deactivation. For example, the P_{LlacO1} promoter contains two operators with one directly between the RNA binding boxes. However, this phage λ promoter requires viral RNA polymerase (RNAP) to be encoded in the *E. coli* genome[35]. Viral RNAPs are widely used in gene circuits and for gene expression because they bind specific promoters different from those of bacterial RNAPs. The most popular viral RNAP is T7 phage RNAP, which is so efficient that when bacterial cells contain T7 RNAP, up to half of all proteins inside the cell may be translated from the genes expressed from the T7 promoter.[36]. For example, BL 21 (DE3) *E. coli* strain has T7 RNAP gene inserted into chromosome and this strain is widely used for highly efficient gene expression. Nowadays, T7 RNAP gene is also expressed in eukaryotic cells[37], and the eukaryotic nuclear gene encoding mitochondrial RNA polymerase was found to be homologous to viral T7 and T3 enzymes[38]. This may represent indirect evident that the nucleus of eukaryotes has a symbiogenetic origin.

There are many different ways to regulate gene expression using sugar operons. In the example of lactose induction, many different promoters are known, including PlacUV5, Ptac, PLac/ara, PLacO, PAlacO and PT7lacO[35]. However, it is not necessary to use sugar operons because riboswitches have similar functions, and can be synthesised specifically for the molecule

of interest[39]. However, the library of modules which could be used in biosynthetic circuits remains very limited[40].

Synthetic gene oscillators – several genetic genetic systems with interconnected expression – are a promising tool for synthetic biology and thus sugar sensing. These systems provide insight on gene regulation and gene interactions with the different signalling molecules such as inducers. Many oscillators are based on positive or negative inducible operons such as Ara or Lac, respectively[34]. Repressor or activator genes are inserted together with fluorescent protein genes downstream of a specific promoter regulated by some of these genes. Usually, a small stable messenger RNA A degradation tag is added to each protein sequence to decrease mRNA lifetime and oscillation period[41]. As a result, oscillators can be easily constructed as plasmids and transformed into cells or used in *in vitro* transcription-translation systems. Oscillators can be tuned by changing the concentration of the inducer[42]. Consequently, this can be used for specific sugar sensing if a sugar or its derivatives act as inducers of the oscillator. Allolactose is a natural inducer of the Lac operon, and IPTG is a synthetic inducer. The main advantage of IPTG is that it is not metabolised by most cells[43]. Nearly 20% of lactose is converted to allolactose by β-galactosidase (LacZ), which is responsible for breaking down lactose to glucose and galactose. The affinity of allolactose for the lac repressor protein is even higher than that of IPTG[44,45,46]. The LacY product is responsible for the uptake of lactose and others inducers (IPTG, TMG, etc) inside the cell; however, at high concentrations, IPTG can diffuse into the cell through the membrane[47]. Unfortunately, IPTG at a concentration greater than 50 mM is toxic to both prokaryotic and eukaryotic cells, and no more than 1 mM is usually added in bacterial expression systems[48].

Theoretically, any nucleic acid-encoded enzyme can be used as a component of a synthetic biology circuit. There are several limitations of course, such as the ability of gene transcription to be initiated and regulated, the ability of a gene to be expressed and a protein

properly folded in a particular cellular type, protein toxicity, and functional aspects. Enzymatic components that amplify a specific molecule can be included into synthetic biological sensors to allow the sensing of low concentrations of this molecule. For example in terms of sugar amplification, the bovine galactosyltransferase gene could be expressed in *E. coli* for lactose amplification.

1.4. β-1, 4-Galactosyltransferase 1

All living cells express specific glycosyltransferase genes to produce sugars, grow polysaccharide chains, and link sugars to proteins and others molecules[49]. β-1,4-Galactosyltransferase is one of the most well-studied glycosyltransferases[50]. Free galactosyltransferase performs the catalytic function of making glycosidic bonds between glucose and galactose synthesising lactose (Figure 2); however, synthesis of lactose from glucose as a sugar acceptor and UDP-galactose as a sugar donor is quite inefficient due to the high Km of β-1,4-galactosyltransferase 1 with glucose. The β-1,4-galactosyltransferase gene can be routinely expressed in bacterial cells with appropriate optimisation. For example, the native bovine gene encoding GALT has a membrane domain that must be deleted for efficient translation of active β-1,4-galactosyltransferase 1 in *E. coli* cells[51].

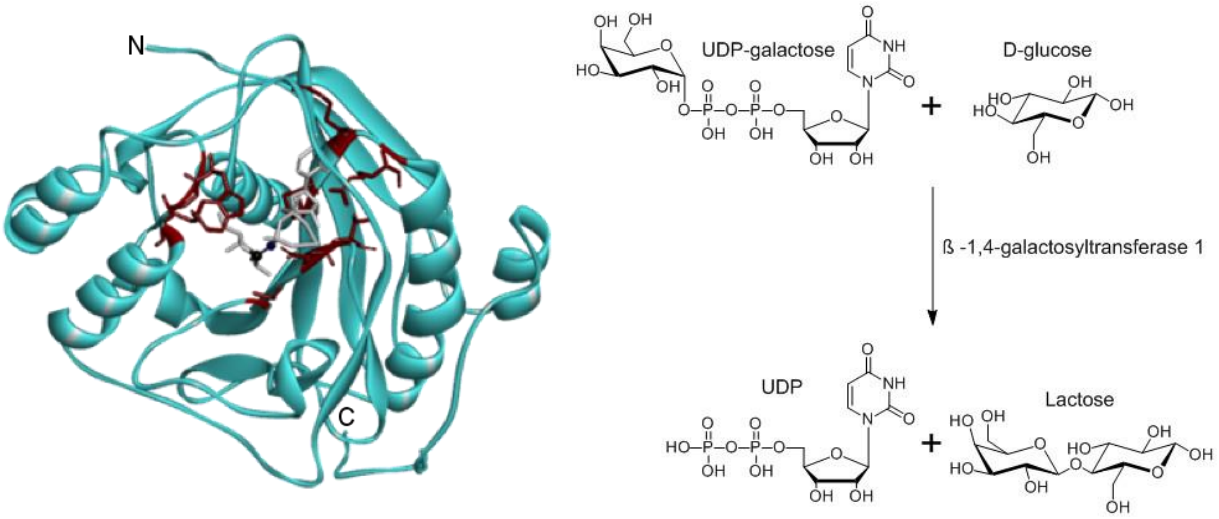


Figure 2. View from a crystal structure of the B4GALT1 monomer showing UDP-gal and the catalytic site [52]. Active site residues are shown as red sticks. UDP-gal is coloured light grey. The galactose anomeric carbon is highlighted in black, and the glycosidic bond oxygen is highlighted in dark blue. The structure is derived from RCSB PDB[53] code 1O0R.

Notably, β-1,4-galactosyltransferase 1 forms a protein complex with α-lactalbumin in mammalian cells called lactose synthase. α-Lactalbumin was the first protein from lactose synthase complex to be studied[54]. It changes the structure of β-1,4-galactosyltransferase 1, increasing its binding affinity to glucose nearly 1000-fold[55]. The lactose synthase complex is the main cellular lactose production mechanism.

1.5. Sucrose as an unnatural sugar in the *E. coli* Csc operon

Sucrose is known as a reserve disaccharide mainly produced by plants and other photosynthetic organisms in very large amounts, whereas animals do not produce much sucrose, and their satellite bacteria such as *E. coli* do not need to break down this sugar. Sucrose is metabolised by only 50% of all wild type *E. coli* strains, and even fewer laboratory strains[56]. However, no attempts have been made to create sucrose-inducible vectors for *E. coli*, even

though the regulatory operon CscAKB was known for a long time[25]. Moreover, the sucrose operon repressor protein CscR is in the same family as the LacI repressor protein, and binds to a similar DNA site[57]. The inducer for CscR has not been identified, but is probably sucrose or fructose[31]. To date, there is no crystallographic or NMR structure of the CscR protein, which is 331 amino acids long, has a molecular weight (MW) of 36,470.74 g/mol, is expected to be stable according to predictions (ExPASy ProtParam tool)[58]. The CscR protein is a LacI family member based on sequence alignment (Table 1).

Score	Expect	Method	Identities	Positives	Gaps
80.1 bits(196)	2e-21	Compositional matrix adjust.	90/346(26%)	158/346(45%)	38/346(10%)

```
CscR  3    SLKDVARLAGVSMMTVSRVMHNAESVRPATRDRVLQAIQTLNYVPDLSARKMRAQGRKPS  62
           +L DVA  AGVS  TVSRV++ A  V   TR++V  A+  LNY+P+  A+++     K S
LacI  5    TLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNYIPNRVAQQLAG---KQS  61

CscR  63   TLAVLAQDTATTPFSVDILLAIEQTASEFGWNSFLINIFSED--DAARAA-RQLLAHRPD  119
            L  +A  +          I+ AI+  A + G  S ++++       +A +AA   LLA R
LacI  62   LLIGVATSSLALHAPSQIVAAIKSRADQLG-ASVVVSMVERSGVEACKAAVHNLLAQRVS  120

CscR  120  GIIYTTMGLRHITLPESLYGENIVLANC---------VADDPALPSYIPDDYTAQYESTQ  170
           G+I        I  P        V A C         V+D   + S I           +
LacI  121  GLI--------INYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVE  172

CscR  171  HLLAAGYRQPLCFWLPESALATGYRRQGFEQAW-RDAGRDLAEVK-QFHMATGDDHYTDL  228
           HL+A G++Q      P S+++   R  G+ +   R+  + +AE +  +   +G   +
LacI  173  HLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSG---FQQT  229

CscR  229  ASLLNAHFKSGKPDFDVLICGNDRAAFVAYQVLLAKGVRIPQDVAVMGFDNLVGVGHLFL  288
             +LN            ++  ND+ A  A + +   G+R+  D++V+G+D+       ++
LacI  230  MQMLNEGIVP-----TAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTED-SSCYI  283

CscR  289  PPLTTIQLPHDIIGREAALHIIEGREGGGRVTR---IPCPLLIRCST  331
           PPLTTI+    ++G+ +   +++  +G  V     +P  L+ R +T
LacI  284  PPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTT  329
```

Table 1. Alignment of CscR and LacI protein sequences obtained by BLAST[59] search.

Two important sites are present in the CscR amino acid sequence – the DNA-binding site near the N-terminus, and the sugar-binding site near the C-terminus. According to the existing literature and based on LacI as an example, the CscR protein likely binds a specific DNA operator site (Table 2)[60].

LacI operator sequence	AATTGTGAGCGGATAACAATT
CscR operator sequence	AGTTAACGTTAACA

Table 2. Comparison of the CscR and LacI operator sequences. According to data from Porcheron and colleagues[60], CscO is ~five base pairs shorter than LacI, but both sequences are palindromic and occupy neighbouring positions within the promoter site[61].

The sugar-binding site of the CscR protein has not been thoroughly studied as indicated by a literature search. According to the existing literature, CscR should possess high affinity for sucrose or fructose[31]; however, the CscR protein may potentially bind others sugars such as glucose because of their chemical and structural similarity. Thus, the true inducer of CscR is still unknown, and this problem requires further investigation (Figure 3).

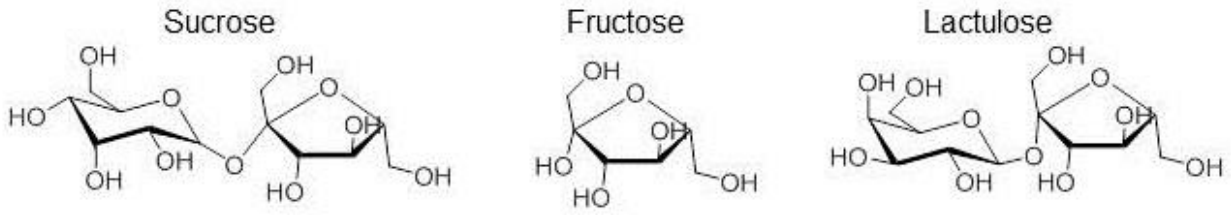


Figure 3. Potential CscR inducers. Sucrose or Fructose should be natural inducers of CscR and lactulose is beneficial molecule for Lac/Csc system because it is expected to induce both Lac and Csc operons.

The CscR protein is interesting in evolutionary terms because it is quite distant from homologous sucrose regulators such as the ScrR repressor in other organisms[57]. Moreover, phylogenetic analysis shows that the CscR branch could be more evolutionarily ancient than the sister LacI branch[57]. Thus, it could be speculated that ancestors of *E. coli* were mostly feeding on sucrose, which is produced in vast quantities by plants, and only later switched to modern lactose metabolism due to a change in environment to mammalian organisms, which are rich in lactose but not sucrose.

1.6.Sucrose-inducible recombinant plasmids

Sucrose is a cheap source of carbon, and cultivating microorganisms on sucrose as a substrate is an attractive proposition. Nielsen and colleagues have made an impact on the study of sucrose metabolism in *E. coli*[62,28,25,63,64,31]. They studied the Csc operon, which is present in some wild type *E. coli* strains, and contains genes for sucrose transportation and breakage. However, researchers have studied the application of this operon in industrial *E. coli* strains to make them grow on sucrose substrates, rather than studying sucrose as a transcriptional regulator. As a result, the sucrose operon regulator protein, which probably interacts with sucrose, is poorly characterised; there is no direct evidence that sucrose itself binds to this protein, no crystal structure of CscR, and no evidence that *E. coli* recombinant expression systems are induced by sucrose[31]. A sucrose metabolism plasmid for *E. coli* K12 was first described by Charles and colleagues in 1979[65], and Lengeler and colleagues later made and tested several sucrose-inducible plasmids based on DNA from the EC3132 strain[66]. For example, plasmid pJBL101 contained the CscR gene together with other Csc operon genes (CscA, CscK and CscB), but a plasmid containing CscR with recombinant genes such as GFP was not reported. The CscR gene has been used only as a part of the whole CscRABY operon, but not separately, even in a recent study on sucrose metabolism in *Pseudomonas putida* published in May 2018[67]. Although the lack of sucrose metabolism in *E. coli* is usually deemed a disadvantage, this was useful for designing and testing sucrose induction systems in this project as a part of sugar amplification system.

2. CHAPTER II. The “sugar PCR” concept

2.1. Aims of this project

The main goal of the work described in this project was to create a system capable sensing and amplifying a specific poli- or oligosaccharide of interest. Oligosaccharides are built from modules: furanose or pyranose carbohydrate rings[68]. They can form long, branched structures linked by glycosidic bonds. Sugars are biopolymers and the growth of their chains is catalysed by enzymatic reactions. It was envisaged that this fact could be used to create an amplification system for sugars based on glycosyltransferases.

2.2 Motivation and relevance

Complex sugars are usually synthesised *in vitro* using synthetic organic chemistry methods, which involve many protection and deprotection steps[69], resulting in numerous side products. This is often impractical for long polysaccharide chains. However, there are many natural glycosyltransferase enzymes that allow us to build complex sugars from simple sugar donor building blocks. Enzymatic synthesis is potentially more efficient because of the selectivity of glycosyltransferases [70].

Many attempts have been made to synthesise specific sugars for medical needs, both chemically and enzymatically. For example, there has been major interest in heparin synthesis because heparins are involved in many cellular processes, especially prevention of thrombus formation. One group synthesised heparin tetrasaccharides by chemical sequential glycosylation[71], but this chemical approach is expensive and time-consuming in comparison with enzymatic sugar synthesis by living cells. An interesting alternative to chemical synthesis is enzymatic synthesis based on sensor chips, as demonstrated using sulfotransferases and C5-

epimerase in the synthesis of heparins and heparin sulphates. However, this screening method is useful for modifying existing oligosaccharides, rather than their *de novo* synthesis[72].

Using synthetic biology methods to create specific gene circuits inside living cells, systems could potentially be made for the synthesis of practically any natural polysaccharide. For example, the cellular machinery could be harnessed to make important parts of glycans for antibodies. In such a system glycosyltransferases would have an analogous function as DNA polymerases in polymerase chain reaction (PCR), hence the term “Sugar PCR” has been used.

2.3. “Sugar PCR”

The main idea of the “sugar PCR” method is to use biosynthetic enzyme libraries to sense and amplify any desired sugar. Because polysaccharides are polymers synthesised by specific glycosyltransferases, and because DNA is also a polymer synthesised by DNA polymerase, PCR amplification of DNA can be compared with hypothetical sugar synthesis (Figure 4). A more specific example of “sugar PCR” is presented in Figure 5.

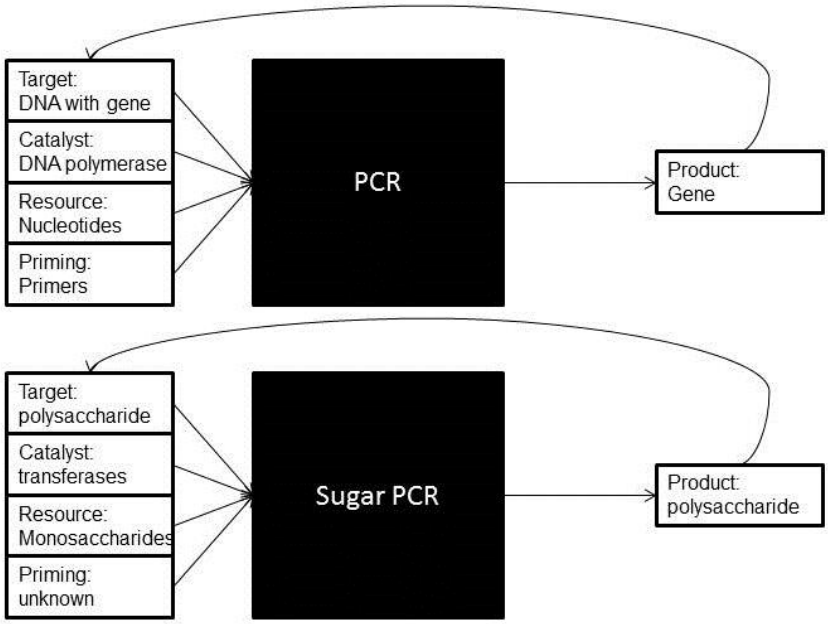


Figure 4. Comparison of PCR and “sugar PCR”. The fundamental difference is the use of primers, which are required to initiate the transcription of DNA and are responsible for the specificity of products in PCR. By contrast, “sugar PCR” is relatively unspecific because the catalysis by glycosyltransferases is not linked with the structure of the polysaccharide amplified. Preferential polysaccharide production is based on specific enzymes and their interactions.

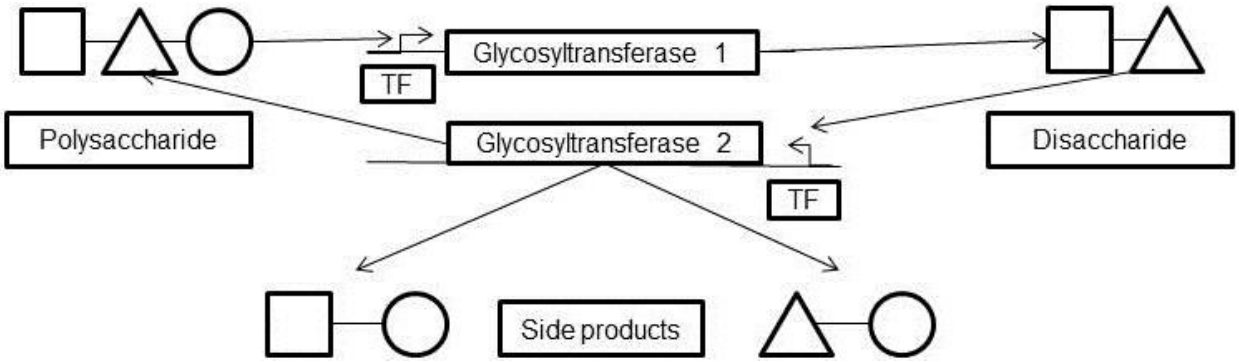


Figure 5. Amplification of a trisaccharide from three different monomers. Firstly, a trisaccharide such as raffinose, simple trisaccharide built from three different monomers (galactose, glucose and fructose), binds a specific riboswitch (transcriptional/translational factor) that induces the expression of the first gene for an enzyme making a disaccharide such as sucrose, which is not catabolised by this specific cell. The amplified disaccharide then binds a specific transcription factor such as the sucrose repressor, which induces the expression of a second glycosyltransferase gene, resulting in the addition of a specific monosaccharide to the amplified disaccharide. This enzyme connects a sugar donor to any available sugar acceptors, and not only the amplified disaccharide, resulting in side products.

2.4. Practical implications of "sugar PCR"

Unfortunately, if longer oligosaccharides are being amplified, there are more side products for each step, resulting in lower product yield. Each polysaccharide requires its own sensors and enzymes. However, even DNA polymerase in PCR is not specific, but amplification of desired product is achieved by primer specificity and complementation. As a result, there is no direct analogy with DNA amplification because there is no analogue of nucleotide complementation in sugars.

Another challenge of this approach is that living cells usually metabolise sugars as a carbon source by essential cellular glycosidases. Due to the complexity of cellular systems, it is

not fundamentally possible to efficiently synthesize any sugar in natural cells. To prevent this, initial building blocks which are not metabolised by cells should be chosen. For example, 50% of natural *E. coli* strains do not metabolise sucrose [24]. If *E. coli* could synthesise and amplify sucrose, the first step in amplification of many sucrose-containing oligosaccharides such as raffinose would be achieved (Figure 5).

The second problem is that enzyme specificity and even glycosyltransferase synthesis specificity is limited to some extent. There are two participants in oligosaccharide formation by glycosyltransferases: a sugar donor and a sugar acceptor. Even if the sugar donor is quite specific for a particular glycosyltransferase, there are many possible sugar acceptors[73]. As a result, each glycosyltransferase can use more than one sugar acceptor as a substrate, consequently forming more than one product. This is a major drawback because cellular resources are not infinite, and cells can produce a limited amount of sugar. If cells make many side products alongside the production of polysaccharide of interest, yield of the target sugar decreases exponentially during each step of amplification.

Additionally, glycoside 'primers' can be used to control glycosylation, but they are very polar and struggle to pass through the cellular membrane[74]. Glycoside primers are artificial glycoside molecules which serve as a substrate for glycosyltransferases. These primers occupy glycosyltransferases, preventing unnatural product formation by these enzymes while natural substrates for glycosyltransferases are not involved in oligosaccharide synthesis. Some groups found a way to make glycoside primers pass through the Golgi apparatus membrane and inhibit specific natural glycosylation pathways[75]. For example, 4-methylumbelliferyl-β-xyloside is a popular glycoside primer for *N*-linked glycan chains, and GalNAc-α-phenyl functions for *O*-linked forms[76]. To date, glycoside primers have not been studied in bacterial cells.

There are many *N*-linked glycans with sialic acid residues at the end of the oligosaccharide chain that could be used for sensing and amplifying signals from specific

oligosaccharides (Figure 6). A successful "sugar PCR" model system could in principle sense and amplify any sugar from the glycome of a specific organism. Certainly, there are many difficulties regarding sugar amplification due to the complex nature of cellular sugar metabolism. Synthesis of *N*-linked glycans is particularly interesting because they are responsible for protein glycosylation. GlcNAc is therefore a logical choice for the first step in *N*-linked polysaccharide synthesis, but unfortunately GlcNAc is metabolised by *E. coli* and is naturally present in most cells as a component of the murein cell wall [77].

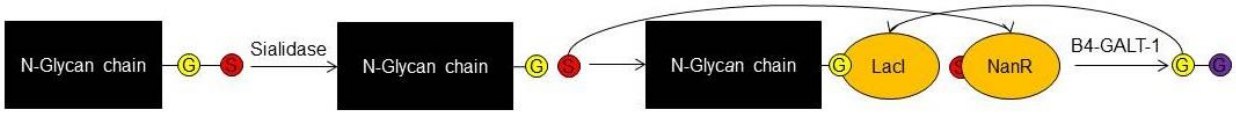


Figure 6. Sensing and amplification of *N*-linked glycans ending with galactose-sialic acid chains using sialidase, the LacI (lactose) repressor, and the NanR (sialic acid) repressor[78]. Amplification of the signal is performed by galactosyltransferase. However, this sensor will likely also amplify signals from *N*-linked sialic acid glycans without galactose due to galactosyltransferase autoinduction. This could be prevented by addition of an intermediate synthesis step, or using another cell type for compartmentalisation.

2.5. Initial design

A simple model system that could subsequently be transformed into a more applicable construct had to be developed. At the beginning of this project it was decided to attempt to use isolated glycosyltransferase, similar to isolated DNA polymerase in PCR, to amplify the sugar of interest. Bacteria cultures are used to produce sugar acceptors and donors for this enzyme. Key molecules should flow through connected compartments and induce the synthesis of reaction substrate step by step. The amplification of lactose was considered as a model objective, because lactose is the most biologically studied sugar (Figure 7).

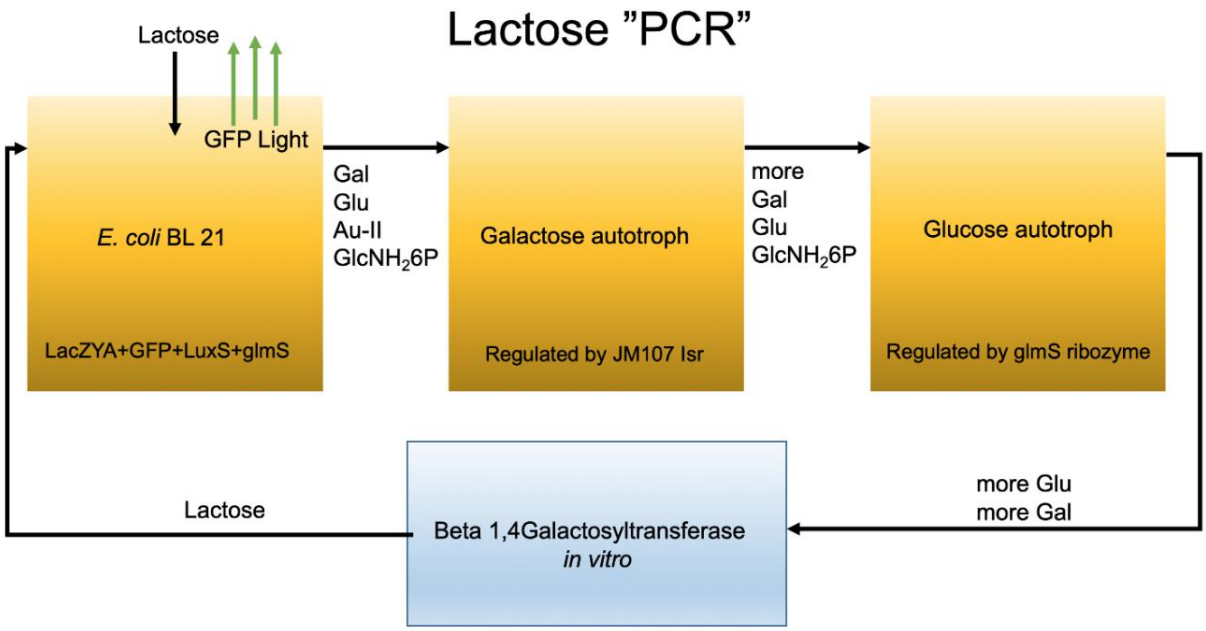


Figure 7. Compartmentalized lactose amplification system is regulated by glmS riboswitch and Lux operon. Three bacterial cultures are put in different compartments. The first culture expresses GFP gene in presence of lactose and synthesize small molecules inducers for the next two cultures. Second and the third cultures synthesize galactose and glucose. Isolated beta-1,4-GALT protein catalyses the synthesis of lactose in fourth compartment.

3. CHAPTER III. Results and Discussion

3.1. Lactose synthesis autoinduction plasmids

The synthetic oscillator designed in this project is based on induction caused by the synthesis of a regulatory molecule by a recombinant enzyme. The first part of the lactose-sucrose oscillator is lactose synthesis. To make this oscillator work, production and consumption of the inducer must be regulated, which is not possible with IPTG, hence the biological inducer was used instead. To produce lactose, the β-1,4-galactosyltransferase I enzyme was chosen, along with red fluorescent protein gene expression under the control of the lac promoter as an indirect signal of lactose production level by β-1,4-galactosyltransferase I in *E. coli* cells (Figure 8).

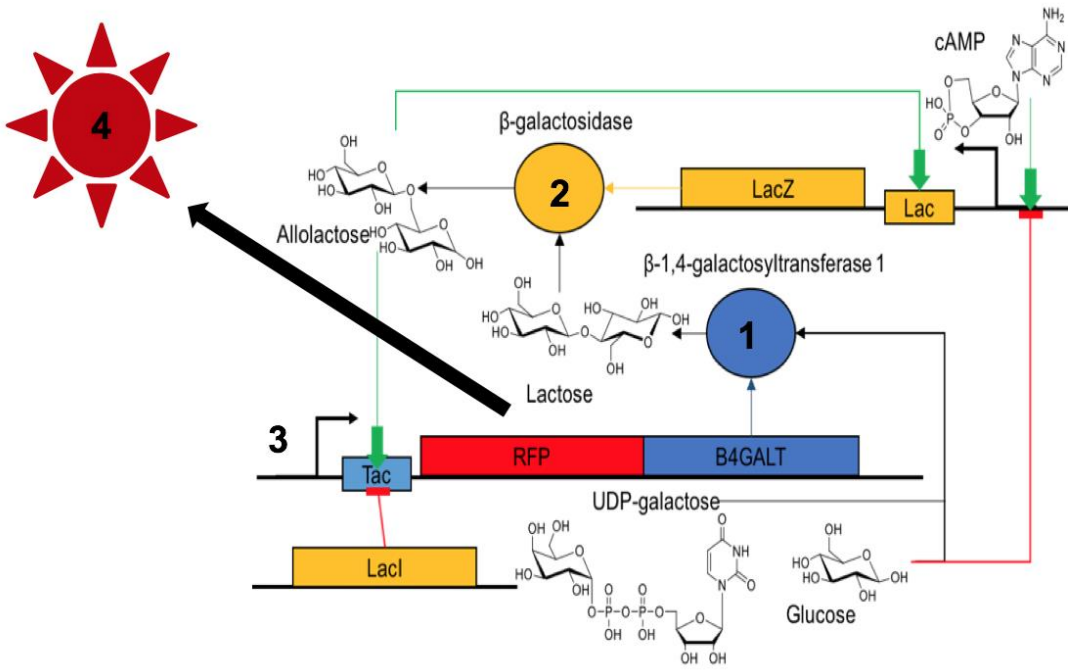


Figure 8. The lactose-inducible B4GALT1 expression system. 1 – The B4GALT1 gene is basally expressed from a plasmid and catalyses the synthesis of lactose from glucose and UDP-galactose. 2 – Lactose is transformed into allolactose by the basally expressed and translated chromosomally encoded β-1,4-galactosyltransferase 1. 3 – Allolactose binds the Lac repressor, inducing the expression of B4GALT1. 4 – The red fluorescent protein (RFP) signal is observed following accumulation of RFP fused with B4GALT1. cAMP can be added to induce expression in the presence of glucose.

The pCW-LIC vector[79] (plasmid 1) was considered suitable for construction of the synthetic system because it contains three sequential Tac promoters followed by suitable restriction sites. The Tac promoter is a synthetic hybrid of the Trp and Lac promoters[80]. Unlike the natural Lac promoter, the Tac promoter is not repressed by glucose, and has a higher affinity for RNA polymerase[80]. Starting from empty pCW-LIC, Plasmids pCW-RFP and pCW-RFP-GALT were constructed (Figure 9) by inserting B4GALT1 directly after the mCherry[42] RFP gene, and B4GALT1 and RFP were fused to directly link the fluorescence signal with the expression level of B4GALT1.

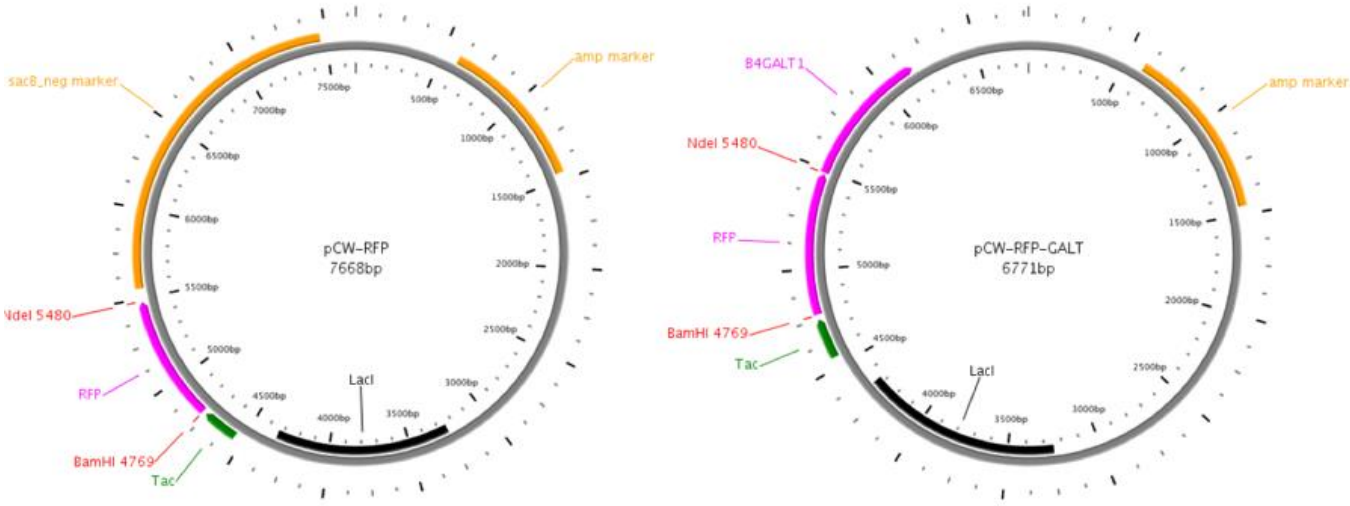


Figure 9. Plasmid maps for pCW-RFP and pCW-RFP-GALT constructs. For the full sequences, see the Appendices.

The construct containing the RFP-B4GALT1 fusion protein was successfully translated in *E. coli* BL21 cells cultured in 0.4% glucose M9 medium. Additionally, 10 mM cAMP was added to activate transcription of LacZ from the chromosomally encoded *E. coli* Lac promoter[81].

A similar approach has been described by Chen et al.[82] who used a plasmid containing the UDP-galactose-4′-epimerase gene fused with the β-1,4-galactasyltransferase gene under the control of a lactose-inducible promoter. UDP-galactose-4′-epimerase is also naturally present in

most *E. coli* strains, and was presumably chosen after testing the conversion of UDP-glucose to UDP-galactose in *E. coli* cells to make different oligosaccharides[82]. Despite the similarities, the main research question addressed by Chen et al.[82] was different, but they happened to use similar enzymes to those employed in the present work. Their results showed that IPTG addition increased the concentration of lactose in *E. coli* cells, and higher concentrations of IPTG yielded more lactose product. However, they did not discuss the role of synthesised lactose itself as an inducer of gene expression, whereas the present project aimed to explain the role of newly synthesised lactose in gene expression, and the expected autoinduction. Thus, the RFP gene was fused with the β-1,4-galactasyltransferase gene, placed under the under control of a lactose-inducible promoter, and gene expression was measured based on RFP fluorescence intensity. Surprisingly, the expression level of RFP was maximal without IPTG induction. Moreover, addition of higher IPTG concentrations led to a lower fluorescence intensity, presumably due to IPTG toxicity (Figure 10b).

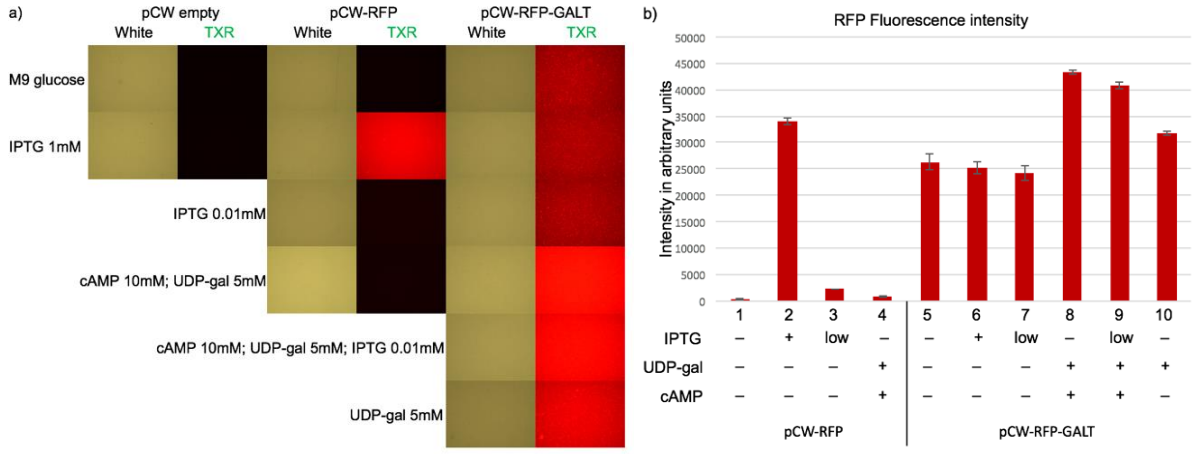


Figure 10. a) Fluorescence microscopy of *E. coli* BL21 cells harbouring pCW, pCW-RFP, and pCW-RFP-GALT plasmids following growth in 0.4% glucose M9 medium. **b)** Fluorescence intensity (excitation = 587 nm, emission = 610 nm) of *E. coli* BL21 cells harbouring pCW-RFP and pCW-RFP-GALT plasmids following growth overnight in M9 medium. *E. coli* BL21 cells transformed with empty pCW-LIC vector were used as a control. Confidence intervals were calculated based on ten measurements for each sample with significance 0.05.

This result shows that a sufficient amount of lactose was produced by β-1,4-galactosyltransferase and transformed into allolactose by chromosomally encoded β-galactosidase, even at the initial basal level of expression when all lactose LacO sites are constantly occupied by lactose repressor proteins. This means that addition of IPTG at any concentration would not improve gene expression, and could even indirectly decrease it due to toxicity of IPTG at high concentrations.

For additional control of expression in *E. coli* BL21 cells harbouring pCW-RFP-GALT, culturing was repeated in M9 medium containing 1% glycerol as the only carbon source. However, a strong fluorescence signal was observed without any induction in cultures harbouring the pCW-RFP-GALT plasmid (Figure 11). This could be because the gluconeogenesis pathway is always active in *E. coli* cells, and glucose is constantly being produced from any carbon source[83].

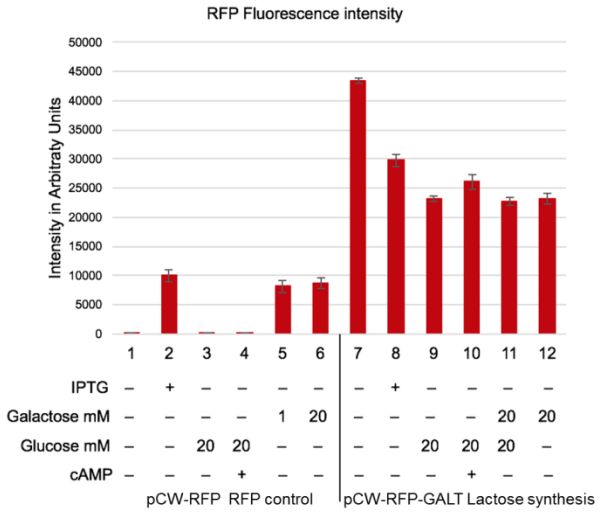


Figure 11. Fluorescence intensity (excitation = 587 nm, emission = 610 nm) of *E. coli* BL21 cells harbouring pCW-RFP and pCW-RFP-GALT plasmids following culturing overnight in M9 medium containing 1% glycerol. BL 21 *E. coli* BL21 cells harbouring empty pCW-LIC vector were used as a control. Confidence intervals were calculated based on ten measurements for each sample with significance 0.05.

This result appeared to be inconsistent with existing data from Chen et al.[82] showing that IPTG addition increased the lactose level. However, their experiment did not provide enough information about the expression level of β-1,4-galactosyltransferase and other recombinant genes. One possible explanation could be that addition of IPTG somehow changes the dynamics of lactose metabolism in the cells. The epimerase in the work of Chen and colleagues transforms UDP sugars, and a two-plasmid system was used. By contrast, the autoinduction system employed in the present project is different and simpler because only one enzyme (B4GALT1) is involved.

During the studies of the lactose-inducible B4GALT1 basal amplification vector pCW-RFP-GALT it was observed expression of β-1,4-galactosyltransferase 1 gene in *E. coli* resulted in the constant induction of all Lac-inducible promoters inside the same cell, even if the B4GALT1 gene was under the control of a Lac-inducible promoter itself (Figure 10). It could be argued that this autoinduction is caused by promoter leakiness, but this explanation is not sufficient, and all promoters are leaky to some extent. It can be argued that this autoinduction is not only due to promoter leakiness, but rather the dynamic nature of the induced transcription-translation process. Even the strongest and least leaky promoter can have some mRNA translated from it due to the low probability that RNAP will initiate transcription from a promoter when the operator site is not occupied by a regulator protein. When this basal transcription occurs frequently, a promoter is considered leaky, but basal transcription happens even for non-leaky promoters; the probability could be calculated by modelling based on promoter strength, transcription initiation speed, operator strength, RNA polymerase concentration, and repressor protein concentration[84]. However, when the synthetic protein gene for the inducer is expressed under the control of the inducible promoter, the possible basal inducer concentration should be added to the model. This additional exponential increase of the inducer concertation has the potential to amplify promoter induction. This leakiness factor is usually neglected in biosynthetic

circuit modelling, but is actually very important, because high leakiness lowers expression noise[85].

In summary, lactose autoinduction loop system based on pCW-RFP-GALT plasmid was designed and tested (Figure 8, 9). It was shown experimentally that lactose is being produced by *E. coli* cells at any conditions independent of IPTG additions (Figures 10, 11).

3.2. Sucrose based sugar amplification system and its benefits

The present project showed that instead of focusing only on different ways to achieve IPTG induction, different sugars can be used as inducers in parallel. Sucrose has many advantages: it is not metabolised by nearly half of *E. coli* strains, including BL21, XL-10 gold, and many others bioengineering strains [56]. There is also still a lack of diversity in synthetic circuit modules, and a sucrose module will likely find wide use in synthetic biology [40]. Finally, sucrose is one of the cheapest oligosaccharides. According to Sigma Aldrich (www.sigmaaldrich.com), 1 gram of sucrose is nearly 500 times cheaper than 1 gram of IPTG – the most expensive component of most current *E. coli* expression systems[86].

It is common knowledge that plants, and cyanobacteria are the main manufacturers of organic matter[87], while other bacteria and animals are the main metabolisers of plant products. There are many plant enzymes for sucrose synthesis that could be useful for the design of sucrose synthetic circuit. For example, fructose phosphate synthase, first identified in spinach leaves, is efficient at synthesising sucrose[88]. If this enzyme is inserted in *E. coli*, it is possible to accumulate sucrose for further polysaccharide synthesis, but sucrose synthesis by *E. coli* involving this cyanobacterial enzyme is quite inefficient; the intracellular concentration of sucrose accumulated only reaches 7 mM[89].

E. coli is the most studied model for biosynthetic circuits, and construction of a sugar amplification system using sucrose as an initial disaccharide is a logical starting point. Moreover, the sucrose transcription regulation system in *E. coli* is poorly understood. Even though the sequences of the Csc operator site and the CscR gene are known, there is no experimental evidence that sucrose binds the CscR protein and induces transcription from the Csc operon[31]. This project aimed to address this while expanding the list of available biosynthetic circuit modules. It may then be possible to add additional sugar blocks to sucrose and amplify oligosaccharides of the interest inside cells. Further, an initial disaccharide synthesis step could serve as a starting point for amplification of many oligosaccharides containing this disaccharide (Figure 12).

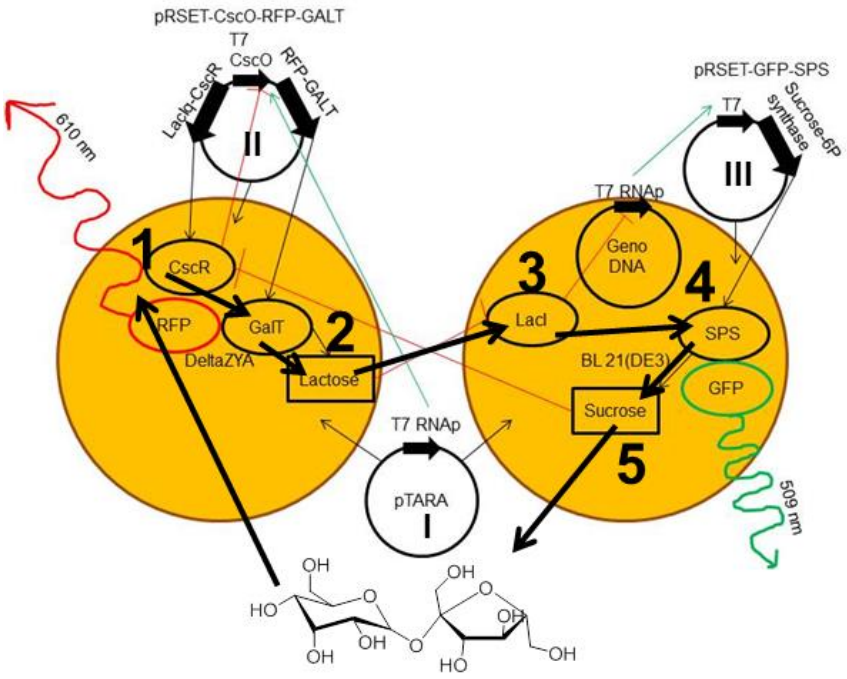


Figure 12. Experimental plan for a sucrose amplification system using the *E. coli* Csc operon and *Synechocystis sp.* SPS gene. Two different *E. coli* strains are transformed with T7 lactose and sucrose-inducible vectors carrying sucrose phosphate synthase and B4GALT1 genes, respectively. 1 - Sucrose binds CscR inducing the expression of GALT, 2 - Lactose is synthesized by GALT, 3 - Lactose diffuses to BL 21 (DE3) cells and binds LacI inducing the expression of SPS, 4 - Sucrose is synthesized by SPS and diffuses to DeltaZYA cells.

Sugar metabolism inside living cells represents a particular challenge in the design of the system. For example, glucose is involved in most metabolic pathways inside *E. coli.* Moreover, glucose is constantly produced by *E. coli* from all carbon sources during gluconeogenesis[90]. Thus, using a disaccharide not containing glucose would be ideal, such as lactulose, a synthetic sugar composed of galactose and fructose. Unfortunately, lactulose is metabolised by *E. coli*[91], hence sucrose is probably a better sugar to initiate amplification because it is naturally not metabolised by *E. coli*.

Two-cell systems provide additional compartmentalisation, reducing autoinduction of separate operons that could come from small amounts of basal translated compounds. For example, it was shown experimentally that β-1,4-galactosyltransferase 1 is always present in cells in small amounts due to inevitable basal expression. As a result, cells with the B4GALT1 gene would likely contain some lactose. If these cells also contain a lactose-inducible element together with B4GALT1, this element would be constantly autoinduced (Figure 10). However, when lactose is made in different cells, a low initial lactose concentration becomes important because it takes significant time for lactose to travel through membranes, diffuse through the media, enter the lactose-inducible cells, and trigger induction.

This project aimed to demonstrate that strains regulated by lactose and sucrose together can be used to make complex synthetic systems. This idea is also based on the diversity of RNA polymerases which could be supplied to bacteria to perform promoter-specific transcription[92]. In addition to natural bacterial RNAPs, there are several viral polymerases, such as T7, which has a set of very specific promoters that are very different form natural bacterial RNAP promoters[36].

Thus, a system using four different plasmids sharing the same antibiotic resistance and two sucrose-deficient *E. coli* strains (deltaLacZYA[93] and BL21 DE3) were constructed. In this system, deltaLacZYA strain serves as an inducer producer, and BL21 DE3 as a sugar

synthesiser. Moreover, the inducer producer is regulated by the sucrose produced by the synthetic strain. This allows both amplification of sucrose, and measurement of its initial concentration, and can be termed ‘sugar amplification’.

The most limiting step in this system is the construction of a new PT7cscO sucrose inducible plasmid. Thus, the pRSET vector containing the T7 promoter and terminator sites was used. The Csc operator site was added directly after the T7 promoter, in line with a similar study on the T7 lac operator[94] (table 3).

T7LacO promoter sequence[94]

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCC

Constructed T7CscO promoter

TAATACGACTCACTATAGGTTAACGTTAAC

Table 3. A comparison of the existing T7LacO promoter and the T7CscO promoter constructed in this project.

However, a CscR repressor protein for this operator is still required, and it was placed under the control of the LacIq (plasmid 6) promoter and terminator, as for LacI in the pCW-LIC vector. The Csc operon is less studied than the Lac operon, and although the sequence of the Csc operator is known, there are no studies on the interactions between CscR and CscO. There is also no X-ray crystallography or NMR structure of the CscR protein available. In theory, a sucrose-inducible *E. coli* plasmid could be used for the simple and cheap inducible expression of genes in *E. coli* because sucrose is the cheapest disaccharide, and is not consumed by *E. coli*, unlike the more expensive lactose. Less than 1 mM IPTG is required to induce the lac operon, compared with nearly 6 mM sucrose to induce the Csc operon[24].

Thus, the insight on lactose synthesis inside *E. coli* was combined by using the pCW-RFP-GALT plasmid in the sucrose-inducible recombinant system. If sucrose is added to such a system, it is likely to remain present at a constant concentration, similar to when a constant concentration of IPTG is added to Lac-regulated transformants. The current results showed the successful expression of RFP from a modified pRSET vector induced by sucrose at concentrations ranging between 10 and 100 mM.

The pRSET vector was selected for the construction of sucrose regulated plasmid because it contains an efficient T7 RNAP promoter for recombinant gene transcription, and a Lac operator site was successfully introduced directly after the T7 promoter sequence, along with a reverse LacI sequence as a lactose-inducible modification. The same strategy was employed, but with Csc operator site instead, and the resulting pRSET(CscO)-RFP-GATL_CscR vector was successfully designed and tested (Figure 13).

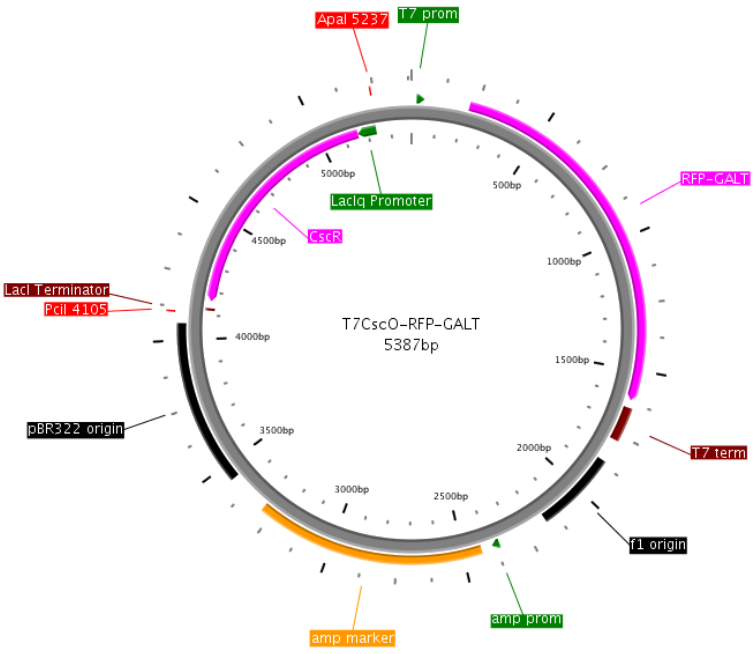


Figure 13. Plasmid map of the pRSET-RFP-GALT_CscR construct. The gene encoding the RFP-GALT fusion protein from the pCW-RFP-GALT plasmid is used in this construct. The CscO operator site was added by site-directed mutagenesis (SDM) directly after the T7 promoter sequence to make the sucrose-inducible recombinant plasmid. For the full sequences, see the Appendices.

There are no commercially available vectors containing the CscR gene. For example, the pCSCX (Addgene) sucrose utilisation phenotype plasmid for *E. coli* does not contain the regulatory gene[25]. Preliminary, this plasmid was also used in this project to amplify sucrose permease (CscB) for sucrose transport studies. Only three different CscR sequences were found in the NCBI gene database; from *E. coli*, from *Shigella flexneri*[95], and from *Yersinia pestis*[96]. To further simplify the plasmid construction steps, the CscR gene was synthesised in reverse and placed between the LacIq promoter and terminator sites. The stronger LacIq promoter is a modified version of the Lac promoter that is used in IPTG inducible expression vectors[97].

As a result, the CscR gene was synthesized and placed in reverse upstream of the T7 promoter in the pRSET expression vector. The regulatory protein sequence is usually placed in expression vectors under its own promoter and terminator, and cells transformed with such vectors constantly express the regulatory protein gene from each copy of the plasmid, thereby achieving efficient regulation. Following the existing examples of LacI-regulated vectors, and the knowledge that CscR is homologous to LacI, the sucrose-inducible pRSET vector was constructed (Figure 14).

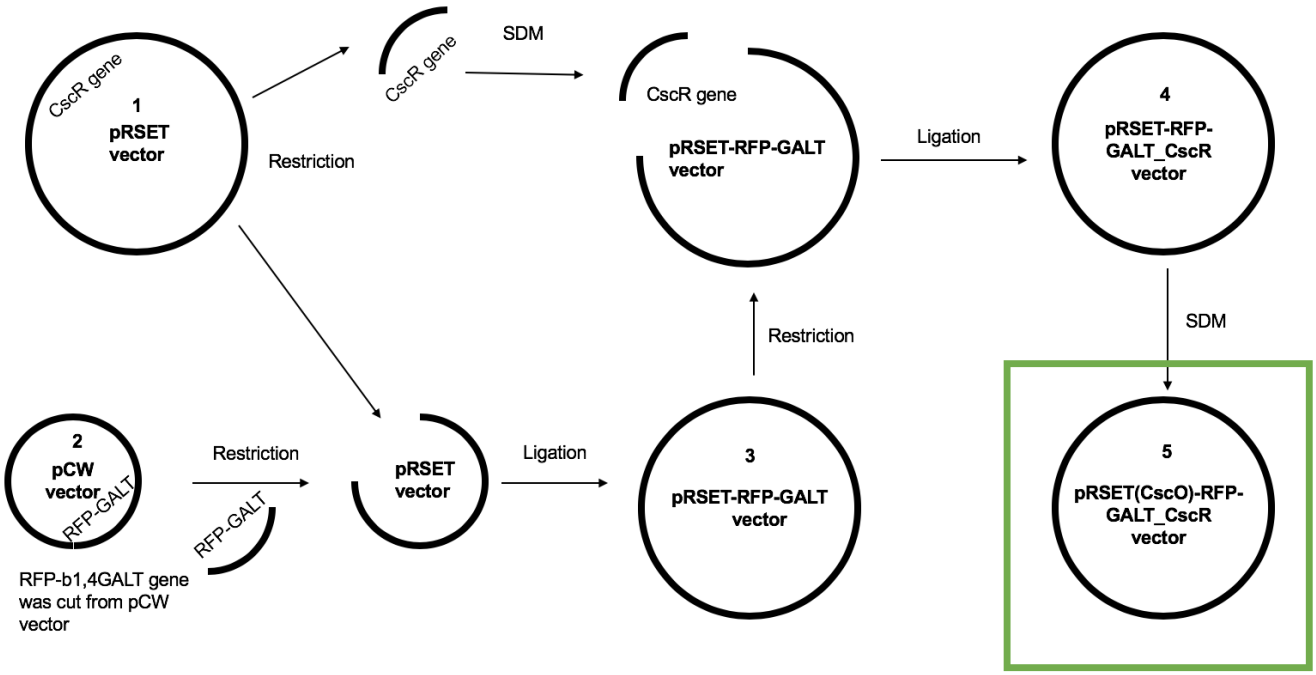


Figure 14. Construction of the sucrose-inducible plasmid. 1) The CscR gene was inserted into the pRSET vector by GeneArt. 2) The RFP-GALT gene from the pCW-RFP-GALT plasmid was inserted into the empty pRSET vector. 3) The CscR gene with new restriction sites added by SDM was inserted into the pRSET-RFP-GALT vector. 4) The CscO operator site was added by SDM to the pRSET-RFP-GALT_CscR gene directly after the T7 promoter. 5) The final pRSET(CscO)-RFP-GALT_CscR sucrose-inducible plasmid.

For future design of a lactose-sucrose oscillator strain of *E. coli*, silencing of the LacZ gene is required to prevent autoinduction inside cells responsible for lactose synthesis. The LacZ gene is responsible for conversion of intracellular lactose to the biological inducer allolactose. The *E. coli* strain delta-ZYA MG1655 in which the entire Lac operon is deleted from the chromosomal DNA was used. However, delta-ZYA does not contain the T7 RNAP gene within the chromosomal DNA. As a result, this strain was co-transformed with two plasmids: the new sucrose-inducible plasmid described above, and a T7 RNAP gene-expressing plasmid. The pTARA arabinose-inducible vector contains a fully functional T7 RNAP gene, and is compatible with the sucrose-inducible pRSET vector. Because the MG1655 strain is not competent for

transformation, it was prepared for electroporation by washing and freezing in ultrapure water. The delta-ZYA strain was subsequently successfully transformed via electroporation with the sucrose-inducible pRSET(CscO)-RFP-GALT_CscR (pRSET(CscO)) plasmid and pTARA containing the T7 RNAP gene. This strain was later used as a control for autoinduction tests because it is completely unregulated by lactose.

To demonstrate a role of B4GALT1 in autoinduction of T7 RNAP expression in *E. coli* BL21 (DE3) cells, the BL21 (DE3) Tuner strain (Novagen) was employed in which the LacZYA operon is deleted, making it more sensitive to inducer concentration. Because the natural inducer of the Lac operon is allolactose, which is converted from lactose by LacZ, any induction by lactose could be prevented if the LacZ gene is silenced. Both the pCW-RFP-GALT and pRSET-RFP-GALT plasmids are subject to autoinduction by lactose produced by B4GALT1. As a result, addition of IPTG did not improve expression of RFP in transformed *E. coli* BL21 (DE3) cells. However, when BL21 (DE3) Tuner cells were transformed with the same plasmids, addition of IPTG significantly increased expression of RFP (Figure 15). This result also showed that addition of both IPTG (or allolactose) and sucrose is required to induce gene expression from the pRSET(CscO) vector.

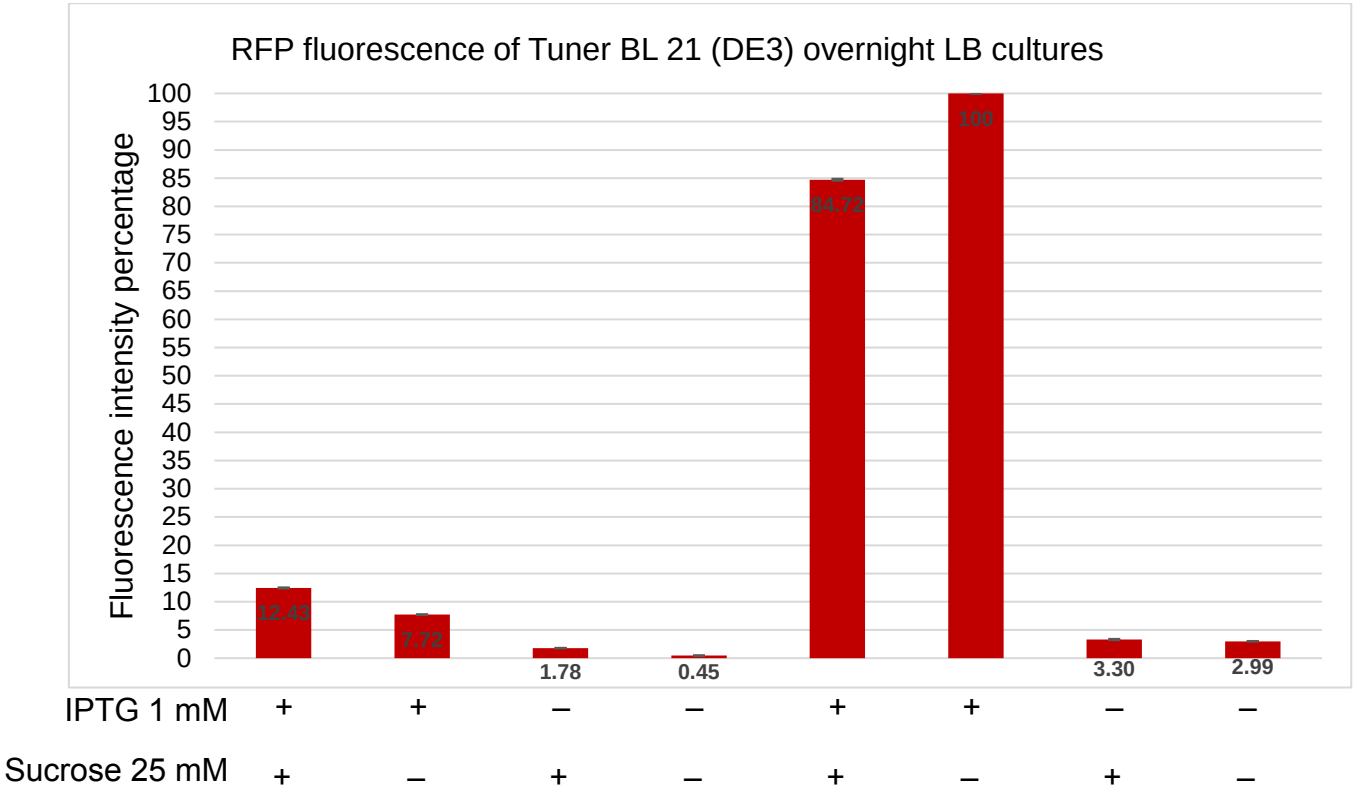


Figure 15. The role of IPTG induction in *E. coli* BL21 (DE3) Tuner cells transformed with sucrose regulated pRSET(CscO)-RFP-GALT_CscR (1,2,3,4) and non-regulated by sucrose pRSET-RFP-GALT_CscR (5,6,7,8) plasmids. Confidence intervals were calculated based on ten measurements for each sample with significance 0.05.

The amount of RFP produced in BL21 (DE3) cells transformed with different pRSET-RFP-GALT vectors was so high that RFP could be easily detected by visual observation of colonies in daylight (Figure 16). This simple method was also used to compare sucrose-induced and non-induced colonies transformed by sucrose-inducible plasmids. It was found that sucrose-inducible cells do not express as much RFP as cells without sucrose regulation, even in the presence of 100 mM sucrose in agar. However, the visual colour difference between induced and non-induced colonies was still detectable (Figure 17). The actual fluorescence of RFP was detected using appropriate excitation and emission filters.

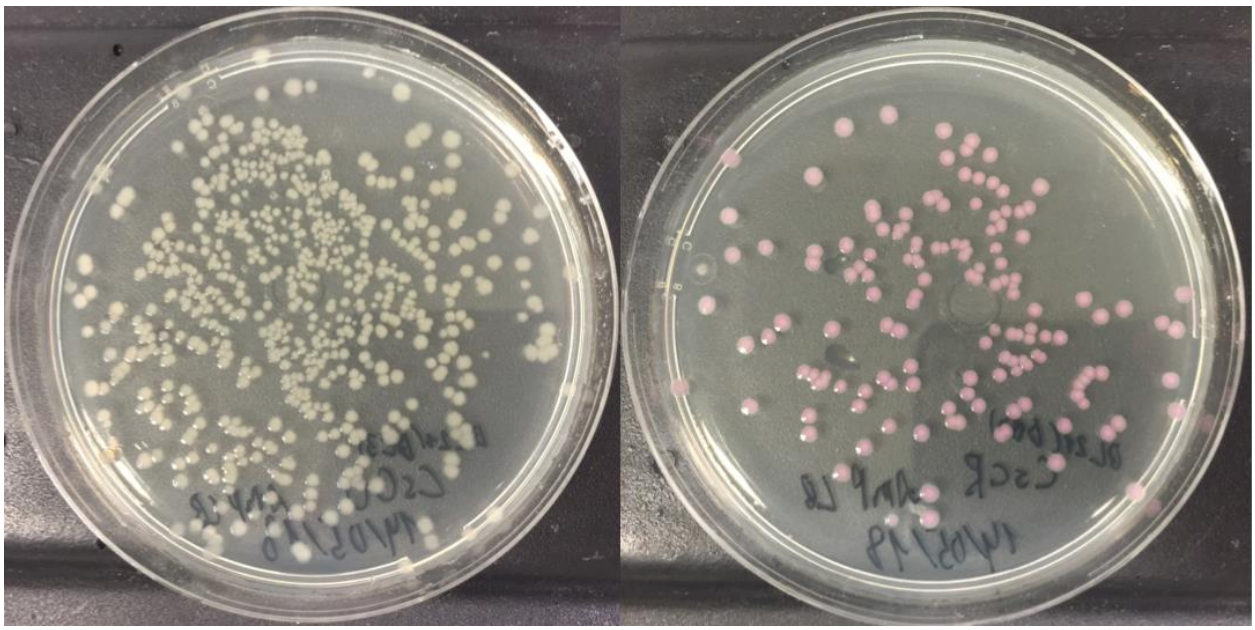


Figure 16. LB ampicillin agar plates 3 days after inoculating with BL21 (DE3) cells harbouring pRSET(CscO)-RFP-GALT_CscR (left) or pRSET-RFP-GALT_CscR (right) plasmids.

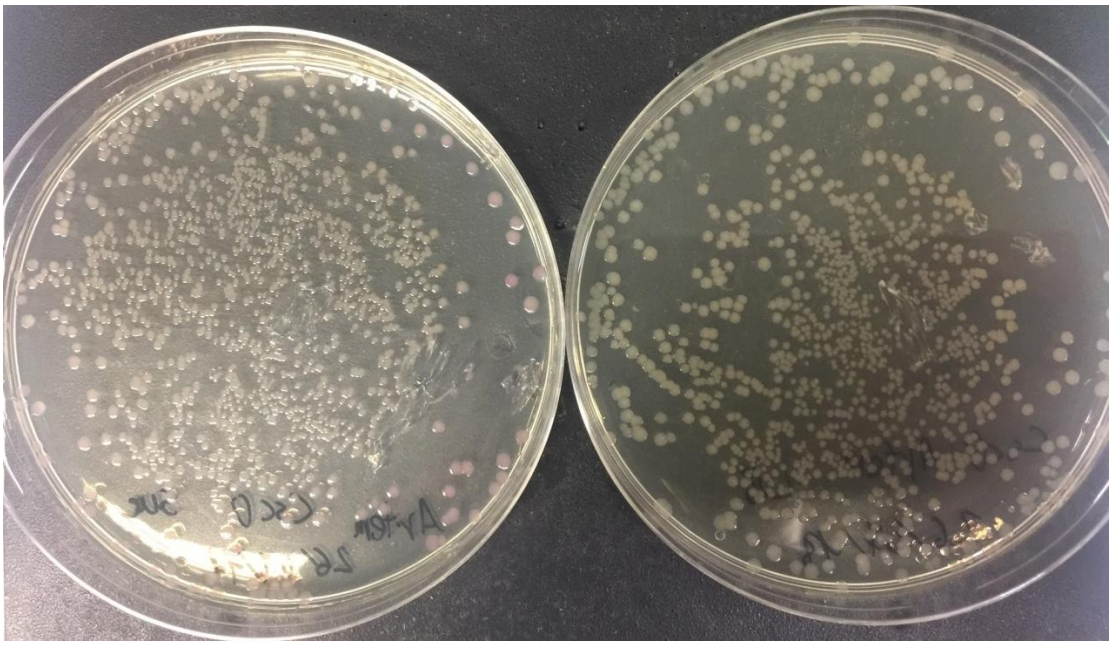


Figure 17. LB ampicillin agar plates 7 days after inoculating with BL21 (DE3) cells harbouring the pRSET(CscO)-RFP-GALT_CscR plasmid, with 100 mM sucrose added to the plate on the left.

Following these visual experiments, colonies were placed under light close to the excitation wavelength of RFP, and observing under a red light filter. Agar plates containing sucrose-inducible colonies were placed under a fluorescence microscope, and the fluorescence intensity of colonies with and without sucrose were compared (Figure 18).

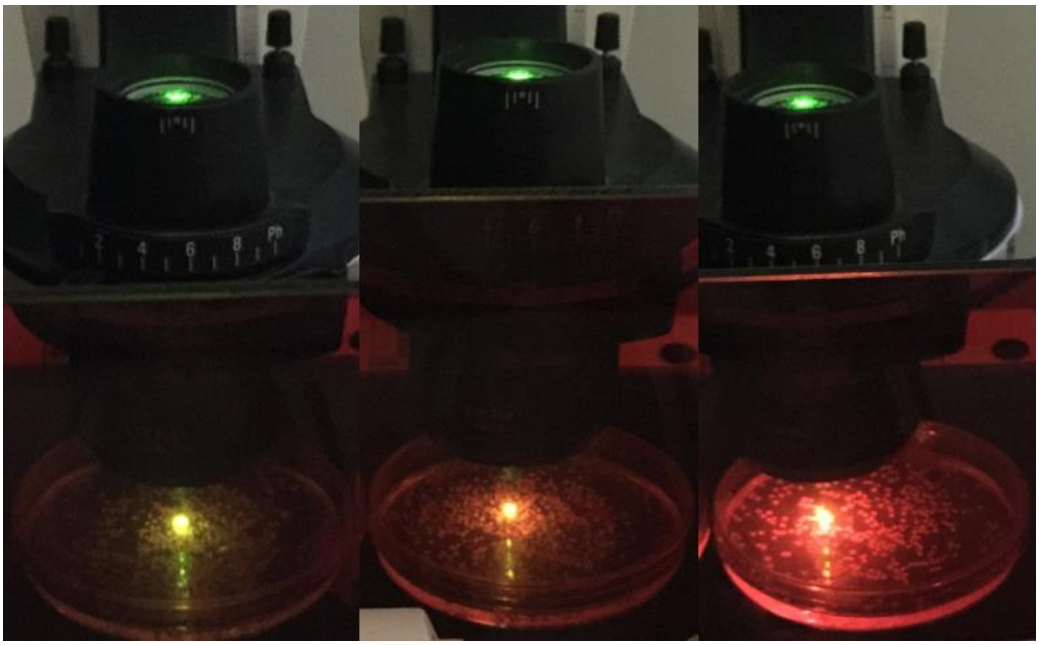


Figure 18. Red fluorescence comparison of different colonies under a fluorescence microscope following excitation. On the left is an LB ampicillin agar plate with colonies harbouring the sucrose-inducible pRSET plasmid. In the middle is an LB-ampicillin agar plate with colonies harbouring the sucrose-inducible pRSET plasmid, but with 100 mM sucrose added. On the right is an LB ampicillin agar plate with colonies harbouring the non-sucrose-regulated pRSET plasmid.

In summary, the novel sucrose inducible recombinant plasmid (pRSET(CscO)-RFP-GALT_CscR) based on pRSET vector was constructed (Figures 13,14). An increase in RFP fluorescence of pRSET(CscO) transformed BL 21 (DE3) cells growing in presence of sucrose was observed, indicating functionality of the system (Figures 15,16,17,18).

3.3. Sucrose transport into *E. coli* cells

The new sucrose-inducible plasmid allowed measurement of sucrose transport into *E. coli* cells because a specific intracellular concentration of sucrose is required to bind all CscR protein molecules. Such experiments allow the identification of the minimal necessary sucrose concentration for plasmid induction. A liquid LB culture of BL21 (DE3) cells transformed with the pRSET(CscO)-RFP-GALT_CscR plasmid was incubated for 6 h and induced with different concentrations of sucrose. The fluorescence intensity was measured after 3 h of induction (Figure 19).

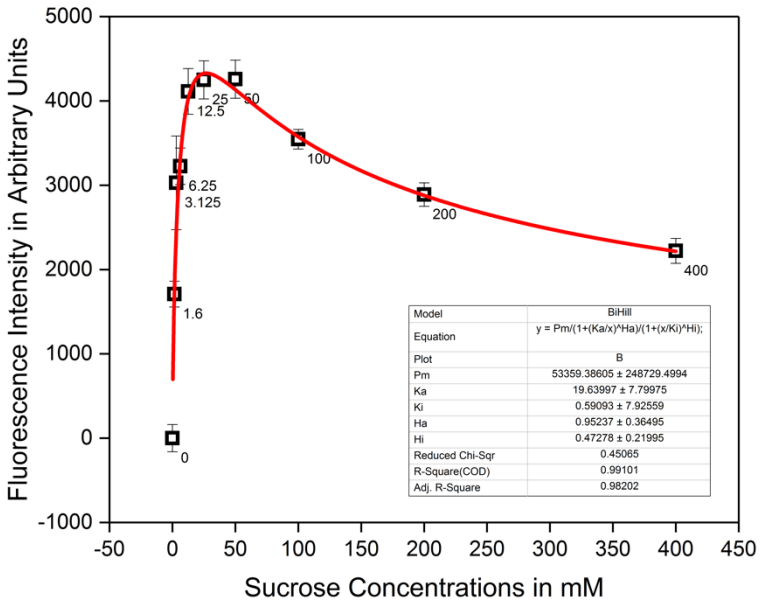


Figure 19. RFP fluorescence intensity curve using different sucrose concentrations ranging from 0 to 400 mM. *E. coli* BL21 (DE3) cells were grown in LB for 6 h, and induced for 3 h. The maximum fluorescence intensity was observed at ~25 mM sucrose. Confidence intervals were calculated based on ten measurements for each sample with significance 0.05.

The gradual decrease in fluorescence intensity above 50 mM sucrose in BL 21 (DE3) cells (Figure 19) could possibly be explained by plasmolysis due to a high sucrose concentration inside cells[98]. However, this decrease in fluorescence was not observed for *E. coli* BL21 (DE3) Tuner cells, in which maximal fluorescence was reached at 1.6 mM sucrose. As Tuner cells require addition of IPTG for the recombinant gene expression, the tight regulation could help cell viability. On the other hand, BL 21 (DE3) cells experience autoinduction before the actual induction by addition of sucrose.

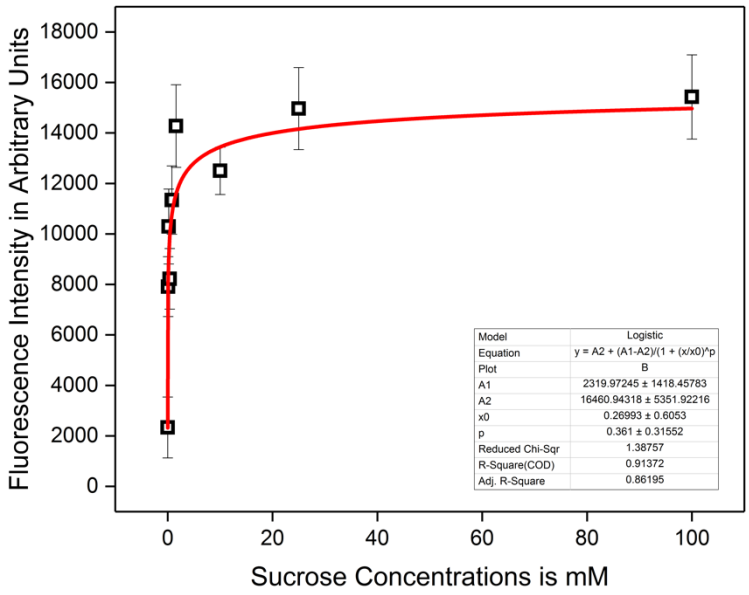


Figure 20. RFP fluorescence intensity curve at different sucrose concentrations ranging from 0 to 100 mM. *E. coli* BL21 (DE3) Tuner cells were grown in LB for 6 h, and induced for 3 h. The maximal fluorescence intensity was observed at ~1.6 mM sucrose (1 mM IPTG was added to each culture). Confidence intervals were calculated based on ten measurements for each sample with significance 0.05.

Some weak fluorescence was detected without the addition of sucrose, but this could be because clear LB medium was used as a control. A significant increase in fluorescence was detected even after addition of 0.1 mM sucrose (Figure 21).

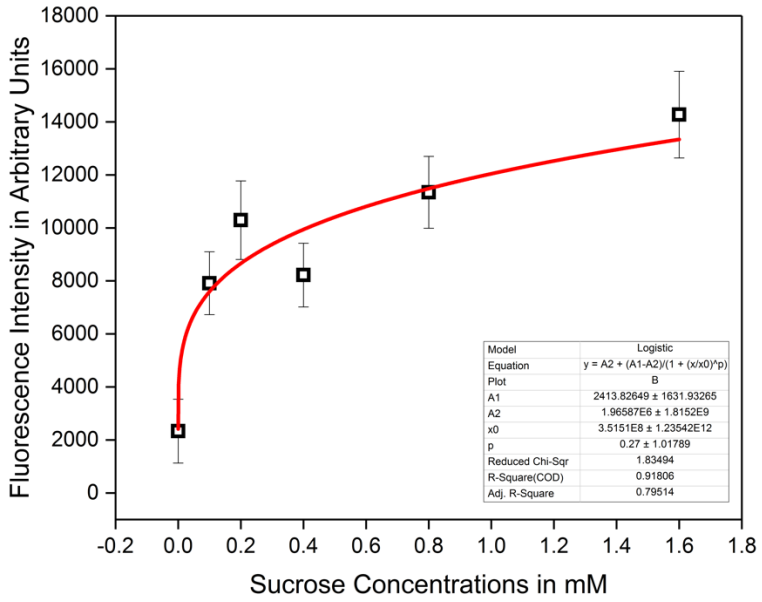


Figure 21. RFP fluorescence intensity curve at 0–1.6 mM sucrose. Confidence intervals were calculated based on ten measurements for each sample with significance 0.05.

E. coli BL21 (DE 3) cells grew very poorly in M9 media and did not adapt to these conditions (see Appendix). As a result, sucrose induction in M9 minimal media was not studied.

In summary, the pRSET(CscO) sucrose regulated plasmid was used to test sucrose transport efficiency into *E. coli* cells. It was found experimentally, that the sucrose transport rate increases with the sucrose concentration and that about 1 mM of sucrose medium is probably enough to induce expression from pRSET(CscO) plasmid. However, this result was unexpected and is contradictory to a literature report[31].

4. CHAPTER IV. Conclusions

Herein, an autoinduction loop with a biosynthetic enzyme synthesising the inducer of its own gene expression under the control of its product was investigated. B4GALT1 served as a model enzyme, and the inducer was lactose. Lactose autoinduction was so efficient that addition of IPTG did not improve the gene expression level of RFP-GALT fusion proteins. Indeed, IPTG addition even decreased the expression level. The results showed that this type of system could be used to construct synthetic oscillators for sensing and amplification of specific sugars. Furthermore, the results imply that autoinduction does not depend solely on promoter strength and can be observed even using the least leaky promoters.

Successful lactose amplification and autoinduction of lactose operon is a proof-of-principle result showing that sugars could be efficiently amplified from zero concentrations. This result suggests that glycosyltransferases could potentially be used to amplify oligosaccharides in a PCR like manner. The next step in this project is to attempt amplification of longer oligosaccharides, for example *N*-linked glycans.

Additionally, a novel sucrose-regulated recombinant plasmid for *E. coli* was constructed and tested. Sucrose induction was not as efficient as IPTG induction. However, sucrose is cheap, and less toxic than IPTG. In addition, poor transport of sucrose could be beneficial for more sensitive regulation.

The first part of the lactose-sucrose synthetic oscillator based on sucrose-inducible lactose synthesis was constructed and tested. The amount of lactose produced by this system was sufficient for induction of Lac-regulated promoters, but the sucrose concentration required for induction of lactose synthesis was significantly higher, and the efficiency of sucrose induction was lower.

The sucrose-inducible plasmid tests indicated that sucrose transport in *E. coli* BL21 cells is inefficient. However, the intracellular sucrose concentration from the diffusion processes was apparently sufficient to induce recombinant protein gene expression. Thus, sucrose transport in *E. coli* appears to be related to the sucrose concentration in medium.

4.1. Future perspectives and applications

A sucrose induction plasmid was successfully established and tested. Thus, an obvious next step is the construction of a sucrose phosphate synthase gene expression plasmid. A sucrose-lactose synthetic oscillator for sucrose concentration sensing and amplification could then be constructed, which could be a useful model based on the proof-of-principle of sugar sensing and amplification (Figure 21).

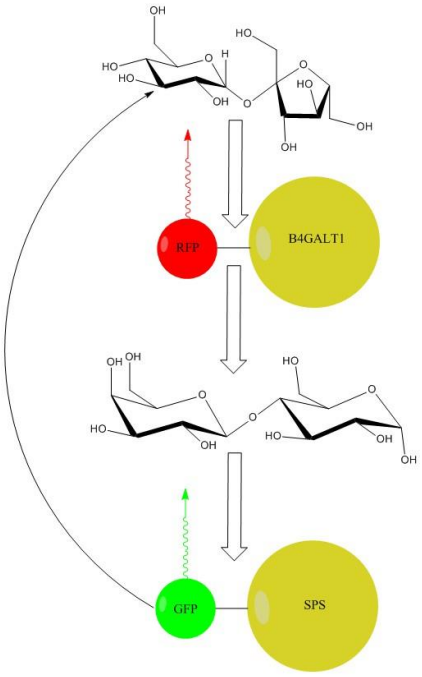


Figure 21. A model sucrose-lactose oscillator. Sucrose induces the expression of the RFP-GaLT gene, and B4GALT1 synthesises lactose. Allolactose induces the expression of the SPS gene, and SPS synthesises sucrose. The two biosynthetic parts are separated in different compartments to prevent immediate induction of the synthesis of one product by the other product.

Intracellular sucrose concentration measurements could be monitored by NMR. To do this cultures of *E. coli* cells harbouring a sucrose inducible plasmid should be incubated with sucrose for several hours. After appropriate washing procedure, ^{1}H NMR spectra of cell lysate should provide insight about sucrose transport into the cells.

A T7 sucrose-inducible strain could be designed by analogy to the existing BL 21 (DE3) strain with a T7 RNA gene inserted in the chromosomal DNA the control of a promoter with a CscO site. This strain could be beneficial for large volume recombinant gene expression because sucrose is a cheaper inducer than IPTG.

In principle, any sucrose-inducible vector based on bacterial RNAP could be constructed similarly to the pRSET(CscO) vector in this project.

Specific sugar sensing libraries could be used in medicine for diagnostic purposes. For example, a glycation test could be made. While oligosaccharides from the healthy glycosylation pattern would be detected and amplified, any abnormal glycosylation would not be detected. The ketoamine glycosylation sensing system could also be added to analyse the glycation rate of haemoglobins and albumins in blood. This method could be beneficial in comparison with direct blood sugar tests as glycosylated proteins represent long term information, because haemoglobin in blood degrades only after 120 days[99]. This could be used as a simple and cheap test for glycosylation in the case of diabetes.

5. Experimental details

5.1. General protocols

All experiments involving bacterial cultures were conducted in a Herasafe Kendro KS12 Laminar flow hood. Bacterial media components and equipment such as tips were sterilised in an LTE Touchclave-R 40H autoclave prior to use. All bacterial liquid cultures were incubated in a New Brunswick Scientific G25 shaker at 37ºC with shaking at 250 rpm. All agar plates were incubated in a Heraeus D-6450 Hanau incubator at 37ºC. If required, agar plates and liquid cultures were stored at 4ºC in s cold room. All primers were ordered from Sigma-Aldrich (St. Louis, Missouri, United States) or Invitrogen (Carlsbad, California, United States). All restriction enzymes were ordered from New England Biolabs (Ipswich, Massachusetts, United States) or Clontech TAKARA BIO INC. (Kusatsu, Shiga Prefecture, Japan). All plasmids were ordered from Addgene (Cambridge, Massachusetts, United States).

5.2. Glycerol stocks

500 µl of autoclaved 80% glycerol was added to 500 µl LB bacterial culture, mixed, and centrifuged to homogenise. All glycerol stocks were stored in a -80ºC freezer.

5.3. Plasmid extraction from *E. coli* cells

Lysogeny broth agar (400 ml ddH_2O, 10 g granulated LB, 6 g bacteriological agar powder) was prepared and autoclaved. If required, 400 µl of antibiotic (ampicillin) was added from a 100 mg/ml stock in warm LB agar (40–50ºC). Twenty plastic agar plates were filled with liquid LB agar (~20 ml of agar in each), incubated in a laminar for 1 h for freezing, and stored at 4ºC in a cold room. Plasmid-containing bacteria from agar slabs were transferred to prepared agar plates

with a plastic bacterial coil and incubated at 37ºC overnight (~16 h). Plates with single colonies were stored at 4ºC. LB broth liquid medium (400 ml ddH_20, 10 g granulated LB) was prepared and autoclaved, and where required, 400 µl of warm LB medium (40–50ºC) was added from a 100 mg/ml ampicillin stock. A single cloning colony was transferred to 10 ml LB medium into a 50 ml tube using a 'yellow' tip from the agar plate. Liquid cultures were incubated overnight at 37ºC with shaking at 220 rpm.

DNA was extracted using a QiaSpin miniprep kit 27104. Liquid cultures were centrifuged for 3 min at 7850 rpm. The supernatant was discarded, the cell pellet was suspended in 250 µl of ice-cold P1 buffer, and 250 µl of P2 buffer was added. The cell suspension was swirled and incubated for 4 min at room temperature. Immediately after this, 350 µl of N3 buffer was added and the sample was applied to a 1.5 ml microcentrifuge tube and centrifuged for 10 min at 13,000 rpm. The supernatant was applied to a blue Qiaprep spin column at a volume of 750 µl per column. Columns were centrifuged for 1 min at 13,300 rpm, the flown-trough was removed, and 500 µl of buffer PB was added to the column for an additional wash. Columns were centrifuged for 1 min at 13,300 rpm, the flow-through was again removed, and 750 µl of buffer PE (with ethanol) was added. Columns were centrifuged for 1 min at 13,300 rpm, the flown-through was discarded, and the column was centrifuged for 1 min at 13,300 rpm, and this was repeated one more time. Columns were then transferred to 1.5 ml microcentrifuge tubes and 33 µl of preheated (50ºC) ddH_2O was added to the middle of each column and incubated for 5 min at room temperature. After centrifuging for 2 min at 13,300 rpm, columns were removed, microcentrifuge tubes were mixed by vortexing and centrifuged using a table-top centrifuge. The DNA concentration in ng/µl was measured using a Nanodrop spectrophotometer at 260 nm.

5.4. Bacterial transformation of *E. coli* BL21, BL21 (DE3) and BL21 (DE3) Tuner competent cells

Super Optimal broth with Catabolite repression (SOC) medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) was kept at room temperature. *E. coli* BL21 cells from -80ºC stocks were thawed on ice, and 20 µl of thawed cells were aliquoted in pre-chilled 14 ml falcon tubes. Next, ~100 ng of transformation DNA was added to cells, tubes were swirled, incubated on ice for 5 min, cells were heat-shocked at 42ºC for 30 s, then incubated on ice for 2 min. An 80 µl sample of preheated SOC medium was then added and incubated at 37ºC with shaking at 250 rpm for 1 h. Finally, 10 µl of SOC culture was spread on an agar plate with appropriate antibiotic and incubated at 37ºC overnight.

5.5. Bacterial transformation of XL10-Gold ultracompetent cells

SOC medium was kept at 37ºC. XL10-Gold cells from -80ºC stocks were thawed on ice and 45 µl of thawed cells were aliquoted in pre-chilled 14 ml falcon tubes, with 2 µl of β-mercaptoethanol added to each aliquot. Cells were incubated for 10 min on ice with swirling every 2 min. Next, 1 µl (~100 ng) of plasmid or 10 µl of ligation product was added, tubes were swirled, and incubated for 30 min on ice. Cells were heat-shocked at 42ºC for 30 s, then incubated on ice for 2 min. A 450 µl sample of preheated SOC medium was added and incubated at 37ºC with shaking at 250 rpm for 1 h. Transformation mixtures were spread on agar plates and incubated at 37ºC overnight. If no single colonies were obtained, some cells were transferred to a new agar plate with a bacterial coil.

5.6. Bacterial transformation of Stellar competent cells

SOC medium was kept at 42ºC. Stellar cells from -80ºC stocks were thawed on ice, and 50 µl of thawed cells were aliquoted in pre-chilled 14 ml falcon tubes and mixed with 5 µl of In-fusion product, swirled, and incubated for 30 min on ice. After heat-shocking at 42ºC for 45 s, tubes were incubated on ice for 2 min, 445 µl of preheated SOC medium was added, and cultures were incubated at 37ºC with shaking at 250 rpm for 1 h. Transformation mixtures were spread on agar plates and incubated at 37ºC overnight. If no single colonies were obtained, some cells were transferred to a new agar plate with a bacterial coil.

5.7. M9 minimal expression medium preparation

M9 salts stock solution (Na_2HPO_4 34 g, KH_2PO_4 15 g, NaCl 2.5 g, NH_4Cl 5 g per l) was prepared and autoclaved, and 100 ml 1 M $MgSO_4$ (19.8 g $MgSO_4$*H_2O 70%), 100 ml 1 M $CaCl_2$ (11,11 g $CaCl_2$), 100 ml 20% glucose (20 g glucose + 80 ml ddH_2O) or 50% glycerol (glycerol 100 g + ddH_2O 100 ml), and 700 ml dH_2O stocks were similarly prepared and autoclaved. In a laminar hood, 200 ml M9 salts + 2 ml $MgSO_4$ + 20 ml 20% glucose or 20 ml 50% glycerol + 100 µl $CaCl_2$ were mixed with 700 ml dH_2O, and sterile dH_2O was added to 1 l.

5.8. Fluorescence microscopy protocol

A 200 µl sample of each culture was transferred to a Greiner CELLSTAR 96-well transparent plate, and a Leica DMi 8 Fluorescence optical microscope was used for optical fluorescence measurement with a TXR 540-580 nm fluorescence filter; a 20× and 40× zoom, a camera exposure time of 200 ms, and an optical gain of 3. Observation was performed under both white light and light at the excitation wavelength (Figure 5.1).

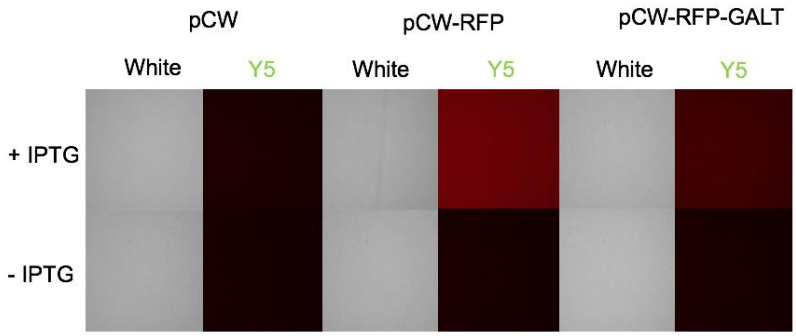


Figure 5.1. Microscopy analysis of the fluorescence signal from different *E. coli* BL21 samples.

5.9. Fluorescence intensity measurement

A 200 µl sample of each culture was transferred to a Greiner CELLSTAR 96-well transparent plate, and a TECAN infinite M 1000 PRO plate reader was used with top/bottom reading, an excitation wavelength of 587 nm, a 5 nm bandwidth, an emission wavelength of 610 nm, a 5 nm bandwidth, gain optimal, a flash frequency of 400 Hz/100 Hz, 50/10 flashes, and an integration time of 20 µs. Cells free growth medium was used as a background.

5.10. Optical density (OD) measurement

OD was measured using a Novaspec III visible spectrophotometer for more than 10 ml culture volumes, or a TECAN infinite M 1000 PRO plate reader for smaller volumes. The absorbance at 600 nm was measured using identical clean medium as a standard. For continuous expression, the OD was measured every 1 or 0.5 h.

5.11. β-1,4-galactosyltransferase 1

The B4GALT1 gene lacking the membrane domain in the pMAL-c2x plasmid[105] (plasmid 11) was provided by Dr James Errey. Specific primers pMalC2x_lac and 100bpBamHI (Supplementary 1) were used to amplify the sequence of B4GALT1 (Supplementary 2) containing part of the plasmid. The B4GALT1 gene was amplified by Q5 Hi-fi DNA polymerase using *Kpn*I R and *Nde*I F primers (Appendix 1).

5.12. DNA agarose gel electrophoresis

Agarose gels (0.8%) were prepared (390 ml ddH_20, 10 ml 50× TAE, 3.2 g agarose) and melted in a microwave oven for 3 min on max power. An 80 ml sample of melted gel was placed in a beaker and 8 µl SYBR safe was added and mixed. After setting at room temperature for 1 h, the gel was placed in an electrophoresis chamber and 500 ml 1× TAE buffer (490 ml ddH_20, 10 ml 50× TAE) was added. A 10 µl 1 kb DNA ladder was added to one lane in each gel, and electrophoresis stock (~100 ng DNA, 1× gel loading dye) was prepared. Gels were run for ~35 min (U = 100 V). If there were products to extract, gels were cut into two parts, and bands for analysis were imaged under UV with an exposure time of 1 s. DNA for extraction was not held under UV longer than 5 min. DNA bands of interest were excised and DNA was extracted using according to the Qiaprep kit protocol. A 3-fold volume of QG buffer (relative to the extracted band) was added and incubated at 50ºC for 15 min, and 1 gel volume of isopropanol was added and mixed. Samples were transferred to an extraction column, centrifuged at 13,000 rpm, 750 µl of buffer PE with ethanol was added, and samples were centrifuged twice at 13,300 rpm for 1 min. Columns were transferred into 1.5 ml Eppendorf tubes, 31 µl of preheated (50ºC) ddH_2O was added and incubated 5 min, columns were centrifuged at 13,300 rpm for 2 min, Eppendorfs were vortexed, and the DNA concentration was measured using a Nanodrop at 260 nm.

5.13. Expression from the pCW-RFP plasmid

Competent *E. coli* BL21 cells were transformed with the pCW-RFP plasmid, and a 10 ml LB liquid culture containing a single transformed colony was incubated overnight (~17 h) at 37ºC with shaking at 250 rpm. Cultures were centrifuged at 7350 rpm for 3 min, the supernatant medium was removed, and the cell pellet was washed with 1 ml M9 minimal medium and centrifuged at 7350 rpm for 3 min. The supernatant was removed and 10 ml M9 medium was added. The cell pellet was mixed until homogeneous and aliquoted into fresh 20 ml M9 medium to achieve an OD of ~0.2. M9 cultures were incubated at 37ºC with shaking at 250 rpm for nearly 5 h, and the OD was measured every 1 or 0.5 h until ~0.6. Three 20 ml M9 cultures were prepared from cells harbouring pCW, pCW-RFP or pCW-RFP-GALT. After reaching OD ~0.6 (Figure 5.2), 2 × 5 ml of each culture was aliquoted into 14 ml falcon tubes and 5 µl 1 M IPTG was added to half of the tubes and incubated for 3 h at 37ºC with shaking at 250 rpm. Droplets from these cultures were exposed under UV 254/345 nm to detect fluorescence. However, nothing was detected, probably because UV light does not excite RFP, which has an excitation wavelength of 587 nm[100]. A 200 ml sample of each culture was transferred to a Greiner 96 black transparent bottom plate and fluorescence was measured using a PHERASTAR plate reader. The 5 ml cultures were then incubated overnight (~18 h) at 37ºC with shaking at 250 rpm, and 200 ml samples were transferred to a Greiner 96 transparent plate and observed under a Leica DMi 8 Fluorescence microscope.

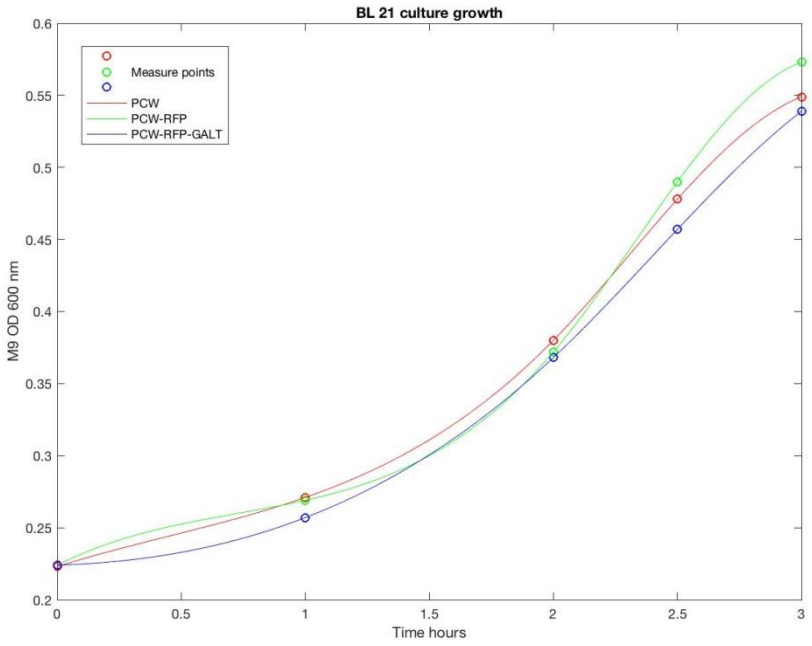


Figure 5.2. *E. coli* cultures in M9 medium were grown at 37ºC with shaking at 250 rpm for 3 h to an OD at 600 nm of ~0.6 (measured vs. clear M9 medium as a blank). The interpolated graph was plotted using MATLAB_R2017a.

5.14. DNA enzymatic restriction

DNA was extracted from *E. coli* or agarose gels and restriction digests were performed at the appropriate temperature[101]. Agarose gels were run to separate DNA fragments, and bands were extracted using a Qiaprep Gel extraction kit. Restriction digests were performed as follows:

1) *Xba*I restriction (T/CTAGA; *Xba*I, 5 µl, 20 u/µl): DNA 10 µg; 10× CutSmart Buffer 5µl (50 mM potassium acetate, 20 mM TRIS-acetate, 10 mM magnesium acetate, 100 µg/ml BSA, pH 7.9, 25°C); ddH_2O to 50 µl. Incubated for 6 h/15 min at 37ºC. This incubation was performed to cut the pRJ-Paph-LacZYA plasmid and the GFP insert for ligation (Figure 5.3).

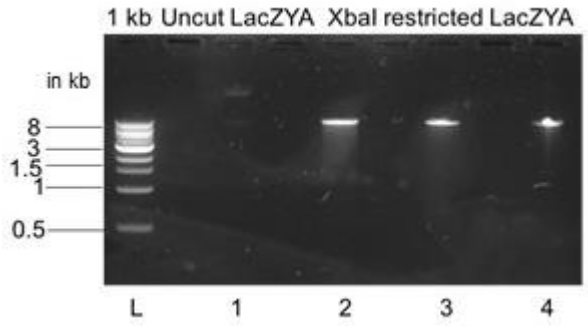


Figure 5.3. 1% agarose gel image of pRJ-Paph-LacZYA plasmid; (1) uncut plasmid; (2,3,4) restricted plasmid.

2) *Apa*I restriction (GGGCC/C; *Apa*I 5 µl, 50 u/µl): DNA 10 µg; 10× CutSmart Buffer 5 µl; ddH_2O to 50 µl. Incubated at 25ºC for 6 h/15 min. This incubation was performed to cut the pRJ-Paph-LacZYA plasmid and GFP insert for ligation (Figure 5.4).

3) *Afl*II restriction (C/TTAAG; *Afl*II 5 µl, 10 u/µl): DNA 10 µg; 10× MBuffer 5 µl (100 mM TRIS-HC, 100 mM MgCl2, 10 mM dithiothreitol, 500 mM NaCl, pH 7.5); 0.1% BSA 5 µl; ddH_2O to 50 µl. Incubated at 37ºC for 16 h. This incubation was performed to cut the pRJ-Paph-LacZYA and LacZ-GFP plasmids and compare their lengths (Figure 5.4).

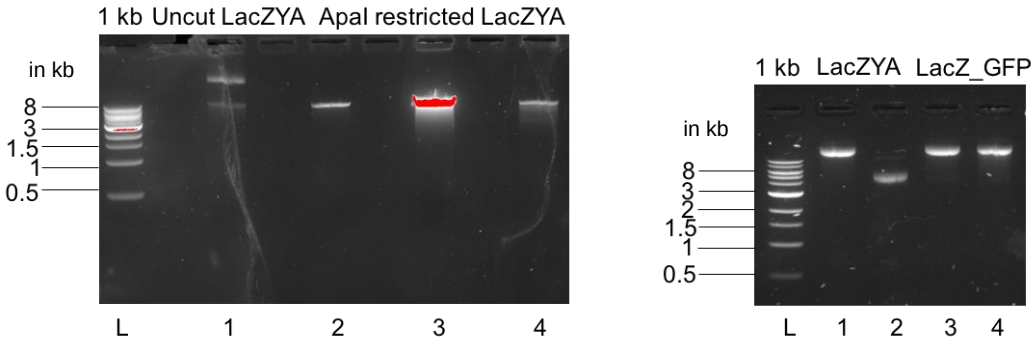


Figure 5.4. 1% agarose gel images of pRJ-Paph-LacZYA plasmid, and LacZ_GFP plasmid; (1 left) uncut plasmid; (2,3,4 left) restricted plasmid; (1 right) uncut plasmid; (2 right) restricted plasmid; (3 right) uncut plasmid; (4 right) restricted plasmid.

4) *Nde*I (CA/TAATG) and *Kpn*I-HF (GGTAC/C) double digestion: *Nde*I 1 µl (20 u/µl); *Kpn*I-HF 1 µl (20 u/µl); DNA 5 µg; 10× CutSmart Buffer 5 µl; ddH_2O to 50 µl. Incubated at 37ºC for 16 h. This incubation was performed to make the pCW-GALT plasmid (Figure 5.5).

5) *Bsr*GI-HF (T/GTACA) and *Bbv*CI (CC/TCAGC) double digestion: *Bsr*GI-HF 5 µl (20 u/µl); *Bbv*CI 5 µl (2 u/µl); DNA 10 µg; 10× CutSmart Buffer 5 µl; ddH_2O to 50 µl. Incubated at 37ºC for 16 h. This incubation was performed to cut the pRJ-Paph-LacZYA plasmid and GFP insert for ligation (Figure 5.5).

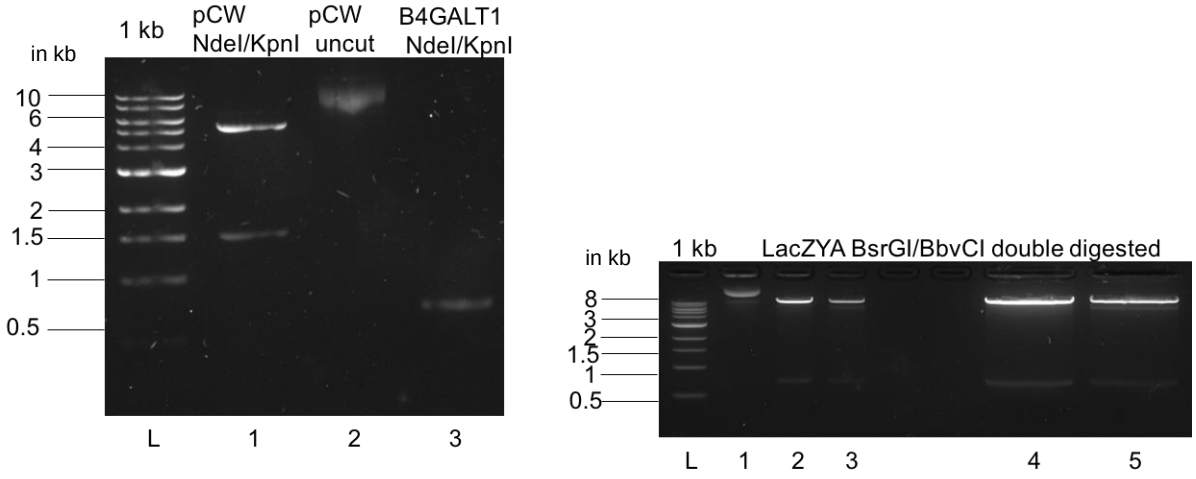


Figure 5.5. 1% agarose gel images of pCW vector, B4GALT1 gene, and pRJ-Paph-LacZYA plasmid; (1 left) restricted vector; (2 left) uncut vector; (3 left) restricted gene; (1,2,3,4,5 right) restricted plasmid.

6) *Bam*HI (G/GATCC) and *Nde*I (CA/TAATG) double digestion: *Bam*HI 1 µl (20 u/µl); *Nde*I 1 µl (20 u/µl); DNA 5 µg; Neb 3.1 Buffer 5 µl (100 mM NaCl, 50 mM TRIS-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA, pH 7.9, 25°C); ddH_2O to 50 µl. Incubated at 37ºC for 16 h (Figure 5.6). This was the best result; the most important plasmids (pCW-RFP and pCW-RFP-GALT) were obtained after restriction and Quick T4 ligation (Table 5.2).

7) *Nco*I-HF (C/GATGG) digestion: *Nco*I-HF 1 µl (20 u/µl); DNA 300 ng; 10× CutSmart Buffer 2 µl, ddH_2O to 20 µl. Incubated at 37ºC for 3 h. The *Nco*I restriction site is only present in RFP and B4GALT1 genes, and absent in the pCW-LIC vector. Thus, plasmids are cut by this enzyme only when these inserts are present. The lengths of plasmids were also confirmed using this digest (Figure 5.7).

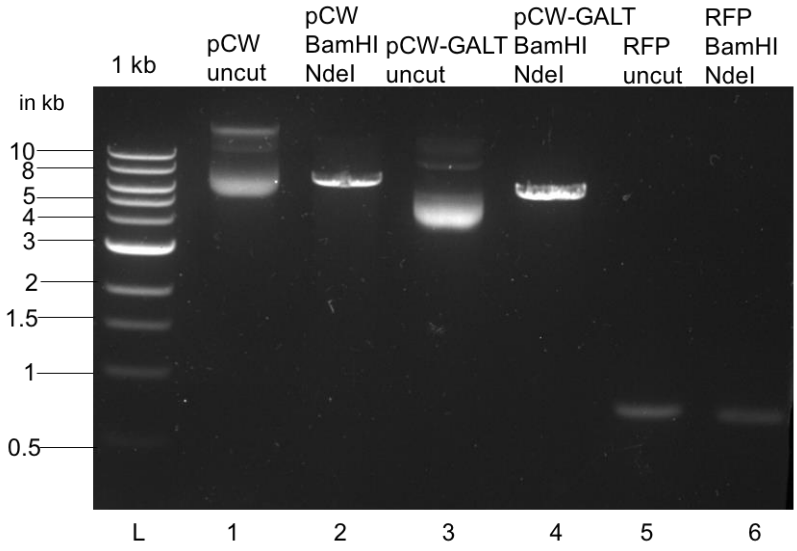


Figure 5.6. 1% agarose gel image of pCW vector, pCW-GALT plasmid, and RFP gene; (1) uncut vector; (2) restricted vector; (3) uncut plasmid; (4) restricted plasmid; (5) uncut gene; (6) restricted gene.

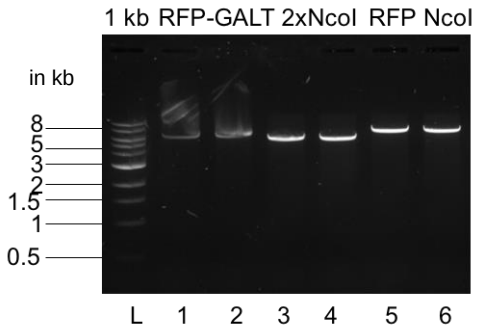


Figure 5.7. 1% agarose gel image of pCW-RFP-GALT plasmid and pCW-RFP plasmid; (1,2)uncut plasmid; (3,4) restricted plasmid; (5,6) restricted plasmid.

8) *Bam*HI (G/GATCC) and *Kpn*I-HF (GGTAC/C) double digestion: *Bam*HI 1 µl (20 u/µl); *Kpn*I-HF 1 µl (20 u/µl); DNA 5 µg; Neb 3.1 Buffer 5 µl (100 mM NaCl, 50 mM TRIS-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA, pH 7.9, 25°C); ddH_2O to 50 µl. Incubated at 37ºC for 16 h (Figure 5.8).

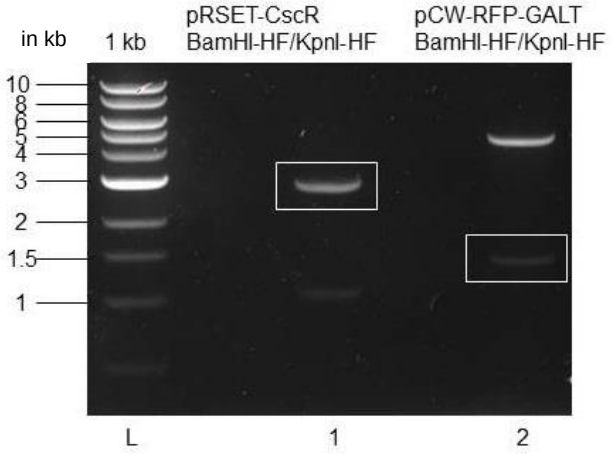


Figure 5.8. 1% agarose gel image of pRSET-CscR plasmid and pCW-RFP-GALT plasmid; (1) restricted plasmid; (2) restricted plasmid.

5.15. Ligation experiments

Due to the low efficiency of T4 and T4 quick ligations, an alternative method was found. It was decided to use the In-fusion cloning method because it is quite efficient and less complex than T4 ligation. The same PCR primers and restriction enzymes were used. Moreover, there is no need to cut the insert; only the vector is cut during In-fusion cloning. Different ligation conditions were tried, including varying the insert to vector ratio, ligation time, and DNA concentration. The conditions yielding the best results are presented below (Table 5.2).

T4 quick ligation	pCW-(GALT)
Vector	200 ng
B4GALT1	60 ng (3:1)
2× Quick T4 buffer	10 µl
Quick T4 ligase	1 µl
ddH_2O	to 20 µl
V = 20 µl T = 25ºC	t = 5 min

Table 5.1. T4 ligation reaction conditions for insertion of the B4GALT1 gene into the pCW-LIC vector. This was the first successful ligation to give the working plasmid pCW-GALT. This plasmid was used for the next steps of this project.

T4 quick ligation	pCW-(RFP)-GALT	In-fusion	pCW-(RFP)	pCW-(RFP)-GALT
Vector	200 ng	Vector	200 ng	200 ng
RFP	70 ng (3:1)	RFP	40 ng (2:1)	46 ng (2:1)
2× Quick T4 buffer	10 µl	5x In-fusion premix	2 µl	2 µl
Quick T4 ligase	1 µl	ddH_2O	to 10 µl	to 10 µl
ddH_2O	to 20 µl			
V = 20 µl T = 25ºC	t = 5 min	V = 10 µl	T = 50 ºC	t = 15 min

Table 5.2. T4 and In-fusion ligation cloning conditions for insertion of the RFP gene into the pCW-LIC and pCW-GALT vectors. The T4 ligation method was slightly more efficient in terms of the number of colonies and the plasmid concentration.

5.16. PCR

To amplify the genes of interest and to analyse the cloning attempts, the Q5 Hi-fi DNA polymerase (New England Biolabs, Ipswich, MA) was employed, which achieved a higher efficiency than Taq and CloneAMP Hi-fi polymerases (Figure 5.9). As a result, it was decided to use Q5 polymerase for all PCR experiments unless otherwise stated. The general PCR method is presented below (Table 5.3b). The extension time was dependent on DNA length, and varied for different samples. The annealing temperature was dependent on the primer melting temperature. Most primers in this project had a melting temperature ≥72°C (Appendix 1). As a result, annealing and extension steps were usually combined.

a) PCR components

DNA	5 ng
10 mM F-primer	2.5 µl
10 mM R-primer	2.5 µl
ddH_2O	to 25 µl
Q5 premix	25 µl
	V = 50 µl

b) PCR process

Denaturation	98ºC 30 s
35 cycles:	
1) Denaturation	98ºC 10 s
2) Annealing	72ºC 30 s
3) Extension	72ºC 30 s/kb
Final extension	72ºC 120 s

Table 5.3. **a)** Q5 Hi-fi PCR components. For single cloning analysis, a 20 µl volume was used. **b)** Q5 PCR process. The annealing temperature was 72ºC for most primers due to their high melting temperature. The extension time was 30 s for most inserts because they were shorter than 1 kb. DNA was held at 10ºC after reactions were completed.

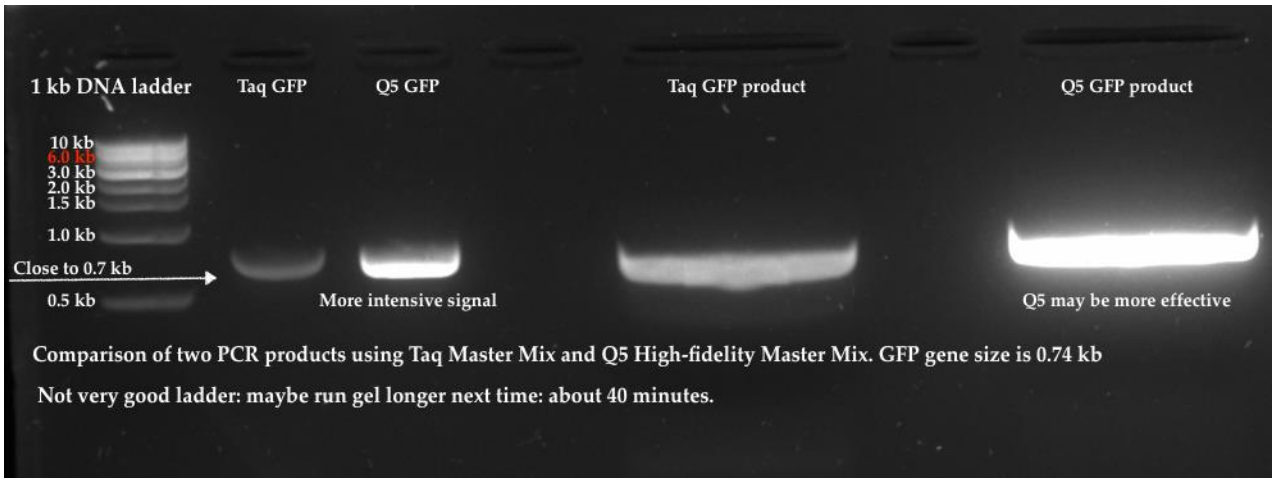


Figure 5.9. Comparison of Taq and Q5 Hi-fi DNA polymerases. The same primers were used to amplify the GFP gene from the pAAV-GFP vector. Q5 polymerase was nearly 2-fold more effective than Taq, and more DNA was extracted. A high DNA concentration is important because a significant amount of DNA is lost during restriction enzyme digestion.

5.17. Site-directed mutagenesis of the pCW-RFP-GALT plasmid

Deletion of one cytosine was detected at the end of the B4GALT1 gene inserted in the pCW-GALT and pCW-RFP-GALT plasmids by Sanger sequencing (Figure 5.10). This probably occurred due to the long exposure under the short UV wavelength (254 nm) during gel extraction, or during amplification or restriction digestion of the B4GALT1 insert.

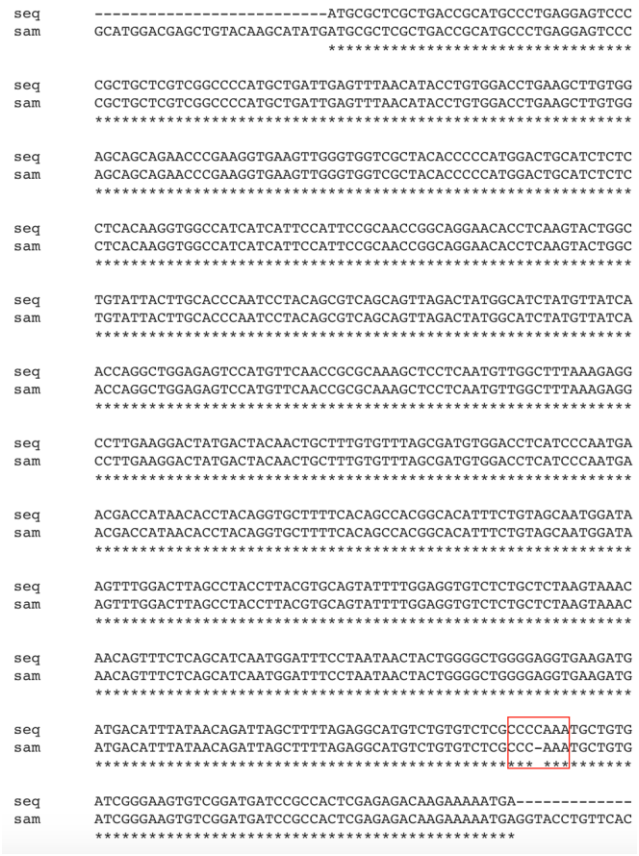


Figure 5.10. Alignment of the B4GALT1 gene sequence from the pCW-RFP-GALT plasmid with the B4GALT1 database reference sequence. The cytosine deletion site is highlighted in the red box. This deletion did not affect the RFP part of the protein because B4GALT1 gene is translated after RFP gene; However, the B4GALT1 protein was translated incorrectly, and there was no fluorescence without IPTG induction. This additionally indicates that B4GALT1 is responsible for 'autoinduction' in M9 glucose medium.

a) PCR mutation components

pCW-RFP-GALT	10 ng
10× Buffer	5 µl
GALT-Mut primers	250 ng
dNTP mix	1 µl
Quick Solution	3 µl
ddH_2O	to 50 µl

b) PCR mutation process

Denaturation	95ºC 60 s
18 cycles:	
1) Denaturation	95ºC 50 s
2) Annealing	60ºC 50 s
3) Extension	68ºC 420 s
Final e*xtension*	68ºC 420 s

Table 5.4. a) Mutation PCR components. Specific primers with an additional cytosine in the middle covering the B4GALT1 cytosine deletion site were used (Appendix 1). **b)** Process used for PCR mutation.

5.18. Synthesis of CscR gene.

The reverse CscR gene under the control of the LacIq promoter was synthesised by GeneArt service (Invitrogen) and delivered inside the pRSET A expression vector (Figure 5.11). The pRSET expression vector was employed because it is a small and efficient T7 promoter vector.

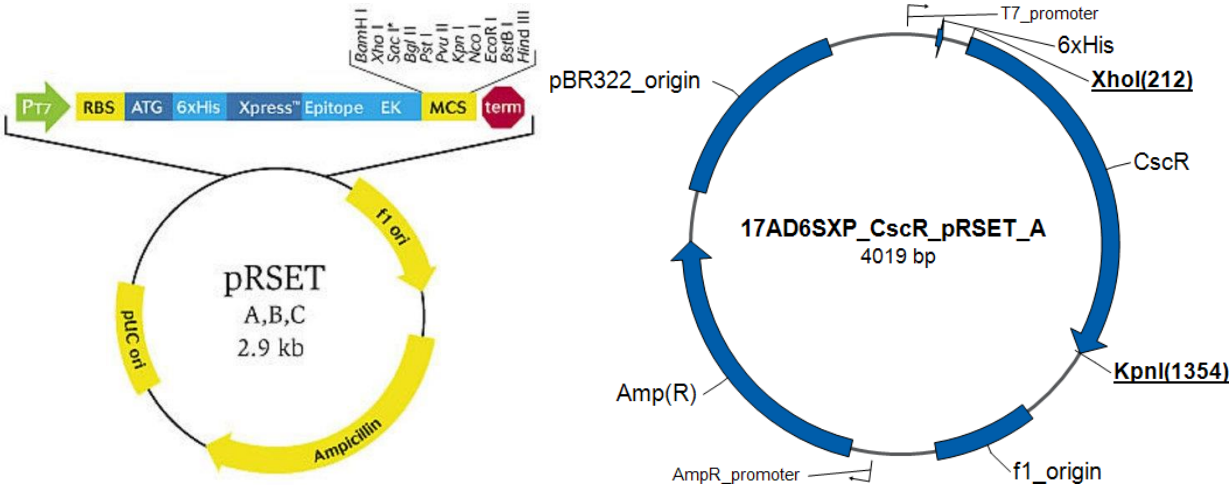


Figure 5.11. A schematic map of the pRSET vector is available at https://www.thermofisher.com.

5.19. Sucrose inducible recombinant plasmid construction.

*Bam*HI and *Kpn*I restriction sites were chosen because the pCW-RFP-GALT plasmid has the same restriction sites. This allows single-step digestion of the pRSET vector and ligation of the RFP-GALT genes (Figure 5.12, see Appendix for plasmid maps).

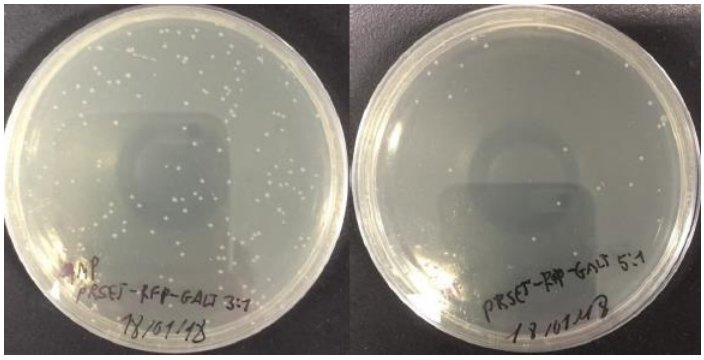


Figure 5.12. Colonies of BL 21 (DE3) cells on LB ampicillin agar after T4 ligation of RFP-GALT gene inside pRSET vector.

E. coli XL-10 cells were transformed with the pRSET-CscR vector and grown in LB ampicillin-selective medium to amplify and extract plasmid DNA. The pRSET-CscR and pCW-RFP-GALT plasmids were double-digested with *Bam*HI and *Kpn*I enzymes (Figure 5.8). The LacIq CscR gene was amplified from the pRSET-CscR plasmid by Q5 Hi-fi PCR (Figure 5.13).

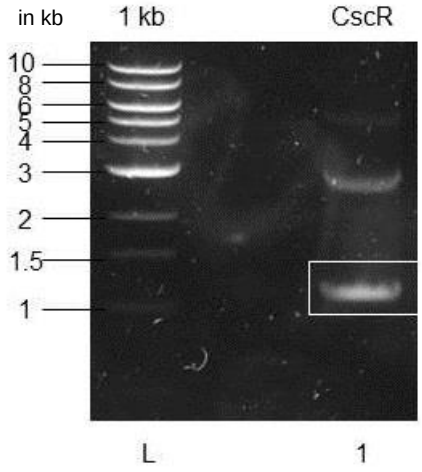


Figure 5.13. 1% agarose gel image of CscR gene. (1 boxed) amplified gene.

The *Apa*I restriction site was added to the non-coding region of the pRSET vector, 160 bp upstream of the T7 promoter, using a QuickChange site-directed mutagenesis kit. The pRSET-RFP-GALT plasmid was sequentially digested by *Pci*I and *Apa*I restriction enzymes (Figure 5.14).

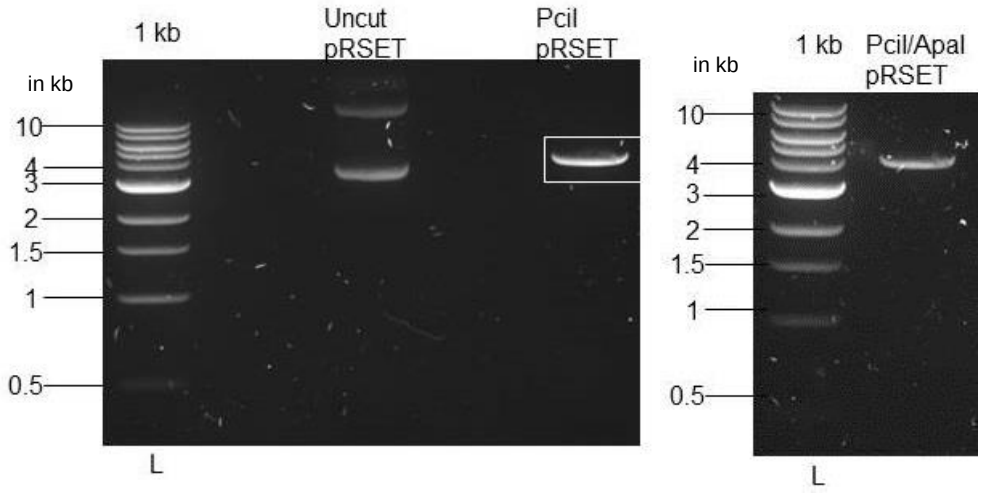


Figure 5.14. 1% agarose gel images of pRSET vector, and pRSET-RFP-GALT plasmid; (left boxed) restricted vector; (right) restricted plasmid.

The amplified CscR gene was inserted into the *Pci*I/*Apa*I restricted pRSET-RFP-GALT vector using an In-fusion cloning kit. As a result, the pRSET-RPF-GALT_CscR plasmid was constructed (Figure 5.15).

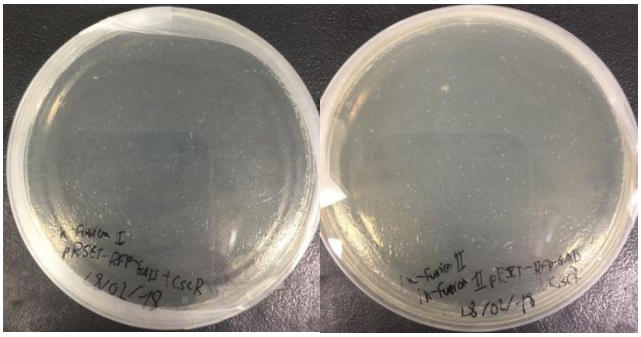


Figure 5.15. Colonies of BL 21 (DE3) cells on LB Ampicillin agar after In-fusion cloning of CscR gene inside pRSET-RFP-GALT vector.

The 12 bp CscO operator site was added to the pRSET-RPF-GALT_CscR vector directly after the T7 promoter using a QuickChange site-directed mutagenesis kit. As a result, the final pRSET(CscO)-RFP-GALT_CscR plasmid was constructed.

The pCW (pCW-RFP; pCW-RFP-GALT)-based plasmids were transformed into *E. coli* BL21 cells, and the pRSET (pRSET-RFP-GALT; pRSET-RFP-GALT_CscR; pRSET(CscO)-RFP-GALT_CscR)-based plasmids were transformed into *E. coli* BL21 (DE3) cells. All the cultures were grown in parallel in LB ampicillin media for 16 h, and their fluorescence intensity was compared (Figure 5.16)

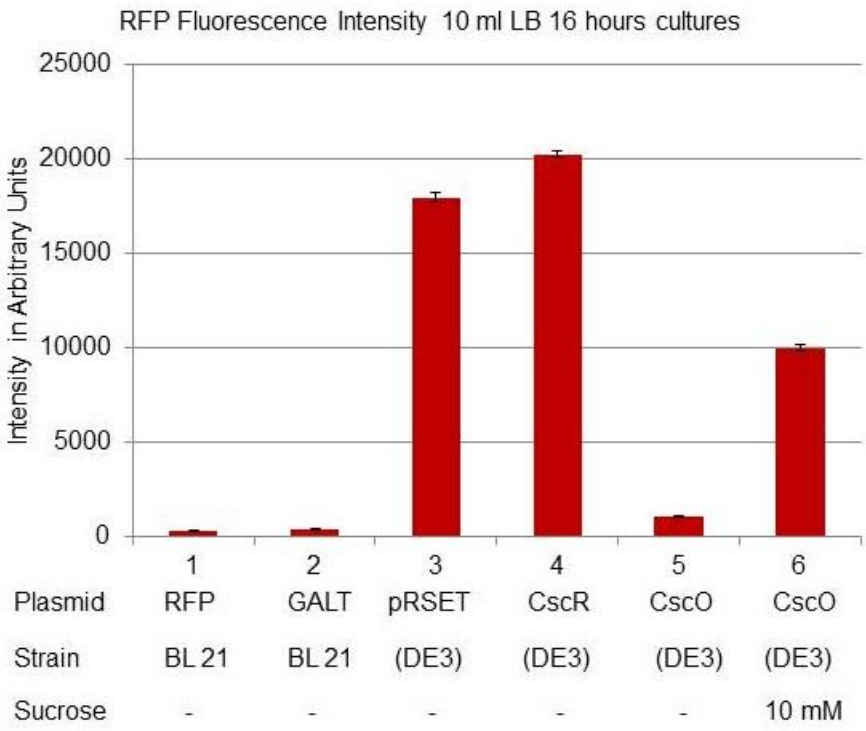


Figure 5.16. A comparison of RFP expression levels based on fluorescence intensity for pCW-RFP (1), pCW-RFP-GALT (2), pRSET-RFP-GALT (3), pRSET-RFP-GALT_CscR (4), pRSET(CscO)-RFP-GALT_CscR (5,6). Confidence intervals were calculated based on ten measurements for each sample with significance 0.05.

The same transformants were grown in M9 0.4% glucose minimal medium. However, none of the *E. coli* BL21 (DE3) cells grew during the first 12 h of incubation in M9 media for unknown reasons. This could be due to features of the *E. coli* BL21 (DE3) strain or the pRSET vector. None of the strains grew on 1% sucrose M9 (Figure 5.17).

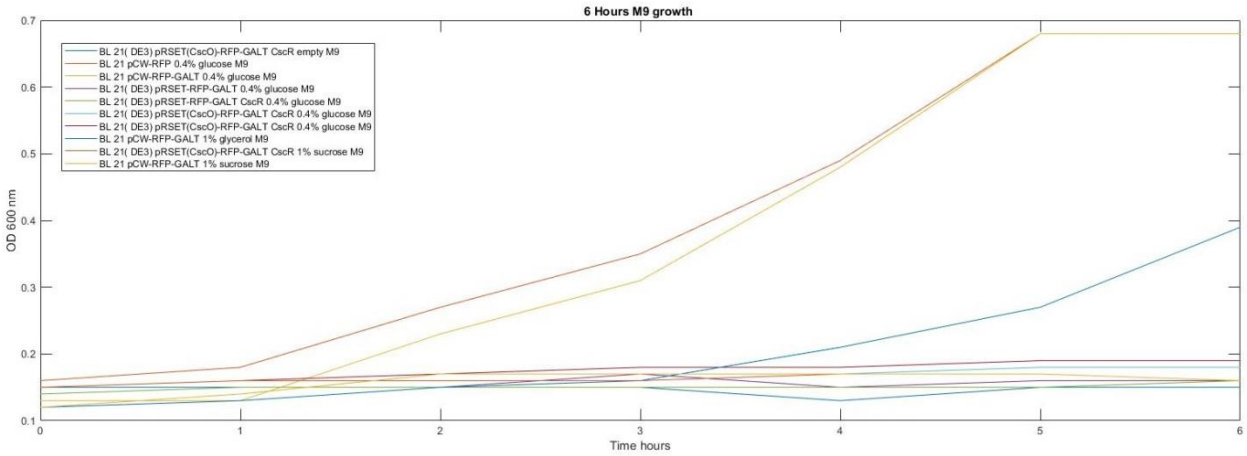


Figure 5.17. 6 hours growth curves of different BL 21 (DE3) cultures based on OD 600 nm measurements.

When *E. coli* BL21 (DE3) colonies harbouring pRSET-RFP-GALT_CscR or pRSET(CscO)-RFP-GALT_CscR plasmids were analysed on LB ampicillin agar, there was a significant difference in their colour due to overexpression of RPF from the T7 promoter. Colonies harbouring the pRSET-RFP-GALT plasmid were pink in colour (Figure 5.18)

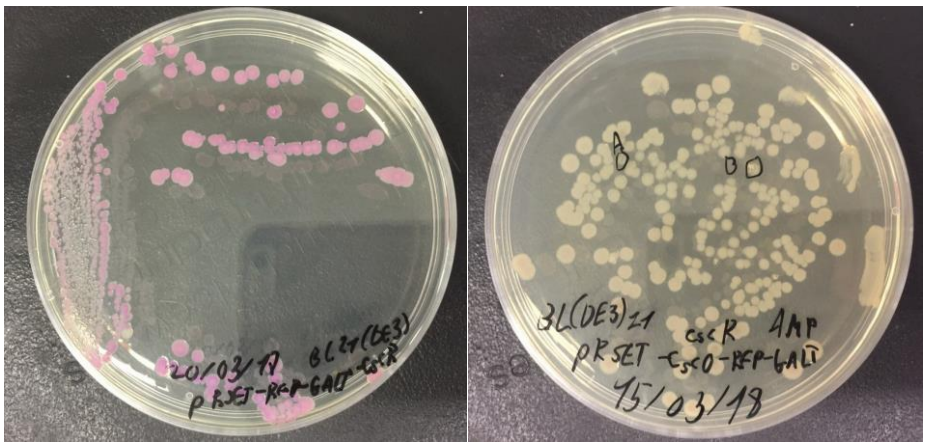


Figure 5.18. LB ampicillin agar plates 1 week after inoculating with BL21 (DE3) cells harbouring pRSET-RFP-GALT_CscR (left) plasmids or pRSET(CscO)-RFP-GALT_CscR (right).

5.20. Electro-competent cells preparation and electroporation of deltaZYA cells

The Delta-ZYA cell culture was a gift from Tanja Kortemme (Addgene plasmid # 52695). 500 ml delta-ZYA liquid LB culture was prepared and incubated overnight at 37ºC with shaking at 200 rpm. The overnight culture OD at 600 nm was 2.0. The culture was cooled on ice for 10 minutes and transferred to 50 ml falcon tube, then was centrifuged at 2000 g for 10 minutes. The supernatant was removed, and cell pellet was washed with 1 ml of ultrapure water. Washed cells were dissolved in 50 ml of ultrapure water and centrifuged at 2000 g for 10 minutes. Cells were washed one more time and centrifuged again. Cell pellet was dissolved in 200 µl of GYT broth (0.125% yeast extract, 0.25% tryptone, 10% glycerol). 10 µl of dissolved cells were transferred to 1 ml of GYT broth for OD measurement. Measured OD at 600 nm was 1.19. Based on this, 200 µl GYT broth cell mixture OD at 600 nm was calculated to be nearly 100. Concentrated cells were aliquoted in Eppendorf tubes in volumes of 35 µl and were put in -80 Cº freezer for storage. Cells could be stored in –80 Cº after each step of this preparation.

LB liquid medium and agar plates with ampicillin (100 µg/ml) and chloramphenicol (25 µg/ml) were prepared for pRSET and pTARA plasmids selection. Delta-ZYA electro-competent cells were thawed on ice and mixed with 100 ng of pTARA and pRSET(CscO)-RFP-GALT_CscR plasmid. Then cells were transferred to prechilled 1 mm transformation cuvette. Electroporation was performed with 10 µF capacitance, 600 Ω resistance, 1800 V voltage and 4 ms pulse in Eppendorf 2510 Electroporator. 700 µl of SOC medium was added immediately after pulse. Cells were incubated at 37Cº for 1 h at 250 rpm. 200 µl of transformed cells were spread on selective agar plates and incubated overnight. Two single colonies were observed after 20 h of incubation (Figure 5.19) and used to prepare 40% glycerol stock of delta-ZYA cells harbouring pTARA and pRSET plasmids.

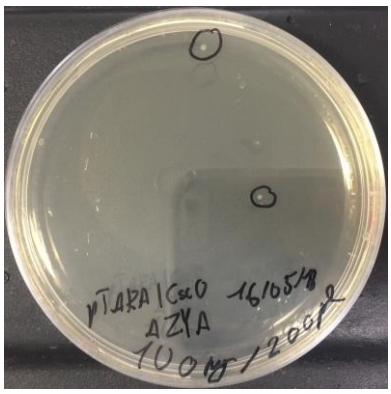


Figure 5.19. Colonies of delta-ZYA strain transformed with pTARA and pRSET(CscO)-RFP-GALT_CscR plasmids on Chloramphenicol and ampicillin selection LB agar.

5.21. Live cell NMR experiments

It was decided to conduct ^{1}H NMR experiments of BL 21 cells harbouring pCW-RFP-GALT plasmid to attempt to detect lactose produced by B4GALT1 enzyme. pCW-RFP harbouring cells were used as a control. Unfortunately, no peaks corresponding to lactose were detected probably because of others metabolites peaks and low lactose concentrations. However, a large glycerol peak was observed for M9 1% glycerol culture and large lactate peak was observed for M9 0.4% glucose culture (Figure 5.20). A comparison of pCW-RFP and pCW-RFP-GALT cultures did not show any significant difference. To summarise, the concentration of lactose produced by pCW-RFP-GALT harbouring cells is too low for the direct ^{1}H cell NMR measurements and existing protocol should be improved for future experiments.

LB or M9 BL 21 cells overnight cultures were used for *in vivo* NMR measurements. 10 ml of overnight cell culture was centrifuged at 2000 g and supernatant was removed. Cell pellet was dissolved in 450 µl of relevant medium and 50 µl of D_2O. NMR ^{1}H spectroscopy of BL 21 *E. coli* cells was performed on Bruker AVB 400 magnet with the center field strength 9.4 tesla. Free induction decay data was analyzed using MestReNova software (Figures 5.20; 5.21).

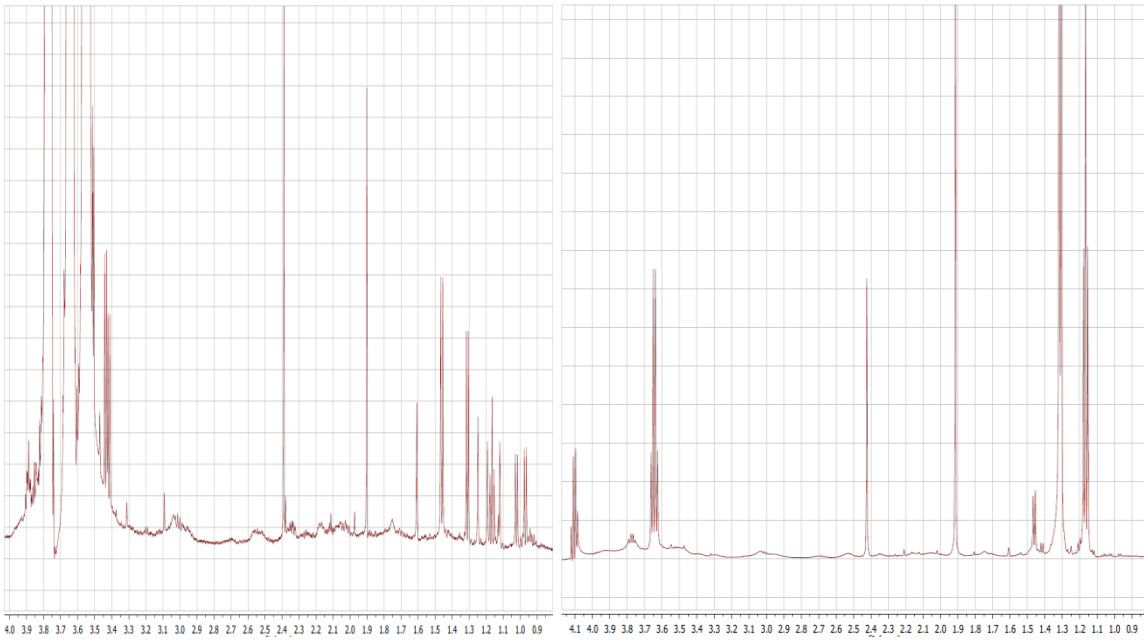


Figure 5.20. ^{1}H Watergate (400 MHz) NMR spectra of BL 21 cells harbouring pCW-RFP-GALT plasmid in 1% glycerol M9 medium (40 OD_{600}) on the left and in 0.4% glucose M9 medium (80 OD_{600}) on the right.

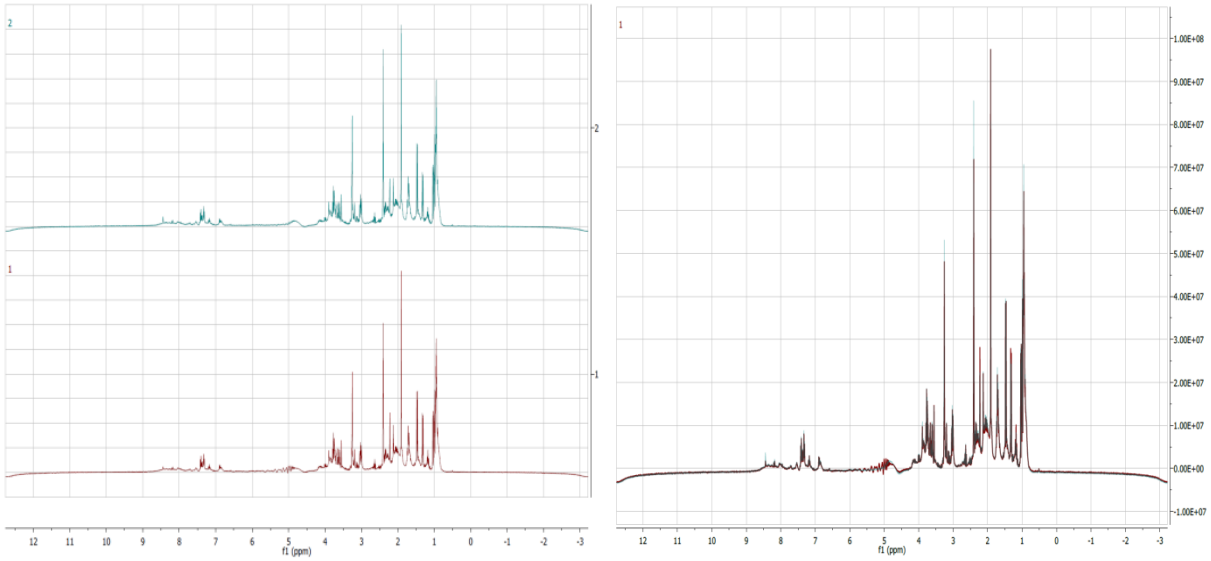


Figure 5.21. ^{1}H Excitation sculpting, water suppression (400 MHz) NMR spectra of BL 21 cells in LB medium. Blue coloured spectrum is for the cells harbouring pCW-RFP plasmid (36 OD_{600}) and red coloured spectrum is for the cells harbouring pCW-RFP-GALT plasmid (34 OD_{600}). Both spectra are stacked on the right for comparison.

6. Appendices

Appendix 1. Sequences of primers used in this project

Primer name	Sequence	Purpose
pCW-F	CTCGTATAATGTGTGGAATTGTGAG	Sequence
pCW-R	AGTTTTGTTCAGCGGCTTGTATGGG	Sequence
Inside_cherry	AAGGTGACCAAGGGTGGCCCCCTGC	Sequence
LacZ_GFP	ACTCATCCTCGCCGTTTTACTCTTTTT	Sequence
GALT_bowine	CGGGCTGCTGAGCCGGCCAGAGTCGCCGAGATCCG	Sequence
GALT_human	TAGCCCACACCCTTCTTAAAGCGGCGGCGGGAAGA	Sequence
PMalC2x_lac_F	GCAAGCTTGGCACTGGCCGTCGTTT	Sequence
PMalC2x_lac_R	AAACGACGGCCAGTGCCAAGCTTGC	Sequence
100bpBamHI_F	ACGCCGCCAGCGGTCGTCAGACTGT	Sequence
100bpBamHI_R	TTACGCCAGCTGGCGAAAGGGGGAT	Sequence
Lacpr_sanger	AAACAGGATCAGCTTACTCCCCATCCCCCT	Sequence
RFP-R	GACAGGATGTCCCAGGCGAAGGGCA	Sequence
lacZYAGFP-F1	GAGCCCGCAGCTCCGGGCCCATGGTGAGCAAGGGCGAG	Sequence
lacZYAGFP-R1	TGGCGGCCGCTCTAGACTACTTGAGCTCGAGATCT	Sequence
GFP	TGCACGGTGCAGATGCGCCGCTTCGAGA	Sequence
F-GALT-Mut	GCATGTCTGTGTCTCGCCCCAAATGCTGTGATCG	SDM
R-GALT-Mut	CGATCACAGCATTTGGGGCGAGACACAGACATGC	SDM
CowGALT_NdeI_F	GCTTAGGAGGTCATATGATGCGCTCGCTGACCGCATGC	Ligation
CowGALT_KpnI_R	GGAGTCAGTGAACAGGTACCTCATTTTTCTTGTCTCTCGAGTGGC	Ligation
NdeI-R	GAACTCATAATCATATGCTTGTACAGCTCGTCCATGCC	Ligation
BamHI-F	ACAGGAAACAGGATCCGTGAGCAAGGGCGAGGAGG	Ligation
BsrGI_F	CGGGACACCCTGTACAATGGTGAGCAAGGGCGAGG	Ligation
BbvCI_R	CACTTCCAGCGCTGAGGCTACTTGAGCTCG	Ligation
F-Primer-GFP	GGCATTTAGCCTTAAATGGTGAGCAAGGGCGAG	Ligation
R-Primer-GFP	ACATTGATTGCTTAACTACTTGAGCTCGAGATCT	Ligation
RFP-R	GACAGGATGTCCCAGGCGAAGGG	Sequence
ApaI_mut-R	CGGTAATACGGTTATGGGCCCACAGAATCAGGGG	SDM
ApaI_mut-F	CCCCTGATTCTGTGGGCCCATAACCGTATTACCG	SDM
CscO_mut-R	ACCGTTGTGGTCTCCGTTAACGTTAACCTATAGTGAGTCGTA	SDM
CscO_mut-F	TACGACTCACTATAGGTTAACGTTAACGGAGACCACAACGGT	SDM
pRSET_CscR-ligation-R	GGTAATACGGTTATGAGGACACCATCGAATGGTGCA	Ligation
pRSET_CscR-ligation-F	CCTTTTGCTCACATGAAGCCTGGGGTGCCTAATGAG	Ligation
CscR-pCW-R	GGAGTCAGTGAACAGGTACCTTATCGCTGTTTACTGAAAAG	Ligation
CscR-pCW-F	GCTTAGGAGGTCATATGATGGTCTCACTGAAAGAAGT	Ligation
inGALT_seq	GGCTTTAAAGAGGCCTTGAA	Sequence
Primer name	Sequence	Purpose

pRSET_upamp_seq	TAAATCGGGGGCTCCCTTTA	Sequence
pRSET_ampstart_seq	TGGTGAAAGTAAAAGATGCT	Sequence
pRSET_Ampmiddle_seq	CCGGCAACAATTAATAGACT	Sequence
pRSET_amp_seq	CTACCAACTCTTTTTCCGAA	Sequence
pREST_upCscR_seq	TTTACGGTTCCTGGCCTTTT	Sequence
pRSET_upT7_seq	TTGGCCGATTCATTAATGCA	Sequence
pRSET_upRFP-GALT_seq	CTGTACGACGATGACGATAA	Sequence
EndAmp	TCGTGCACACAGCCCAGCTT	Sequence
InCscR	TGCAGTGTAATCATCAGGGA	Sequence
pRSET_AMP2	ATTTTGCCGATTTCGGCCTA	Sequence
pRSET_AMP1	ATCTGGCGTAATAGCGAAGA	Sequence
inGALT_2	TCAGCAGTTAGACTATGGCA	Sequence
R_CscB_T4	TTTTTGAATTCCTATATTGCTGAAGGTACAGGCG	Ligation
F_CscB_T4	TTTTTGGTACCTAACTTTAAGAAGGAGATATACATATGGCACTGAATATT	Ligation
pRSET_T7CscO_mut-R	GGAAACCGTTGTGGTCTCCGTTAACGTTAACCTATAGTGAGTCGTATTAA	SDM
pRSET_T7CscO_mut-F	TTAATACGACTCACTATAGGTTAACGTTAACGGAGACCACAACGGTTTCC	SDM
CscB_R	ATCAAGCTTCGAATTCTATATTGCTGAAGGTACAGGCGT	Ligation
CscB_F	GACAAGAAAAATGAGAAGAAGGAGATATACATATGGCACTGAATATTCCA	Ligation
CscB_long-F	TTTTTGGTACCTTTAACTTTAAGAAGGAGATATACATATGGCACTGAA TATTCCATTCAGAAATGCGTACTATCGTT	Ligation
CscB_long-R	TTTTTGAATTCGCAAATAGTTATGGAAACGCCTGTACCTTCAGCAATA TAG	Ligation

Appendix 2. pCW-GALT plasmid sequence based on pCW vector (plasmid 1) sequence submitted by depositor and inserts containing sites (in color) analyzed by the Sanger sequencing using constructed primers (see appendix 1). B4GALT1 gene is highlighted in yellow, important restriction sites are red colored, primer sites are highlighted in green. Ampicillin resistance gene is highlighted in dark-green. LacI gene is highlighted in pink.

GCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCC
GCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTG
CTTTACGGCACCTCGACCCCAAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGA
GTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATT
TCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATCAGGTGGCACTTTTCGGGGA
AATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCT
GGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCC
GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCA
TACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCAT
AACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTA
ACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCA
AACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGC
CCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGT
AACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT
CATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCT
TCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT
GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCG
GGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGC
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTG
AGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTAC
GCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCT
ACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCT
GTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGAACGCCATCAAAAATAATTCGCGTCTGGCCT
TCCTGTAGCCAGCTTTCATCAACATTAAATGTGAGCGAGTAACAACCCGTCGGATTCTCCGTGGGAACAAACGGCGGATTGACCGTAATGGGATAGG
TTACGTTGGTGTAGATGGGCGCATCGTAACCGTGCATCTGCCAGTTTGAGGGGACGACGACAGTATCGGCCTCAGGAAGATCGCACTCCAGCCAGCT
TTCCGGCACCGCTTCTGGTGCCGGAAACCAGGCAAAGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCG
CTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCA
GTGAATCCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAG
CCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATG
AATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCAC
CGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTGACGGCGGGATATAACAT
GAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCT
GATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCG
TTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGC
GCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAG
AGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGAC
GCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCG
CGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTT
GTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCG
GGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGC
TATCATGCCATACCGCGAAAGGTTTTGCACCATTCGATGGTGTCCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTA
ATGTAAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACAC
AGGAAACAGGATCGATCCATCGATGAGCTTACTCCCCATCCCCCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAC
AATTTCACACAGGAAACAGGATCAGCTTACTCCCCATCCCCCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAA
TTTCACACAGGAAACAGGATCCATCGATGCTTAGGAGGTCATATGATGCGCTCGCTGACCGCATGCCCTGAGGAGTCCCCGCTGCTCGTCGGCCCCA
TGCTGATTGAGTTTAACATACCTGTGGACCTGAAGCTTGTGGAGCAGCAGAACCCGAAGGTGAAGTTGGGTGGTCGCTACACCCCCATGGACTGCAT
CTCTCCTCACAAGGTGGCCATCATCATTCCATTCCGCAACCGGCAGGAACACCTCAAGTACTGGCTGTATTACTTGCACCCAATCCTACAGCGTCAG
CAGTTAGACTATGGCATCTATGTTATCAACCAGGCTGGAGAGTCCATGTTCAACCGCGCAAAGCTCCTCAATGTTGGCTTTAAAGAGGCCTTGAAGG
ACTATGACTACAACTGCTTTGTGTTTAGCGATGTGGACCTCATCCCAATGAACGACCATAACACCTACAGGTGCTTTTCACAGCCACGGCACATTTC
TGTAGCAATGGATAAGTTTGGACTTAGCCTACCTTACGTGCAGTATTTTGGAGGTGTCTCTGCTCTAAGTAAACAACAGTTTCTCAGCATCAATGGA
TTTCCTAATAACTACTGGGGCTGGGGAGGTGAAGATGATGACATTTATAACAGATTAGCTTTTAGAGGCATGTCTGTGTCTCGCCCCAAATGCTGTG
ATCGGGAAGTGTCGGATGATCCGCCACTCGAGAGACAAGAAAAATGAGGTACCTGTTCACTGACTCCCGCGGATCAAAAATGACGATTGACGGCATT
ACGTCTAACGATATTTACATGCTTGGTTATGTTTCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAATGGATC
TTGATCCTAACGATGTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAACAATGTCGTGATTACAAGCTATATGACAAACAGAGG
ATTCTACGCAGACAAACAATCAACGTTTGCGCCTAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGTTGTCAAAGACAGCATCCTTGAACAA
GGACAATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATATCCTATTGGCATTGACGTCAGGTGGCACTTTTCGAGGAGATCATGCACAT
GATGACGAAGCTTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATGAGCGGGCTTTTTTTTCAGATCTGCTTGAAGACGAAAG
GGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGGTGCTCGTCAAA

Appendix 3. pCW-RFP plasmid full sequence. RFP gene is highlighted in dark-red, part deleted in pCW-RFP-GALT is highlighted in red, important restriction sites are red colored, primer sites are highlighted in green.

GCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCC
GCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTG
CTTTACGGCACCTCGACCCCAAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGA
GTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATT
TCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATCAGGTGGCACTTTTCGGGGA
AATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCT
GGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCC
GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCA
TACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCAT
AACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTA
ACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCA
AACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGC
CCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGT
AACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT
CATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCT
TCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT
GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCG
GGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGC
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTG
AGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTAC
GCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCT
ACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCT
GTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGAACGCCATCAAAAATAATTCGCGTCTGGCCT
TCCTGTAGCCAGCTTTCATCAACATTAAATGTGAGCGAGTAACAACCCGTCGGATTCTCCGTGGGAACAAACGGCGGATTGACCGTAATGGGATAGG
TTACGTTGGTGTAGATGGGCGCATCGTAACCGTGCATCTGCCAGTTTGAGGGGACGACGACAGTATCGGCCTCAGGAAGATCGCACTCCAGCCAGCT
TTCCGGCACCGCTTCTGGTGCCGGAAACCAGGCAAAGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCG
CTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCA
GTGAATCCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAG
CCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATG
AATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCAC
CGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTGACGGCGGGATATAACAT
GAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCT
GATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCG
TTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGC
GCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAG
AGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGAC
GCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCG
CGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTT
GTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCG
GGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGC
TATCATGCCATACCGCGAAAGGTTTTGCACCATTCGATGGTGTCCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTA
ATGTAAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACAC
AGGAAACAGGATCGATCCATCGATGAGCTTACTCCCCATCCCCCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAC
AATTTCACACAGGAAACAGGATCAGCTTACTCCCCATCCCCCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAA
TTTCACACAGGAAACAGGATCCGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCC
GTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCC
TGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTC
CTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTCCAGGACGGCGAGTTC
ATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGT
ACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGC
CAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAA
CGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGCATATGATTATGAGTTCTCCTCCTGAAAGATCCATAACTTCGTATAGCA
TACATTATACGAAGTTATGCGGCCGCGACGTCCACATATACCTGCCGTTCACTATTATTTAGTGAAATGAGATATTATGATATTTTCTGAATTGTGA
TTAAAAAGGCAACTTTATGCCCATGCAACAGAAACTATAAAAAATACAGAGAATGAAAAGAAACAGATAGATTTTTTAGTTCTTTAGGCCCGTAGTC
TGCAAATCCTTTTATGATTTTCTATCAAACAAAAGAGGAAAATAGACCAGTTGCAATCCAAACGAGAGTCTAATAGAATGAGGTCGAAAAGTAAATC
GCGCGGGTTTGTTACTGATAAAGCAGGCAAGACCTAAAATGTGTAAAGGGCAAAGTGTATACTTTGGCGTCACCCCTTACATATTTTAGGTCTTTTT
TTATTGTGCGTAACTAACTTGCCATCTTCAAACAGGAGGGCTGGAAGAAGCAGACCGCTAACACAGTACATAAAAAAGGAGACATGAACGATGAACA
TCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTGGCAGGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAA
GCCATATAAGGAAACATACGGCATTTCCCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAAAATGAAAAATATCAAGTTCCTGAG
TTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAAAAGGCCTGGACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACT
ATCACGGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCGATTTACATGTTCTATCAAAAAGTCGGCGAAACTTC
TATTGACAGCTGGAAAAACGCTGGCCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTATCCTAAAAGACCAAACACAAGAATGGTCA
GGTTCAGCCACATTTACATCTGACGGAAAAATCCGTTTATTCTACACTGATTTCTCCGGTAAACATTACGGCAAACAAACACTGACAACTGCACAAG
TTAACGTATCAGCATCAGACAGCTCTTTGAACATCAACGGTGTAGAGGATTATAAATCAATCTTTGACGGTGACGGAAAAACGTATCAAAATGTACA
GCAGTTCATCGATGAAGGCAACTACAGCTCAGGCGACAACCATACGCTGAGAGATCCTCACTACGTAGAAGATAAAGGCCACAAATACTTAGTATTT
GAAGCAAACACTGGAACTGAAGATGGCTACCAAGGCGAAGAATCTTTATTTAACAAAGCATACTATGGCAAAAGCACATCATTCTTCCGTCAAGAAA

GTCAAAAACTTCTGCAAAGCGATAAAAAACGCACGGCTGAGTTAGCAAACGGCGCTCTCGGTATGATTGAGCTAAACGATGATTACACACTGAAAAA
AGTGATGAAACCGCTGATTGCATCTAACACAGTAACAGATGAAATTGAACGCGCGAACGTCTTTAAAATGAACGGCAAATGGTACCTGTTCACTGAC
TCCCGCGGATCAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACATGCTTGGTTATGTTTCTAATTCTTTAACTGGCCCATACAAGCCGC
TGAACAAAACTGGCCTTGTGTTAAAAATGGATCTTGATCCTAACGATGTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAACAA
TGTCGTGATTACAAGCTATATGACAAACAGAGGATTCTACGCAGACAAACAATCAACGTTTGCGCCTAGCTTCCTGCTGAACATCAAAGGCAAGAAA
ACATCTGTTGTCAAAGACAGCATCCTTGAACAAGGACAATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATATCCTATTGGCATTGACG
TCAGGTGGCACTTTTCGAGGAGATCATGCACATGATGACGAAGCTTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATGAGCG
GGCTTTTTTTTCAGATCTGCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGGTGCTCG
TCAAA

Appendix 4. pCW-RFP-GALT plasmid (plasmid 5) full sequence. RFP gene is highlighted in dark-red, B4GALT1 gene is highlighted in yellow, important restriction sites are red coloured, primer sites are highlighted in green. Ampicillin resistance gene is highlighted in dark-green. LacI gene is highlighted in pink.

GCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCC
GCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTG
CTTTACGGCACCTCGACCCCAAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGA
GTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGATT
TCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATCAGGTGGCACTTTTCGGGGA
AATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATT
GAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCT
GGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCC
GAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCA
TACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCAT
AACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTA
ACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCA
AACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGC
CCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGT
AACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT
CATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC
GTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCT
TCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCT
GCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCG
GGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGC
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTG
AGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTAC
GCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCT
ACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCT
GTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGAACGCCATCAAAAATAATTCGCGTCTGGCCT
TCCTGTAGCCAGCTTTCATCAACATTAAATGTGAGCGAGTAACAACCCGTCGGATTCTCCGTGGGAACAAACGGCGGATTGACCGTAATGGGATAGG
TTACGTTGGTGTAGATGGGCGCATCGTAACCGTGCATCTGCCAGTTTGAGGGGACGACGACAGTATCGGCCTCAGGAAGATCGCACTCCAGCCAGCT
TTCCGGCACCGCTTCTGGTGCCGGAAACCAGGCAAAGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCG
CTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCA
GTGAATCCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAG
CCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATG
AATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCAC
CGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTGACGGCGGGATATAACAT
GAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCT
GATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCG
TTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGC
GCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAG
AGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGAC
GCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCG
CGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTT
GTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCG
GGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGC
TATCATGCCATACCGCGAAAGGTTTTGCACCATTCGATGGTGTCCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTA
ATGTAAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACAC
AGGAAACAGGATCGATCCATCGATGAGCTTACTCCCCATCCCCCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAC
AATTTCACACAGGAAACAGGATCAGCTTACTCCCCATCCCCCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAA
TTTCACACAGGAAACAGGATCCGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCC
GTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCC
TGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTC
CTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTCCAGGACGGCGAGTTC
ATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGT
ACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGC
CAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAA
CGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGCATATGATGCGCTCGCTGACCGCATGCCCTGAGGAGTCCCCGCTGCTCG

TCGGCCCCATGCTGATTGAGTTTAACATACCTGTGGACCTGAAGCTTGTGGAGCAGCAGAACCCGAAGGTGAAGTTGGGTGGTCGCTACACCCCCAT
GGACTGCATCTCTCCTCACAAGGTGGCCATCATCATTCCATTCCGCAACCGGCAGGAACACCTCAAGTACTGGCTGTATTACTTGCACCCAATCCTA
CAGCGTCAGCAGTTAGACTATGGCATCTATGTTATCAACCAGGCTGGAGAGTCCATGTTCAACCGCGCAAAGCTCCTCAATGTTGGCTTTAAAGAGG
CCTTGAAGGACTATGACTACAACTGCTTTGTGTTTAGCGATGTGGACCTCATCCCAATGAACGACCATAACACCTACAGGTGCTTTTCACAGCCACG
GCACATTTCTGTAGCAATGGATAAGTTTGGACTTAGCCTACCTTACGTGCAGTATTTTGGAGGTGTCTCTGCTCTAAGTAAACAACAGTTTCTCAGC
ATCAATGGATTTCCTAATAACTACTGGGGCTGGGGAGGTGAAGATGATGACATTTATAACAGATTAGCTTTTAGAGGCATGTCTGTGTCTCGCCCCA
AATGCTGTGATCGGGAAGTGTCGGATGATCCGCCACTCGAGAGACAAGAAAAATGAGGTACCTGTTCACTGACTCCCGCGGATCAAAAATGACGATT
GACGGCATTACGTCTAACGATATTTACATGCTTGGTTATGTTTCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTGGCCTTGTGTTAA
AAATGGATCTTGATCCTAACGATGTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAACAATGTCGTGATTACAAGCTATATGAC
AAACAGAGGATTCTACGCAGACAAACAATCAACGTTTGCGCCTAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGTTGTCAAAGACAGCATC
CTTGAACAAGGACAATTAACAGTTAACAAATAAAAACGCAAAAGAAAATGCCGATATCCTATTGGCATTGACGTCAGGTGGCACTTTTCGAGGAGAT
CATGCACATGATGACGAAGCTTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATGAGCGGGCTTTTTTTTCAGATCTGCTTGA
AGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGGTGCTCGTCAAA

Appendix 5. Bovine beta-1,4-galactosyltransferase nucleotide sequence amplified from James Errey pMAL-c2x plasmid (plasmid 3).

ATGCGCTCGCTGACCGCATGCCCTGAGGAGTCCCCGCTGCTCGTCGGCCCCATGCTGATTGAGTTTAACATACCTGTGGACCTGAAGCTTGTGGAGC
AGCAGAACCCGAAGGTGAAGTTGGGTGGTCGCTACACCCCCATGGACTGCATCTCTCCTCACAAGGTGGCCATCATCATTCCATTCCGCAACCGGCA
GGAACACCTCAAGTACTGGCTGTATTACTTGCACCCAATCCTACAGCGTCAGCAGTTAGACTATGGCATCTATGTTATCAACCAGGCTGGAGAGTCC
ATGTTCAACCGCGCAAAGCTCCTCAATGTTGGCTTTAAAGAGGCCTTGAAGGACTATGACTACAACTGCTTTGTGTTTAGCGATGTGGACCTCATCC
CAATGAACGACCATAACACCTACAGGTGCTTTTCACAGCCACGGCACATTTCTGTAGCAATGGATAAGTTTGGACTTAGCCTACCTTACGTGCAGTA
TTTTGGAGGTGTCTCTGCTCTAAGTAAACAACAGTTTCTCAGCATCAATGGATTTCCTAATAACTACTGGGGCTGGGGAGGTGAAGATGATGACATT
TATAACAGATTAGCTTTTAGAGGCATGTCTGTGTCTCGCCCCAAATGCTGTGATCGGGAAGTGTCGGATGATCCGCCACTCGAGAGACAAGAAAAAT
GA

Appendix 6. mCherry nucleotide sequence amplified from pET mCherry LIC cloning vector (plasmid 2).

GTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGA
TCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCT
GTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGG
GAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTCCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCG
GCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAA
GGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCC
GGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCA
CCGGCGGCATGGACGAGCTGTACAAG

Appendix 7. Amino acid sequence of RFP-B4GALT1 fused protein from pCW-RFP-GALT plasmid. Chromophore forming site is highlighted in red. Galactosyltransferase catalytic domain is highlighted in green. RFP part is red coloured and B4GALT1 part is blue coloured.

StartAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQ
FMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTN
FPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGA
YNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKHMMRSLTACPEESPLLVGPMLIEFN
IPVDLKLVEQQNPKVKLGGRYTPMDCISPHKVAIIIPFRNRQEHLKYWLYYLHPILQRQQLDYG
IYVINQAGESMFNRAKLLNVGFKEALKDYDYNCFVFSDVDLIPMNDHNTYRCFSQPRHISVAMD
KFGLSLPYVQYFGGVSALSKQQFLSINGFPNNYWGWGGEDDDIYNRLAFRGMSVSRPKCCDREV
SDDPPLERQEK**Stop**

Appendix 8. pRSET(CscO)-RFP-GALT_CscR plasmid (plasmid 8) full sequence analysed by Sanger sequencing using constructed primers (see appendix 1). RFP gene is highlighted in dark-red, B4GALT1 gene is highlighted in yellow, CscR gene is highlighted in blue, important restriction sites are red coloured, primer sites are highlighted in green. Ampicillin resistance gene is highlighted in dark-green. $T7_{CscO}$ promoter is coloured in green for promoter site and in red for operator site.

GATCTCGATCCCGCGAAATTAATACGACTCACTATAGGTTAACGTTAACGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAG
AAGGAGATATACATATGCGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCTGTACGACGA
TGACGATAAGGATCGATGGGATCCGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGC
TCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCC
CCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCT
GTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTCCAGGACGGCGAG
TTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGA
TGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAA
GGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTAC
GAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGCATATGATGCGCTCGCTGACCGCATGCCCTGAGGAGTCCCCGCTGC
TCGTCGGCCCCATGCTGATTGAGTTTAACATACCTGTGGACCTGAAGCTTGTGGAGCAGCAGAACCCGAAGGTGAAGTTGGGTGGTCGCTACACCCC
CATGGACTGCATCTCTCCTCACAAGGTGGCCATCATCATTCCATTCCGCAACCGGCAGGAACACCTCAAGTACTGGCTGTATTACTTGCACCCAATC
CTACAGCGTCAGCAGTTAGACTATGGCATCTATGTTATCAACCAGGCTGGAGAGTCCATGTTCAACCGCGCAAAGCTCCTCAATGTTGGCTTTAAAG
AGGCCTTGAAGGACTATGACTACAACTGCTTTGTGTTTAGCGATGTGGACCTCATCCCAATGAACGACCATAACACCTACAGGTGCTTTTCACAGCC
ACGGCACATTTCTGTAGCAATGGATAAGTTTGGACTTAGCCTACCTTACGTGCAGTATTTTGGAGGTGTCTCTGCTCTAAGTAAACAACAGTTTCTC
AGCATCAATGGATTTCCTAATAACTACTGGGGCTGGGGAGGTGAAGATGATGACATTTATAACAGATTAGCTTTTAGAGGCATGTCTGTGTCTCGCC
CCAAATGCTGTGATCGGGAAGTGTCGGATGATCCGCCACTCGAGAGACAAGAAAAATGAGGTACCATGGAATTCGAAGCTTGATCCGGCTGCTAACA
AAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCT
GAAAGGAGGAACTATATCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACG
CGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTT
CTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTC
GACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTA
ATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTT
AAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAA
CCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAG
TATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAA
GATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTC
CAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCA
GAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGAT
AACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATC
GTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGG
CGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGC
TGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTA
TCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCA
AGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATC
CCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCT
TGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCA
GATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTA
CCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGG
GTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTC
GGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT
TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCAGGTGGAACAAC
GGATCAACAGCGGGCAAGGGATCCGCGTCACTCTTCCCCCTTCACGACCTTCAATAATATGCAATGCAGCTTCCCGCCCGATAATGTCATGTGGAAG
CTGAATTGTGGTCAGCGGCGGTAAAAACAGATGCCCGACGCCAACCAGATTATCAAAGCCCATTACGGCGACATCCTGCGGGATACGTACCCCCTTC
GCCAAAAGAACCTGATAAGCCACAAAGGCTGCGCGATCGTTACCACATATCAGAACATCAAAATCTGGTTTGCCCGATTTGAAGTGGGCATTGAGTA
AACTTGCGAGATCGGTGTAGTGATCATCACCTGTTGCCATGTGAAATTGTTTCACCTCAGCCAGATCTCGTCCAGCATCACGCCAGGCCTGCTCAAA
TCCCTGCCGACGATACCCTGTTGCCAACGCACTTTCCGGTAGCCAGAAGCATAACGGTTGACGATAGCCCGCCGCGAGCAAATGCTGTGTTGATTCA
TATTGTGCAGTGTAATCATCAGGGATATAACTGGGTAACGCTGGGTCATCCGCCACACAGTTCGCCAATACAATATTTTCACCATACAGAGACTCAG
GCAGCGTGATATGTCGCAGCCCCATTGTAGTATAGATAATGCCATCCGGACGGTGGGCAAGCAGCTGACGTGCCGCGCGGGCAGCGTCATCTTCAGA
AAAAATATTGATTAAAAAACTATTCCAGCCGAACTCGCTGGCGGTTTGCTCAATGGCAAGCAGAATATCAACAGAGAAAGGAGTGGTAGCAGTGTCC
TGCGCCAGCACGGCGAGAGTCGACGGCTTACGTCCTTGAGCGCGCATCTTACGGGCGGAAAGATCAGGAACATAATTCAGGGTCTGGATTGCCTGCA
ATACGCGGTCACGCGTTGCAGGACGCACAGATTCTGCATTATGCATCACCCGGGAGACTGTCATCATCGACACTCCCGCCAGGCGTGCGACATCCTT
TAATGAAGCCATATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCACCATTCGATGGTGTCCTGGGCC
CATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCC
CAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAG

Appendix 9. CscR nucleotide sequence. GenBank: X81461.2[24]

```
ATGGCTTCATTAAAGGATGTCGCACGCCTGGCGGGAGTGTCGATGATGACAGTCTCCCGGGTGATGCATAATGCAGAATCTGTGCGTCCTGCAACGC
GTGACCGCGTATTGCAGGCAATCCAGACCCTGAATTATGTTCCTGATCTTTCCGCCCGTAAGATGCGCGCTCAAGGACGTAAGCCGTCGACTCTCGC
CGTGCTGGCGCAGGACACTGCTACCACTCCTTTCTCTGTTGATATTCTGCTTGCCATTGAGCAAACCGCCAGCGAGTTCGGCTGGAATAGTTTTTTA
ATCAATATTTTTTCTGAAGATGACGCTGCCCGCGCGGCACGTCAGCTGCTTGCCCACCGTCCGGATGGCATTATCTATACTACAATGGGGCTGCGAC
ATATCACGCTGCCTGAGTCTCTGTATGGTGAAAATATTGTATTGGCGAACTGTGTGGCGGATGACCCAGCGTTACCCAGTTATATCCCTGATGATTA
CACTGCACAATATGAATCAACACAGCATTTGCTCGCGGCGGGCTATCGTCAACCGTTATGCTTCTGGCTACCGGAAAGTGCGTTGGCAACAGGGTAT
CGTCGGCAGGGATTTGAGCAGGCCTGGCGTGATGCTGGACGAGATCTGGCTGAGGTGAAACAATTTCACATGGCAACAGGTGATGATCACTACACCG
ATCTCGCAAGTTTACTCAATGCCCACTTCAAATCGGGCAAACCAGATTTTGATGTTCTGATATGTGGTAACGATCGCGCAGCCTTTGTGGCTTATCA
GGTTCTTTTGGCGAAGGGGGTACGTATCCCGCAGGATGTCGCCGTAATGGGCTTTGATAATCTGGTTGGCGTCGGGCATCTGTTTTTACCGCCGCTG
ACCACAATTCAGCTTCCACATGACATTATCGGGCGGGAAGCTGCATTGCATATTATTGAAGGTCGTGAAGGGGGAAGAGTGACGCGGATCCCTTGCC
CGCTGTTGATCCGTTGTTCCACCTGA
```

Appendix 10. Amino acid sequence CscR protein from pRSET(CscO)-RFP-GALT_CscR plasmid. DNA binding domain is highlighted in green.

MASLKDVARLAGVSMMTVSRVMHNAESVRPATRDRVLQAIQTLNYVPDLSARKMRAQGRKPSTL
AVLAQDTATTPFSVDILLAIEQTASEFGWNSFLINIFSEDDAARAARQLLAHRPDGIIYTTMGL
RHITLPESLYGENIVLANCVADDPALPSYIPDDYTAQYESTQHLLAAGYRQPLCFWLPESALAT
GYRRQGFEQAWRDAGRDLAEVKQFHMATGDDHYTDLASLLNAHFKSGKPDFDVLICGNDRAAFV
AYQVLLAKGVRIPQDVAVMGFDNLVGVGHLFLPPLTTIQLPHDIIGREAALHIIEGREGGRVTR
IPCPLLIRCST

T7 promoter sequence TAATACGACTCACTATA

LacI operator sequence AATTGTGAGCGGATAACAATT

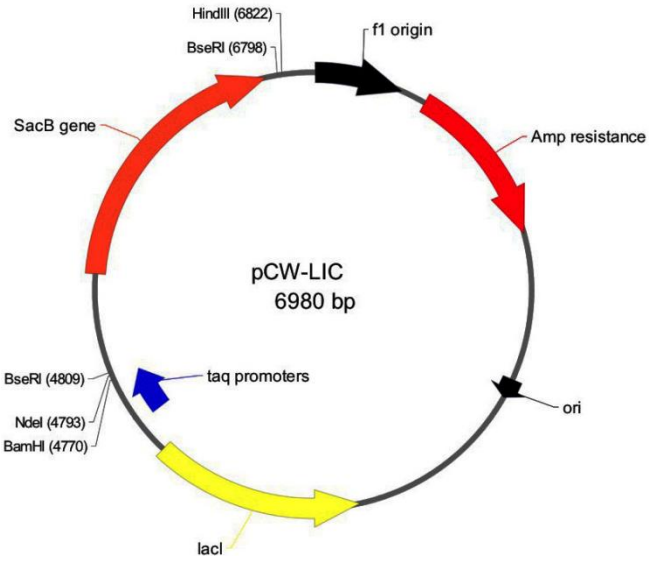
CscR operator sequence AGTTAACGTTAACA[60]

The pCW-LIC vector contains the same LacIq promoter sequence ported in the first LacIq publication.

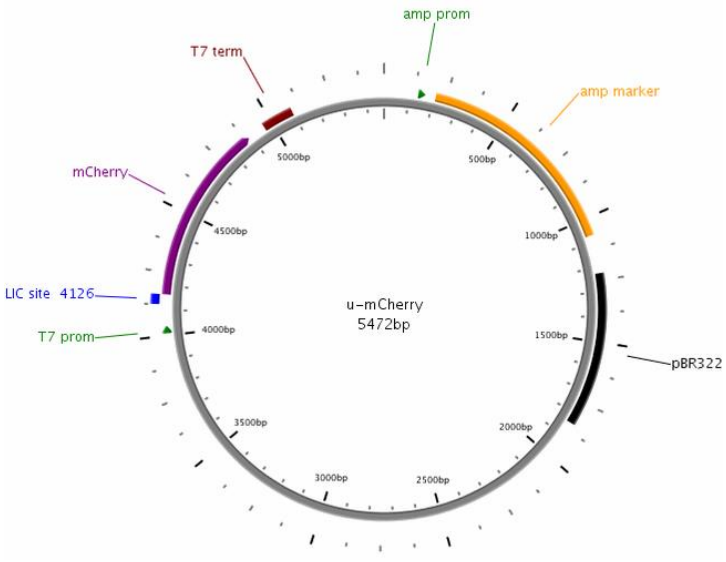
LacIq promoter boxes

GAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGG
TGAAT

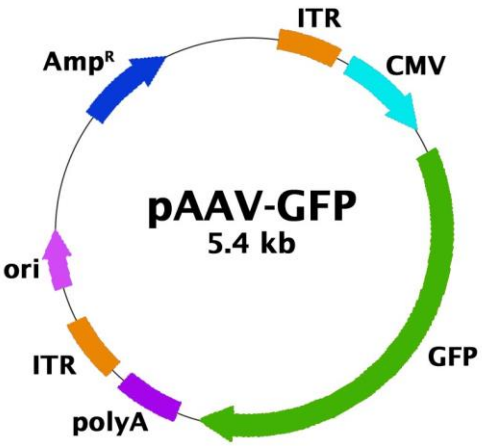
Appendix 11. Plasmids used in this project



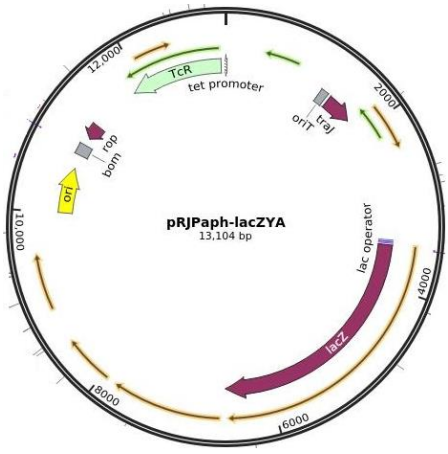
Plasmid 1. The pCW-LIC plasmid (#26098) from Addgene used as a template to make the T7CscO plasmid. It was also used to construct the pCW-RFP-GALT plasmid.



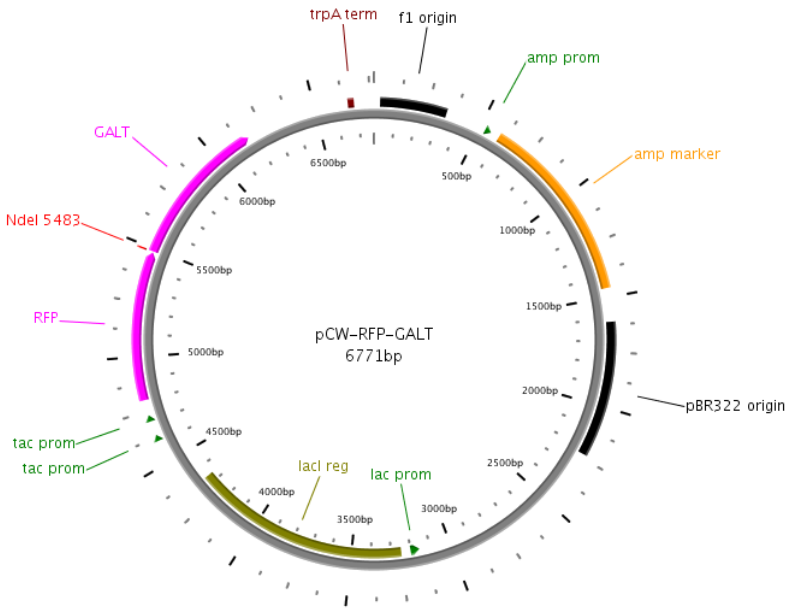
Plasmid 2. The pET mCherry LIC plasmid (#29769) was used to amplify RFP gene for design of fusion RFP-GALT gene.



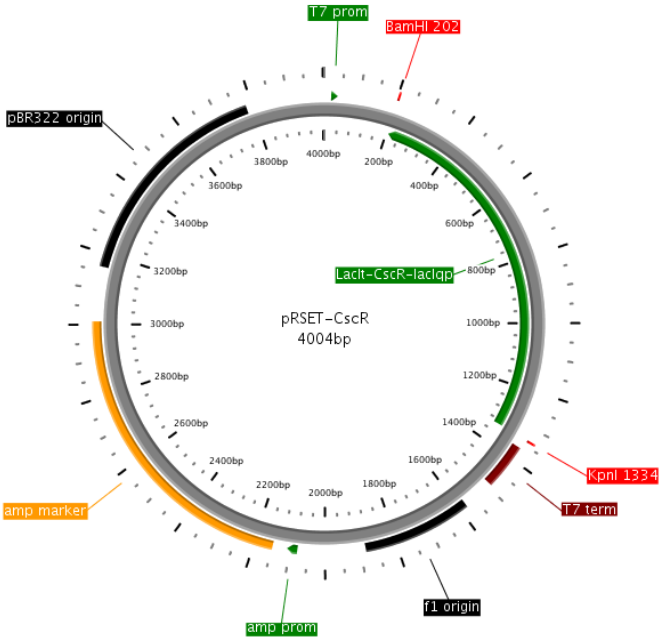
Plasmid 3. The pAAV-GFP[102] plasmid (Cell Biolabs, San Diego, California, United States) was used to amplify GFP gene for LacZ-GFP plasmid (data not shown) and planned to be used for future part of Csc/Lac oscillator design.



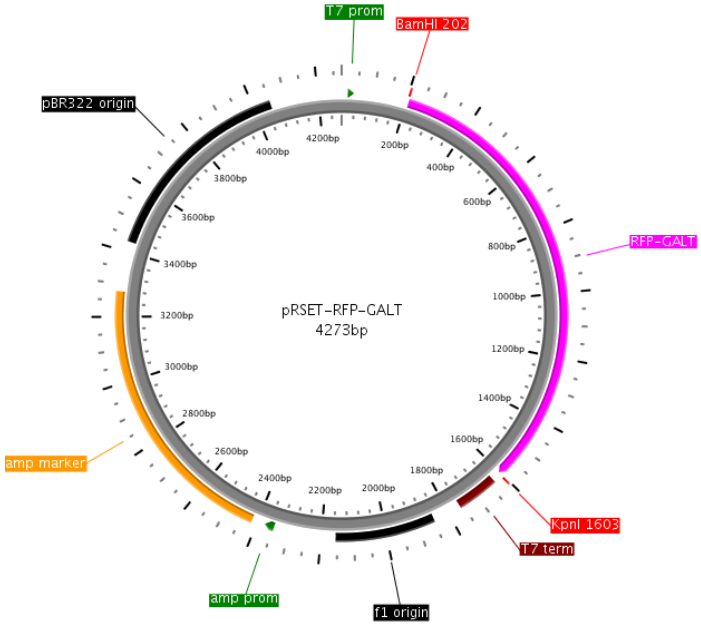
Plasmid 4. The pRJPaph-lacZYA plasmid[103] (#67037) from Addgene was used to insert GFP gene after LacZ gene; however, this was not successful probably because of the big vector size and inefficient transformation of competent cell with ligation products.



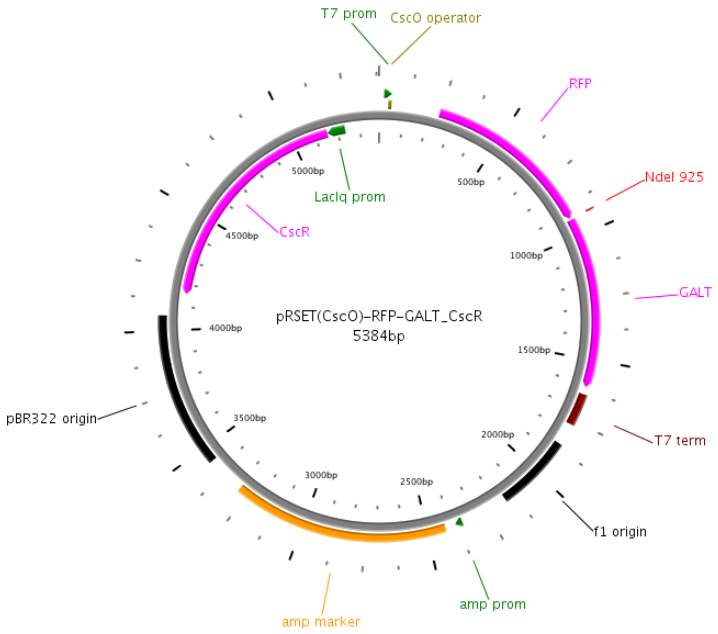
Plasmid 5. The pCW-RFP-GALT plasmid was used to study auto induction and lactose synthesis inside *E. coli*.



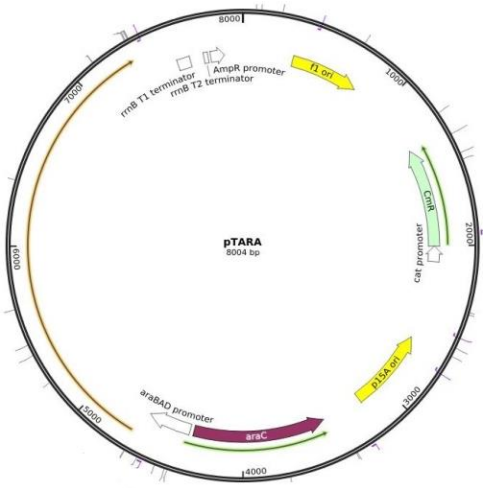
Plasmid 6. Synthesis of the reverse CscR gene under the control of the LacIq promoter and terminator in the pRSET vector was required for this research. Firstly, the pRSET contains a T7 promoter and terminator, and it is not very big (3 kb without insert). Secondly, it allows the easy cutting of the CscR gene following amplification by PCR. RFP-GALT gene was inserted instead of cut CscR gene and reverse CscR gene was inserted before T7 promoter (plasmid 8). The pRSET vector was supplied by Invitrogen as a part of the gene synthesis service.



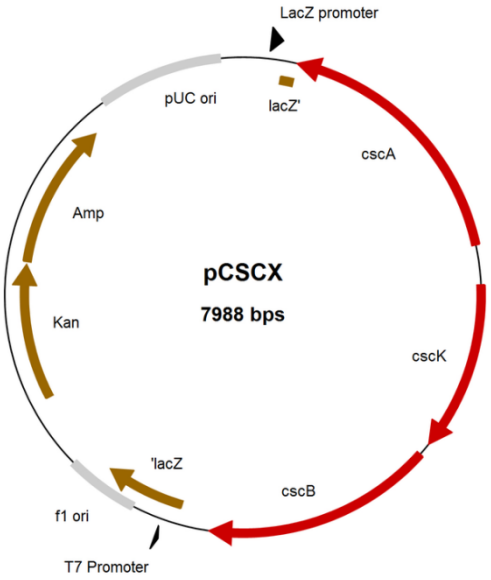
Plasmid 7. The pRSET-RFP-GALT plasmid requires T7 polymerase to synthesise the genes. As a result, this plasmid offers good control because fluorescence would only be expected when cells are producing T7 RNAP. The next step is site-directed mutagenesis so that the T7 promoter is regulated by sucrose.



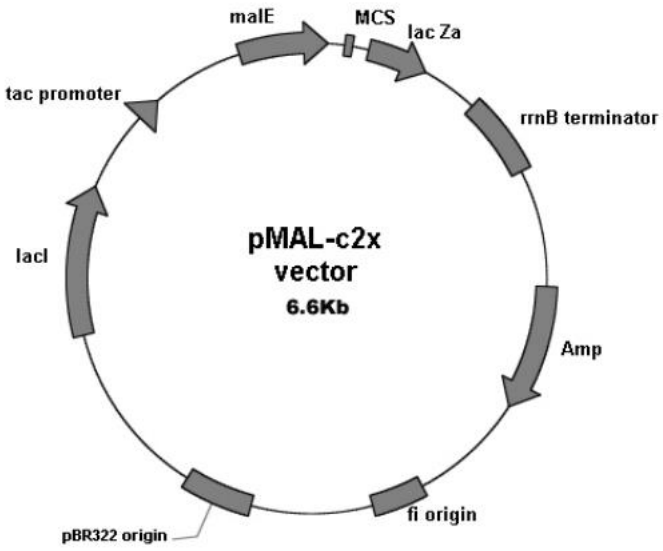
Plasmid 8. The pRSET(CscO)-RFP-GATL_CscR contains the sucrose operator sequence after the T7 promoter. CscR gene synthesised previously in the reverse position of pRSET vector, was inserted before the T7 promoter.



Plasmid 9. The pTara plasmid[104] (#31491) from Addgene. This arabinose-inducible plasmid containing T7 RNA polymerase gene was transformed into the deltaLacZYA strain by electroporation.



Plasmid 10. The pCSCX plasmid[25] (#63918) from Addgene. This plasmid was used to amplify CscB gene to test active sucrose transport in *E. coli.* However, the ligation of CscB inside pRSET(CscO)-RFP-GALR_CscR vector was not successful.



Plasmid 11. The pMAL-c2x plasmid[105] (#75286) from Addgene. This plasmid was used by Dr James Errey. B4GALT1 gene was amplified from this plasmid in this project.

7. References

1. Collins, P. M.; Ferrier, R. J.; Berlin, W., *Monosaccharides: their chemistry and their roles in natural products*. Wiley Chichester: 1995.
2. Breimer, M. E.; Hansson, G. C.; Karlsson, K.-A.; Leffler, H.; Pimlott, W.; Samuelsson, B. E., Sequencing of oligosaccharides by mass spectrometry applied on a 12-sugar glycolipid. *FEBS letters* **1981,** *124* (2), 299-303.
3. Medeiros, P. M.; Simoneit, B. R., Analysis of sugars in environmental samples by gas chromatography–mass spectrometry. *Journal of Chromatography A* **2007,** *1141* (2), 271-278.
4. Both, P.; Green, A.; Gray, C.; Šardzík, R.; Voglmeir, J.; Fontana, C.; Austeri, M.; Rejzek, M.; Richardson, D.; Field, R., Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing. *Nature chemistry* **2014,** *6* (1), 65.
5. Deng, Y.; Chen, L.-X.; Han, B.-X.; Wu, D.-T.; Cheong, K.-L.; Chen, N.-F.; Zhao, J.; Li, S.-P., Qualitative and quantitative analysis of specific polysaccharides in Dendrobium huoshanense by using saccharide mapping and chromatographic methods. *Journal of pharmaceutical and biomedical analysis* **2016,** *129*, 163-171.
6. Varki, A.; Sharon, N., Historical background and overview. **2009**.
7. Crews, P.; Rodriquez, J.; Jaspars, M.; Crews, R. J., Organic structure analysis. **1998**.
8. Sprinzak, D. and Elowitz, M.B., Reconstruction of genetic circuits. *Nature* **2005.** 438(7067), 443.
9. Halford, N. G.; Purcell, P. C.; Hardie, D. G., Is hexokinase really a sugar sensor in plants? *Trends Plant Sci* **1999,** *4* (3), 117-120.
10. Wall, M. E.; Hlavacek, W. S.; Savageau, M. A., Design of gene circuits: lessons from bacteria. *Nat Rev Genet* **2004,** *5* (1), 34-42.
11. Koonin, E. V.; Wolf, Y. I., Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic acids research* **2008,** *36* (21), 6688-6719.
12. Brogna, S.; Ashburner, M., The Adh-related gene of Drosophila melanogaster is expressed as a functional dicistronic messenger RNA: multigenic transcription in higher organisms. *The EMBO journal* **1997,** *16* (8), 2023-2031.
13. Blumenthal, T., Trans-splicing and polycistronic transcription in Caenohabditis elegans. *Trends in Genetics* **1995,** *11* (4), 132-136.
14. Lewis, M.; Chang, G.; Horton, N. C.; Kercher, M. A.; Pace, H. C.; Schumacher, M. A.; Brennan, R. G.; Lu, P., Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **1996,** *271* (5253), 1247-1254.
15. Hirooka, K.; Kodoi, Y.; Satomura, T.; Fujita, Y., Regulation of the rhaEWRBMA Operon Involved in l-Rhamnose Catabolism through Two Transcriptional Factors, RhaR and CcpA, in Bacillus subtilis. *J Bacteriol* **2015,** *198* (5), 830-45.
16. Schlegel, H. G.; Zaborosch, C., *General microbiology*. Cambridge university press: 1993.
17. Schleif, R., Regulation of the L-arabinose operon of Escherichia coli. *Trends in Genetics* **2000,** *16* (12), 559-565.
18. Takeda, Y.; Takase, K.; Yamato, I.; Abe, K., Sequencing and characterization of the xyl operon of a gram-positive bacterium, Tetragenococcus halophila. *Applied and environmental microbiology* **1998,** *64* (7), 2513-2519.
19. Wilson, D. B.; Hogness, D. S., The enzymes of the galactose Operon in Escherichia coli I. Purification and characterization of uridine diphosphogalactose 4-epimerase. *Journal of Biological Chemistry* **1964,** *239* (8), 2469-2482.
20. Kim, C.; Song, S.; Park, C., The D-allose operon of Escherichia coli K-12. *Journal of bacteriology* **1997,** *179* (24), 7631-7637.

21. Biedendieck, R.; Gamer, M.; Jaensch, L.; Meyer, S.; Rohde, M.; Deckwer, W. D.; Jahn, D., A sucrose-inducible promoter system for the intra- and extracellular protein production in Bacillus megaterium. *J Biotechnol* **2007,** *132* (4), 426-30.
22. Penfold, D.; Macaskie, L., Production of H2 from sucrose by Escherichia coli strains carrying the pUR400 plasmid, which encodes invertase activity. *Biotechnology letters* **2004,** *26* (24), 1879-1883.
23. Schmid, K.; Ebner, R.; Altenbuchner, J.; Schmitt, R.; Lengeler, J., Plasmid-mediated sucrose metabolism in Escherichia coli K12: mapping of the scr genes of pUR400. *Molecular microbiology* **1988,** *2* (1), 1-8.
24. Jahreis, K.; Bentler, L.; Bockmann, J.; Hans, S.; Meyer, A.; Siepelmeyer, J.; Lengeler, J. W., Adaptation of Sucrose Metabolism in the Escherichia coli Wild-Type Strain EC3132 *Journal of Bacteriology* **2002,** *184* (19), 5307-5316.
25. Bruschi, M.; Boyes, S. J.; Sugiarto, H.; Nielsen, L. K.; Vickers, C. E., A transferable sucrose utilization approach for non-sucrose-utilizing Escherichia coli strains. *Biotechnology advances* **2012,** *30* (5), 1001-1010.
26. Lunn, J. E., Evolution of sucrose synthesis. *Plant physiology* **2002,** *128* (4), 1490-1500.
27. Salerno, G. L.; Curatti, L., Origin of sucrose metabolism in higher plants: when, how and why? *Trends in plant science* **2003,** *8* (2), 63-69.
28. Arifin, Y.; Sabri, S.; Sugiarto, H.; Kromer, J. O.; Vickers, C. E.; Nielsen, L. K., Deletion of cscR in Escherichia coli W improves growth and poly-3-hydroxybutyrate (PHB) production from sucrose in fed batch culture. *J Biotechnol* **2011,** *156* (4), 275-8.
29. Luesink, E. J.; Marugg, J. D.; Kuipers, O. P.; De Vos, W. M., Characterization of the Divergent sacBKand sacAR Operons, Involved in Sucrose Utilization byLactococcus lactis. *Journal of bacteriology* **1999,** *181* (6), 1924-1926.
30. Weickert, M. J.; Adhya, S., A family of bacterial regulators homologous to Gal and Lac repressors. *Journal of Biological Chemistry* **1992,** *267* (22), 15869-15874.
31. Steen, J. A.; Bohlke, N.; Vickers, C. E.; Nielsen, L. K., The trehalose phosphotransferase system (PTS) in E. coli W can transport low levels of sucrose that are sufficient to facilitate induction of the csc sucrose catabolism operon. *PLoS One* **2014,** *9* (2), e88688.
32. Weickert, M. J.; Adhya, S., The galactose regulon of Escherichia coli. *Molecular microbiology* **1993,** *10* (2), 245-251.
33. Malan, T. P.; Kolb, A.; Buc, H.; McClure, W. R., Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. *Journal of molecular biology* **1984,** *180* (4), 881-909.
34. Stricker, J.; Cookson, S.; Bennett, M. R.; Mather, W. H.; Tsimring, L. S.; Hasty, J., A fast, robust and tunable synthetic gene oscillator. *Nature* **2008,** *456* (7221), 516-9.
35. Lutz, R.; Bujard, H., Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research* **1997,** *25* (6), 1203-1210.
36. Studier, F. W.; Moffatt, B. A., Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology* **1986,** *189* (1), 113-130.
37. Elroy-Stein, O.; Moss, B., Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells. *Proceedings of the National Academy of Sciences* **1990,** *87* (17), 6743-6747.
38. Cermakian, N.; Ikeda, T. M.; Cedergren, R.; Gray, M. W., Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic acids research* **1996,** *24* (4), 648-654.

39. Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R., Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **2004,** *428* (6980), 281.
40. Purnick, P. E.; Weiss, R., The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* **2009,** *10* (6), 410-22.
41. Karzai, A. W.; Roche, E. D.; Sauer, R. T., The SsrA–SmpB system for protein tagging, directed degradation and ribosome rescue. *Nature Structural & Molecular Biology* **2000,** *7* (6), 449-455.
42. Potvin-Trottier, L.; Lord, N. D.; Vinnicombe, G.; Paulsson, J., Synchronous long-term oscillations in a synthetic gene circuit. *Nature* **2016,** *538* (7626), 514-517.
43. Herzenberg, L. A., Studies on the induction of β-galactosidase in a cryptic strain of Escherichia coli. *Biochimica et biophysica acta* **1959,** *31* (2), 525-538.
44. Jobe, A.; Bourgeois, S., Lac repressor-operator interaction: VI. The natural inducer of the lac operon. *Journal of molecular biology* **1972,** *69* (3), 397IN7405-404IN8408.
45. Santillán, M.; Mackey, M. C., Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. *Biophysical journal* **2004,** *86* (3), 1282-1292.
46. Gilbert, W.; Müller-Hill, B., Isolation of the lac repressor. *Proceedings of the National Academy of Sciences* **1966,** *56* (6), 1891-1898.
47. Marbach, A.; Bettenbrock, K., lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *Journal of biotechnology* **2012,** *157* (1), 82-88.
48. Figge, J.; Wright, C.; Collins, C. J.; Roberts, T. M.; Livingston, D. M., Stringent regulation of stably integrated chloramphenicol acetyl transferase genes by E. coli lac repressor in monkey cells. *Cell* **1988,** *52* (5), 713-722.
49. Wong, C. H.; Halcomb, R. L.; Ichikawa, Y.; Kajimoto, T. J. A. C. I. E. i. E., Enzymes in organic synthesis: application to the problems of carbohydrate recognition (part 1). **1995,** *34* (4), 412-432.
50. Compain, P.; Martin, O. R. J. B.; chemistry, m., Carbohydrate mimetics-based glycosyltransferase inhibitors. **2001,** *9* (12), 3077-3092.
51. Yang, M.; Proctor, M. R.; Bolam, D. N.; Errey, J. C.; Field, R. A.; Gilbert, H. J.; Davis, B. G., Probing the breadth of macrolide glycosyltransferases: in vitro remodeling of a polyketide antibiotic creates active bacterial uptake and enhances potency. *Journal of the American Chemical Society* **2005,** *127* (26), 9336-9337.
52. Ramakrishnan, B.; Qasba, P. K., UDP-Gal: BetaGlcNAc Beta 1, 4-Galactosyltransferase, Polypeptide 1 (B4GALT1). In *Handbook of Glycosyltransferases and Related Genes*, Springer: 2014; pp 51-62.
53. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The protein data bank. *Nucleic acids research* **2000,** *28* (1), 235-242.
54. Brew, K.; Vanaman, T. C.; Hill, R. L., The role of alpha-lactalbumin and the A protein in lactose synthetase: a unique mechanism for the control of a biological reaction. *Proceedings of the National Academy of Sciences* **1968,** *59* (2), 491-497.
55. Ramakrishnan, B.; Qasba, P. K., Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the β1, 4-galactosyltransferase-I. *Journal of molecular biology* **2001,** *310* (1), 205-218.
56. Reid, S. J.; Abratt, V. R., Sucrose utilisation in bacteria: genetic organisation and regulation. *Appl Microbiol Biotechnol* **2005,** *67* (3), 312-21.
57. Nguyen, C. C.; Saier Jr, M. H., Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. *FEBS letters* **1995,** *377* (2), 98-102.

58. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, Springer: 2005; pp 571-607.
59. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *Journal of molecular biology* **1990,** *215* (3), 403-410.
60. Porcheron, G.; Kut, E.; Canepa, S.; Maurel, M. C.; Schouler, C., Regulation of fructooligosaccharide metabolism in an extra-intestinal pathogenic Escherichia coli strain. *Mol Microbiol* **2011,** *81* (3), 717-33.
61. Camas, F. M.; Alm, E. J.; Poyatos, J. F., Local gene regulation details a recognition code within the LacI transcriptional factor family. **2010,** *6* (11), e1000989.
62. Arifin, Y.; Archer, C.; Lim, S.; Quek, L.-E.; Sugiarto, H.; Marcellin, E.; Vickers, C. E.; Krömer, J. O.; Nielsen, L. K., Escherichia coli W shows fast, highly oxidative sucrose metabolism and low acetate formation. *Applied microbiology and biotechnology* **2014,** *98* (21), 9033-9044.
63. Bruschi, M.; Krömer, J. O.; Steen, J. A.; Nielsen, L. K., Production of the short peptide surfactant DAMP4 from glucose or sucrose in high cell density cultures of Escherichia coli BL21 (DE3). *Microbial cell factories* **2014,** *13* (1), 99.
64. Sabri, S.; Nielsen, L. K.; Vickers, C. E., Molecular control of sucrose utilization in Escherichia coli W, an efficient sucrose-utilizing strain. *Applied and environmental microbiology* **2013,** *79* (2), 478-487.
65. Alaeddinoglu, N. G.; Charles, H., Transfer of a Gene for Sucrose Utilization into Escherichia colik12, and Consequent Failure of Expression of Genes for d-Serine Utilization. *Microbiology* **1979,** *110* (1), 47-59.
66. Bockmann, J.; Heuel, H.; Lengeler, J., Characterization of a chromosomally encoded, non-PTS metabolic pathway for sucrose utilization in Escherichia coli EC3132. *Molecular and General Genetics MGG* **1992,** *235* (1), 22-32.
67. Löwe, H.; Sinner, P.; Kremling, A.; Pflüger-Grau, K., Engineering sucrose metabolism in Pseudomonas putida highlights the importance of porins. *Microbial biotechnology* **2018**.
68. Harding, S. E.; Tombs, M. P.; Adams, G. G.; Paulsen, B. S.; Inngjerdingen, K. T.; Barsett, H., *An introduction to polysaccharide biotechnology*. CRC Press: 2017.
69. Flowers, H. M., [7] Chemical synthesis of oligosaccharides. In *Methods in enzymology*, Elsevier: 1978; Vol. 50, pp 93-121.
70. Elling, L., Glycobiotechnology: enzymes for the synthesis of nucleotide sugars. In *New Enzymes for Organic Synthesis*, Springer: 1997; pp 89-144.
71. Zhou, Y.; Lin, F.; Chen, J.; Yu, B., Toward synthesis of the regular sequence of heparin: synthesis of two tetrasaccharide precursors. *Carbohydrate research* **2006,** *341* (10), 1619-1629.
72. Munoz, E.; Xu, D.; Avci, F.; Kemp, M.; Liu, J.; Linhardt, R. J., Enzymatic synthesis of heparin related polysaccharides on sensor chips: rapid screening of heparin–protein interactions. *Biochemical and biophysical research communications* **2006,** *339* (2), 597-602.
73. Qasba, P. K.; Ramakrishnan, B.; Boeggeman, E., Substrate-induced conformational changes in glycosyltransferases. *Trends in biochemical sciences* **2005,** *30* (1), 53-62.
74. Esko, J. D.; Bertozzi, C.; Schnaar, R. L., Chemical tools for inhibiting glycosylation. **2017**.
75. Kim, S.; Miura, Y.; Etchison, J. R.; Freeze, H. H., Intact Golgi synthesize complex branched *O*-linked chains on glycoside primers: evidence for the functional continuity of seven glycosyltransferases and three sugar nucleotide transporters. *Glycoconjugate journal* **2001,** *18* (8), 623-633.
76. Etchison, J. R.; Freeze, H. H., A new approach to mapping co-localization of multiple glycosyl transferases in functional Golgi preparations. *Glycobiology* **1996,** *6* (2), 177-189.

77. Uehara, T.; Park, J. T., The N-acetyl-D-glucosamine kinase of Escherichia coli and its role in murein recycling. *Journal of bacteriology* **2004,** *186* (21), 7273-7279.
78. Kalivoda, K. A.; Steenbergen, S. M.; Vimr, E. R.; Plumbridge, J., Regulation of sialic acid catabolism by the DNA binding protein NanR in Escherichia coli. *Journal of bacteriology* **2003,** *185* (16), 4806-4815.
79. López-Abente, G.; Aragonés, N.; Pollán, M.; Ruiz, M.; Gandarillas, A., Leukemia, lymphomas, and myeloma mortality in the vicinity of nuclear power plants and nuclear fuel facilities in Spain. *Cancer Epidemiology and Prevention Biomarkers* **1999,** *8* (10), 925-934.
80. De Boer, H. A.; Comstock, L. J.; Vasser, M., The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences* **1983,** *80* (1), 21-25.
81. Bren, A.; Park, J. O.; Towbin, B. D.; Dekel, E.; Rabinowitz, J. D.; Alon, U., Glucose becomes one of the worst carbon sources for E. coli on poor nitrogen sources due to suboptimal levels of cAMP. *Scientific Reports* **2016,** *6*, 24834.
82. Mao, Z.; Shin, H. D.; Chen, R. R., Engineering the E. coli UDP-Glucose Synthesis Pathway for Oligosaccharide Synthesis. *Biotechnology progress* **2006,** *22* (2), 369-374.
83. Anderson, A.; Cooper, R., Gluconeogenesis in Escherichia coli The role of triose phosphate isomerase. *FEBS letters* **1969,** *4* (1), 19-20.
84. Mukherji, S. and Van Oudenaarden, A., Synthetic biology: understanding biological design from synthetic circuits. *Nature Reviews Genetics* **2009** 10(12), 859.
85. Huang, L.; Yuan, Z.; Liu, P.; Zhou, T., Effects of promoter leakage on dynamics of gene expression. *BMC systems biology* **2015,** *9* (1), 16.
86. Woyski, D.; Cupp-Vickery, J. R. J. A. o. B.; Biophysics, Enhanced expression of cytochrome P450s from lac-based plasmids using lactose as the inducer. **2001,** *388* (2), 276-280.
87. Pinevich, A.; Averina, S.; Velichko, N., Essays in the biology of prochlorophytes. **2010**.
88. Zimmermann, G.; Kelly, G. J.; Latzko, E., Purification and properties of spinach leaf cytoplasmic fructose-1, 6-bisphosphatase. *Journal of Biological Chemistry* **1978,** *253* (17), 5952-5956.
89. Curatti, L.; Folco, E.; Desplats, P.; Abratti, G.; Limones, V.; Herrera-Estrella, L.; Salerno, G., Sucrose-phosphate synthase from Synechocystis sp. strain PCC 6803: identification of the spsA gene and characterization of the enzyme expressed in Escherichia coli. *Journal of bacteriology* **1998,** *180* (24), 6776-6779.
90. Fischer, E.; Sauer, U., Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS. *European Journal of Biochemistry* **2003,** *270* (5), 880-891.
91. Sahota, S. S.; Bramley, P. M.; Menzies, I. S., The fermentation of lactulose by colonic bacteria. *Microbiology* **1982,** *128* (2), 319-325.
92. Shis, D. L.; Bennett, M. R., Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proceedings of the National Academy of Sciences* **2013,** *110* (13), 5028-5033.
93. Eames, M.; Kortemme, T., Cost-benefit tradeoffs in engineered lac operons. *Science* **2012,** *336* (6083), 911-915.
94. Dubendorf, J. W.; Studier, F. W., Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *Journal of molecular biology* **1991,** *219* (1), 45-59.
95. Jin, Q.; Yuan, Z.; Xu, J.; Wang, Y.; Shen, Y.; Lu, W.; Wang, J.; Liu, H.; Yang, J.; Yang, F., Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. *Nucleic acids research* **2002,** *30* (20), 4432-4441.

96. Deng, W.; Burland, V.; Plunkett III, G.; Boutin, A.; Mayhew, G. F.; Liss, P.; Perna, N. T.; Rose, D. J.; Mau, B.; Zhou, S., Genome sequence of Yersinia pestis KIM. *Journal of bacteriology* **2002,** *184* (16), 4601-4611.
97. Calos, M. P., DNA sequence for a low-level promoter of the lac repressor gene and an 'up'promoter mutation. *Nature* **1978,** *274* (5673), 762.
98. Scheie, P. O., Plasmolysis of Escherichia coli B/r with sucrose. **1969,** *98* (2), 335-340.
99. Welsh, K. J.; Kirkman, M. S.; Sacks, D. B., Role of glycated proteins in the diagnosis and management of diabetes: research gaps and future directions. **2016,** *39* (8), 1299-1306.
100. Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. and Remington, S.J., Novel chromophores and buried charges control color in mFruits. *Biochemistry* 2006 45(32), 9639-9647.
101. Kamps-Hughes, N.; Quimby, A.; Zhu, Z.; Johnson, E. A., Massively parallel characterization of restriction endonucleases. **2013,** *41* (11), e119-e119.
102. Grimm, D.; Lee, J. S.; Wang, L.; Desai, T.; Akache, B.; Storm, T. A.; Kay, M. A., In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *Journal of virology* **2008,** *82* (12), 5887-5911.
103. Ledermann, R.; Bartsch, I.; Remus-Emsermann, M. N.; Vorholt, J. A.; Fischer, H.-M., Stable fluorescent and enzymatic tagging of Bradyrhizobium diazoefficiens to analyze host-plant infection and colonization. *Molecular Plant-Microbe Interactions* **2015,** *28* (9), 959-967.
104. Wycuff, D. R.; Matthews, K. S., Generation of an AraC-araBAD promoter-regulated T7 expression system. *Analytical biochemistry* **2000,** *277* (1), 67-73.
105. Walker, I.H., Hsieh, P.C. and Riggs, P.D., Mutations in maltose-binding protein that alter affinity and solubility properties. *Applied microbiology and biotechnology* **2010,** 88(1), 187-197.

Statement of originality

This Msc thesis entitled “Sugar sensing and amplification using bacterial operons” has been written by Artem Shesterikov under the supervision by Professor Benjamin G. Davis and Professor Andrew Baldwin. The English language grammar of this thesis was edited by Oxford Science Editing Limited Company.