1. Engineering Synthetic Mammalian Expression Vectors
2. A Bacterial Strategy to Link Calcium Influx to Cell Survival

Brittany Stewart

Thesis submitted in partial fulfillment of the requirements for the Master's degree in Chemistry

Departments of Chemistry and Biomolecular Science

Faculty of Science

University of Ottawa

Candidate

Brittany Stewart

Supervisor

Professor Corrie DaCosta

Abstract

(1) The field of synthetic biology is rapidly growing, empowered by advancements in Molecular Biology. To express genes of interest, scientists exploit plasmids engineered for bacterial or mammalian expression. Existing plasmids carry superfluous DNA that decreases transformation and transfection efficiencies. Here, we present a novel set of mammalian expression vectors with different selection markers and tunable expression levels. Despite being substantially smaller than traditional vectors, these minimized plasmids display similar levels of gene expression. This set of novel mammalian expression vectors should be useful for a broad range of modern applications.

(2) DREAM is a mammalian calcium-dependent repressor of gene expression, which binds to a downstream DNA sequence, called DRE. The DREAM/DRE system offers a unique way of linking gene expression to calcium influx (44), but the utility of this system has not been assessed in bacteria. Here we develop a simple bacterial cell growth assay, where cell growth is prevented by the expression of *SacB*. This cell growth assay is then exploited to assess the utility of the DREAM/DRE system in bacteria. Unexpectedly, the DRE sequence by itself represses the expression of *SacB*, which can then be de-repressed by the co-expression of DREAM. Should this recovery of *SacB* expression maintain calcium-dependence, the DREAM/DRE system could be exploited in future directed strategies to evolve calcium-permeable ion channels.

Acknowledgements

I would like to acknowledge the contributions of Professor Corrie daCosta, who supervised this research project, providing time and guidance ensuring its success. I would also like to extend gratitude to the members of the daCosta lab for their help, support, and advice throughout this project, as well as acknowledge Johnathan Emlaw, Christian Tessier, Ray LePabic, and Louise Connell for their contributions to the project.

Table of Contents

1.0 Preface

1.1.1 Overview of Synthetic Biology Projects

My master's degree involved two unrelated synthetic biology projects: (1) engineering synthetic mammalian expression vectors based on a set of minimal bacterial expression vectors and (2) development of a bacterial strategy to linking calcium influx to cell survival. The goal of the first project was to assemble a toolbox of vectors with tunable levels of protein expression and different selection markers, to be used in a wide range of synthetic biology applications.

My second project focused on setting up a genetic circuit in bacteria to link calcium influx to cell survival. This involved characterizing the expression of a death gene, and then assessing the effects of a mammalian calcium dependent repressor system on the expression of this death gene upon integration of the repressor system into the genetic circuit. In the future, this genetic circuit could be used in Directed Evolution strategies to select for calcium permeable ion channels with desirable functions.

1.1.2 Statement of Contributions

The work done in chapter 2 is a collaboration between members of the labs of Dr. Campbell-Valois and Dr. daCosta. My undergraduate honours project entailed engineering synthetic mammalian expression vectors based upon a set of minimal pUdO plasmid constructs. This set of bacterial pUdO constructs were assembled by France Manigat in the lab of Dr. Campbell-Valois. An intermediate pUdO plasmid, pUdO1a-CMV, used to make mammalian expression vectors was assembled by Ray LePabic. Upon starting, and throughout my MSc, I continued assembling and testing the mammalian versions of these. During my MSc, I assembled an additional set of vectors with a bovine growth hormone terminator and cloned a luciferase reporter gene into the full panel of vectors. A full list of vectors assembled prior to my MSc and during my MSc can be found in Appendix I.II, Table 6 and Table 7, respectively. Electrophysiology experiments were performed by Christian Tessier and Johnathan Emlaw. Luciferase assays and mammalian cell culture was performed by Louise Connel.

Chapter 1 Engineering Novel Synthetic Mammalian Expression Vectors

1.1.0. Introduction

1.1.1. Overview

Conventionally used mammalian expression vectors contain unnecessary excess DNA that decreases transfection efficiency (1). Based on a set of minimal bacterial expression vectors, we set out to engineer a toolbox of mammalian expression vectors with tunable levels of expression and different selection markers. This toolbox of mammalian expression vectors was based on an identical backbone structure allowing for easy cloning between vectors. I began working on this project during my undergraduate honours research project in the daCosta lab and continued into my master's degree. In my honours project, I developed set of mammalian expression vectors using the SV40 terminator and the CMV promoter and enhancer, with four vectors containing successive deletions to the CMV promoter and enhancer region. During my MSc, I created an additional set of vectors with the bovine growth hormone terminator, aimed to improve expression. I also cloned a luciferase reporter gene into all pUdO mammalian vectors to determine the relative level of gene expression of each vector. A full list of vectors assembled prior to, as well as during, my MSc degree can be found in Tables 6 and 7 in Appendix I.II.

1.1.2. Advancements in Synthetic Biology

Synthetic biology allows scientists to take advantage of biological systems that exist in nature and alter them to achieve many goals. Synthetic biology allows for the redesign of

natural constructs in the pursuit of developing new therapeutic and diagnostic agents or interrogating the natural world.

Modern Cloning Methods

Modern cloning methods rely on the polymerase chain reaction (PCR), which facilitates the inexpensive amplification of almost any DNA sequence. With the invention of PCR (2), small quantities of DNA can be amplified exponentially using oligonucleotide primers, designed based on prior sequencing data. PCR allows for techniques like site-directed mutagenesis to change specific nucleotides in a given DNA sequence. Other DNA manipulation techniques like Gibson assembly, allow for multiple DNA fragments to be ligated to one another in a specified way. These simply yet powerful techniques are enough to not only manipulate existing DNA sequences, but aid in the design of synthetic constructs allowing for the expression of almost any gene of interest.

Gene Synthesis.

Advances in gene synthesis have opened the door to greater freedom in designing new DNA molecules. The first gene-synthesis methods were time consuming and expensive; the first 77 bp of DNA being synthesized over 5 years (3). Now DNA sequences longer than several thousand base pairs of DNA can be routinely synthesized in days (4).

Modern techniques in synthetic biology open the possibility to redesign and optimize bacterial and mammalian expression vectors, all the while minimizing non-essential DNA. Antiquated conventional plasmids tend to be larger in size, due to classical cloning techniques used to assemble them, as well as their natural origin (4). Synthetic

biology allows for plasmids to be constructed from the ground up, which only contain desired elements without any superfluous DNA 'carried over' from donor vectors. Here we describe a series of novel engineered synthetic plasmids for protein expression in mammalian cell lines, which are optimized for modern cloning techniques.

1.1.3. Plasmids

Plasmids, also called expression vectors, are essential tools used in synthetic biology. They allow for exogenous genes of interest to be expressed in bacterial or mammalian cells where they are endogenous. They also facilitate the expression of proteins in large quantities for downstream applications. Plasmids are double-stranded circular DNA molecules that are distinct from chromosomal DNA (5). Plasmids typically range in size from a few kilobases to hundreds of kilobases (6). Genes of interest can be cloned into plasmids and expressed in bacterial or mammalian cells rendering plasmids a pivotal component of the synthetic biology toolbox. All plasmids, both bacterial and mammalian, have elements essential for their maintenance and propagation. These elements include an origin of replication (ORI), and a selection marker (or markers). In addition, there are elements necessary for plasmids to express genes of interest: a multiple cloning site (MCS), a transcriptional promoter, and a terminator.

***Origin of replication*.**

In bacteria, the origin of replication (ORI) controls how many copies of the plasmid are synthesized during DNA replication, which occurs in three steps: initiation, elongation, and termination. The origin of replication, as its name suggests, is where replication of the

plasmid is initiated. Generally, the sequences constituting the ORIs are enriched in adenine and thymine nucleotides, owing to the fact that A-T base pairs only have two hydrogen bonds between them, making it easier to melt double stranded DNA in this region. The lower energy requirement for DNA melting facilitates the process of initiation as the double stranded DNA must separate to allow for replication to occur (7).

Selection markers.

Plasmids usually also contain a gene that encodes a selection marker, which provides antibiotic resistance when expressed. Often these plasmids with a selection marker are transformed into cells that are then grown in the presence of the antibiotic, eliminating cells that have not incorporated the desired plasmids containing the antibiotic resistant gene. For example, ampicillin, a beta-lactam, is a commonly used antibiotic that prevents bacterial cell wall synthesis (8,9). The ampicillin resistance gene expresses a beta-lactamase enzyme that hydrolyses the β-lactam ring of ampicillin preventing ampicillin from interfering with cell wall synthesis. As a result, cells transformed with plasmids containing the ampicillin resistance gene can be identified in the presence of ampicillin.

Multiple Cloning Site.

The multiple cloning site (MCS) contains recognition sites for a variety of restriction enzymes, enabling restriction cloning to insert genes of interest into the plasmid. The multiple cloning site is normally located downstream of a promoter, and upstream of a terminator, both of which are required for transcription of the inserted gene of interest (10).

Promoters and Terminators.

In both prokaryotes and eukaryotes, transcription requires promoters and terminators. However, as the transcriptional machinery used by prokaryotes and eukaryotes is different the promoters and terminators employed by prokaryotes and eukaryotes are different. Eukaryotic expression vectors/plasmids are created from existing bacterial constructs through the addition of eukaryotic-specific promoters and terminators (11).

1.1.4 Transcription

Transcription is a necessary for gene expression of endogenous genes and of genes encoded in plasmids. Transcription is the mechanism by which messenger RNA is transcribed from a DNA template, whereas translation is the synthesis of protein based upon the resulting messenger RNA (mRNA) transcript. Although transcription is similar in prokaryotes and eukaryotes it requires different RNA polymerases and general transcription factors. Transcription is initiated with the help of general transcription factors that help RNA polymerase recognize specific promoter sequences present on the DNA. Double-stranded DNA is then unwound so that a mRNA transcript can be transcribed based upon the template/coding DNA strand. Termination of transcription dissociates the newly synthesized RNA transcript from the polymerase and its DNA template. Although prokaryotic and eukaryotic cells share the same basic principles of transcription, bacteria and eukaryotes employ different RNA polymerases and associated transcription factors.

As such different promoters and terminators are required in mammalian versus bacterial cells, with affinities tailored for their respective transcriptional machinery.

1.1.5. Bacterial pUdO Constructs

We set out to design and synthesize a new set of mammalian plasmids tailored for modern molecular biology techniques. These new plasmids were designed to be as small as possible in order to maximize transfection efficiencies. A set of bacterial plasmids were designed by Dr. Francois-Xavier Campbell-Valois at the University of Ottawa, and denoted 'pUdO' for "plasmid of the Université d'Ottawa". The pUdO plasmids are a set of synthetic plasmids with minimal DNA, and thus contain only the elements essential or useful for gene of interest expression, maintenance, selection, and replication of the plasmid. By minimizing the size of the pUdO plasmids, the goal was to decrease the amount of DNA required for transformations and transfection, and at the same time increase the total number of plasmids in a given mass of DNA (i.e. increase plasmid stoichiometry in a given mass of plasmid). Taking advantage of recent advancements in DNA sequencing and synthesis, pUdO plasmids were assembled to include solely functional components, including: universal sequencing primers, promoters, terminators, and a synthetic MCS.

pUdO Multiple Cloning Site

The MCS is synthetic and designed to contain several restriction sites for commonly used restriction endonucleases, and at the same time is amenable to Golden Gate.

pUdO Promoters and Terminators

The bacterial terminators, thrLABC and L3S2P21, are located on either side of the MCS. Downstream transcription in pUdO plasmids is terminated by either of these terminators. Bacterial terminators contain GC-rich sequences that, once transcribed, form a stem-loop structure leading to dissociation of the newly synthesized RNA from DNA (12). The pUdO bacterial vectors do not have promoters present, allowing for users to insert desired promoters with a gene of interest. This allows for different promoters to be inserted based on the bacterial expression system used.

pUdO ORIs

There are two additional components to the pUdO plasmid backbones that are variable across pUdO constructs: the origin of replication and selection marker. Different origins of replication create a controllable variation in the copy number. For example, p15A has a copy number of about 15 (15), whereas pSC101 has a copy number of about 5 (16) (Table 1). This means that on average each bacterium transformed with p15A or pSC101 has 15 or 5 copies of the respective plasmids.

Table 1. Different Origins of Replication in pUdO Vectors

Vector Name	Origin of Replication
pUdO1	ColE1*
pUdO2	p15A
pUdO3	pSC101
pUdO4	pBBR1

pUdO Selection Markers

Different pUdO plasmids contain different selection markers. Selection markers include antibiotic resistance genes incorporated in plasmids to allow for selection of cells containing the plasmid in the presence of a specified antibiotic. In the case of pUdO plasmids the selection markers are antibiotic resistance genes for either ampicillin, chloramphenicol, trimethoprim, kanamycin, tetracycline, spectinomycin, and zeocin (Table 2). The pUdO plasmids have the same multiple cloning site making it easy to move genes between different pUdOs, and as the pUdOs have a been assembled with various ORIs and selection markers, the pUdO plasmids can be "mixed and matched" for a variety of applications.

Table 2. Different Selection Markers in bacterial pUdO Vectors. 'X' denotes a number representative of the origin of replication present (Table 1).

Vector Name	Selection Marker
pUdOXa	Ampicillin
pUdOXc	Chloramphenicol
pUdOXk	Kanamycin
pUdOXt	Trimethoprim
pUdOXte	Tetracycline
pUdOXs	Spectinomycin
pUdOXz	Zeocin

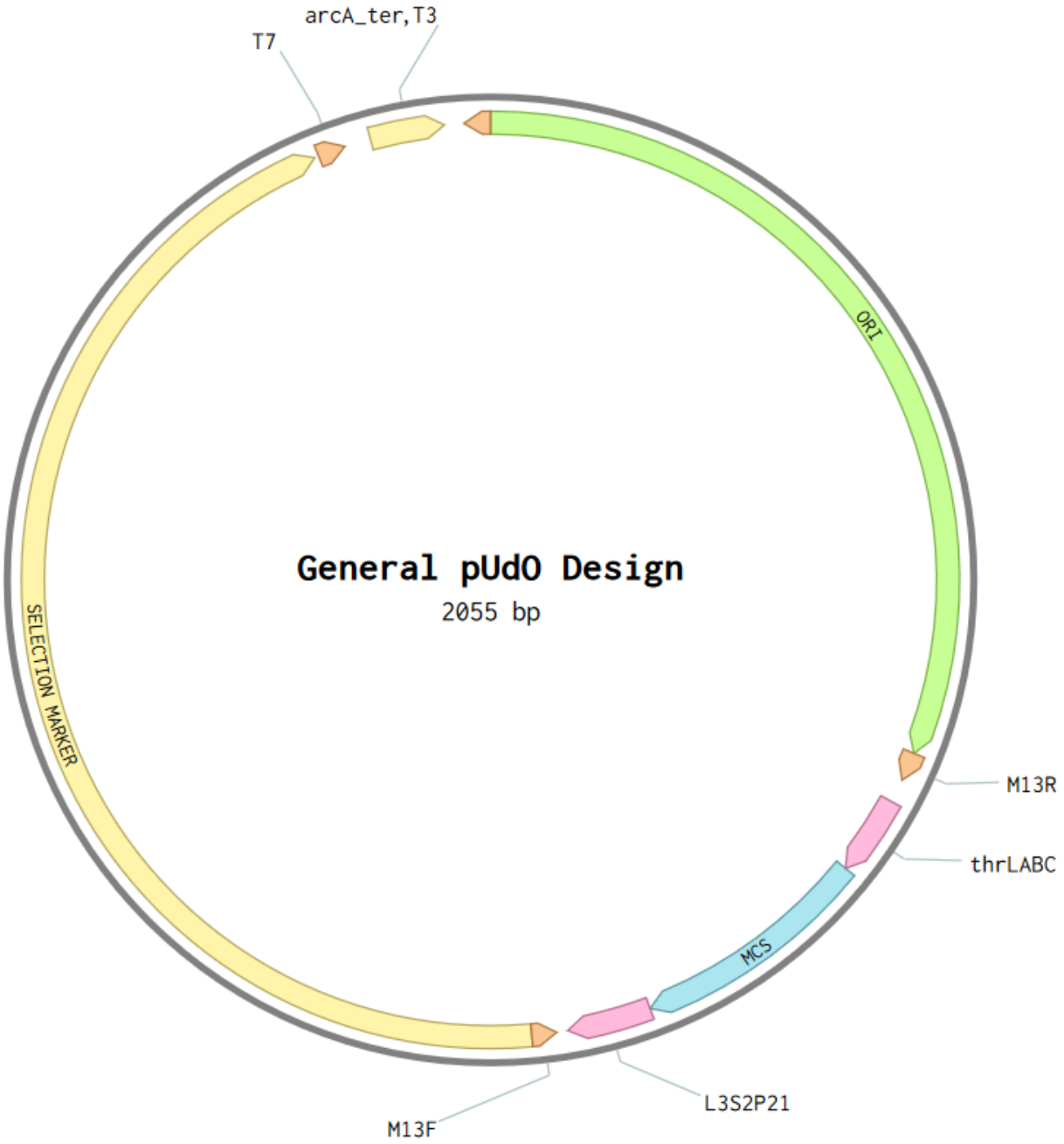


Figure 1. General Schematic of Bacterial pUdO Plasmids. All pUdO plasmids contain a selection marker (yellow), ORI (green), bi-directional arcA terminator, universal sequencing primers (T3, T7, M13R and M13), and the terminators (pink) L3S2P21 and thrLABC on either side of the MCS (blue).

1.1.6. Engineering Mammalian Expression Vectors

The pUdO vectors lack (1) a mammalian promoter and (2) a mammalian terminator necessary to express genes in mammalian cells, and thus can only be used to express genes in bacterial cells. Here, new pUdO-based expression vectors were created to expand the capabilities of the original pUdO vectors, making a limited subset amenable to expression in mammalian cells. As eukaryotic transcription requires eukaryotic promoters and terminators, a eukaryote-compatible promoter and terminator had to be added to make gene expression in mammalian cells possible (Figure 2). pUdO mammalian expression constructs were derived from the original pUdO1 variants, conferring ampicillin, chloramphenicol, trimethoprim, and zeocin resistance (Table 2). The pUdO1 is a high copy vector making it ideal for cloning and maximising the amount of DNA extracted from bacteria prior to transfection into mammalian cells.

pUdO Mammalian Promoter and Terminators

The cytomegalovirus (CMV) promoter was chosen for transcription initiation in mammalian cells. Viruses, like CMV, tend to have small, strong promoters that exploit the host transcriptional machinery (17). Two different mammalian terminators, the Bovine Growth Hormone (BGH) terminator and the Simian Vacuolating Virus 40 (SV40), were chosen to create two different mammalian vector series. The BGH and SV40 terminators are commonly used, strong terminators, that terminate transcription and trigger polyadenylation at the 3'-end of the mRNA. Both the SV40 and BGH terminators contain the AATAAA sequence motif as well as a GC rich region important in recruiting cleavage and polyadenylation factors needed for termination (18,19)).

Derived from the bacterial pUdO1a plasmid, two mammalian expression plasmids were constructed: 1) pUdO1a-C1S expression vector, containing both the full-length CMV promoter and enhancer and the SV40 terminator and 2) the pUdO1a-C1B vector containing the full-length CMV promoter and enhancer and the BGH terminator. These pUdO1a-C1S and pUdO1a-C1B are 2762- bp and 2852-bp synthetic plasmids that can be used to exogenously express proteins in mammalian tissue culture (Figure 2).

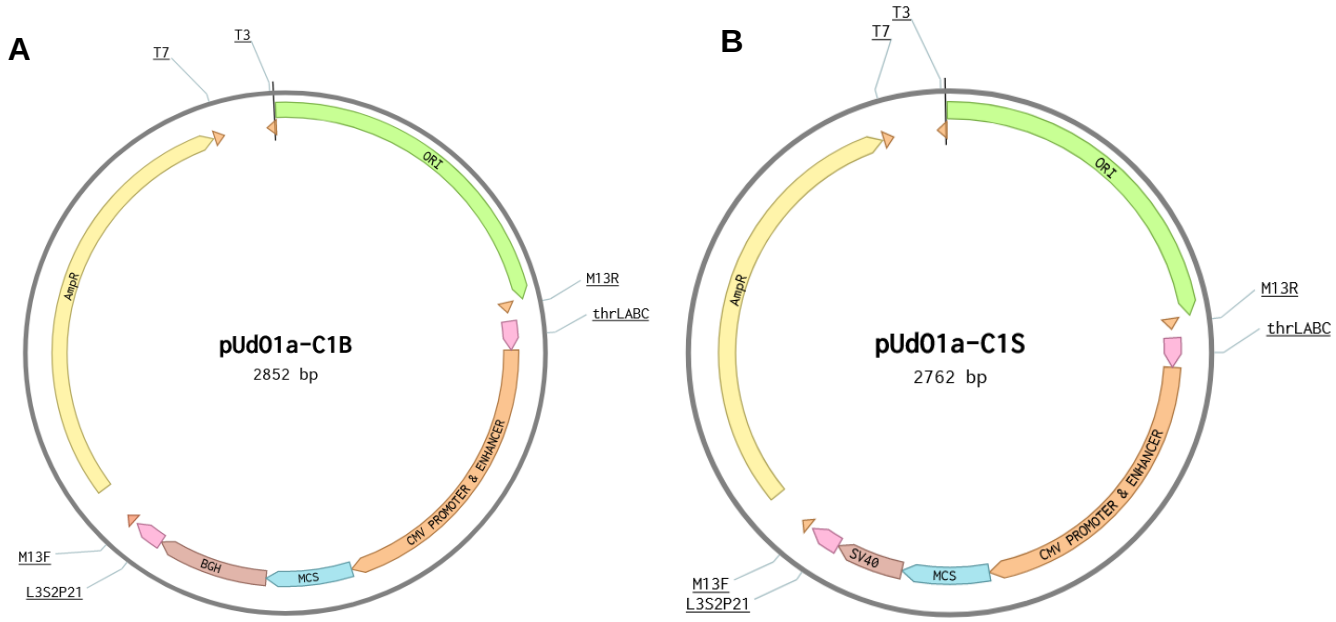


Figure 2. Schematics of the pUdO1a-C1S and pUdO1a-C1B vectors. A. pUdO1a-C1B contains the full-length CMV enhancer and promoter (orange) inserted prior to the MCS (blue) and the BGH terminator (purple) inserted following the MCS. B. pUdO1a-C1S also contains the full-length CMV promoter and enhancer prior to the MCS, but the SV40 terminator (purple) is inserted prior to the MCS.

1.1.7. Successive Deletions to CMV Promoter and Enhancer to Engineer Vectors with Tunable Expression Levels

In bacteria, levels of protein expression can be tuned with the use of pUdOs with different ORIs. However, as the ORI is only relevant in bacteria, tunable levels of protein expression in mammalian cells cannot be achieved this way. Although possible to transfect with lower amounts of DNA this comes with the sacrifice of transfection efficiency. In addition to the pUdO1a-C1S and pUdO1a-C1B vectors eight other versions of the pUdO1a mammalian expression vectors were engineered each with a different truncation of the CMV promoter as denoted by the number following 'C' (Table 3). The full-length CMV promoter and enhancer is 584-bp long, though a shortened, 508-bp version of the CMV promoter and enhancer was identified through an NCBI nucleotide BLAST (nBLAST) against the full CMV. Additionally, three progressively shorter versions of the CMV promoter (100 bp, 52 bp, and 45 bp) have been implemented and characterized in HaloTag® Flexi mammalian expression vectors (Promega) demonstrating decreasing levels of protein expression (20-23). They can be used to express genes in mammalian cells with a range of expression levels depending on which promoter is chosen (20). The full-length CMV promoter, the three versions of CMV promoter identified in HaloTag® Flexi vectors and the shortened version identified on NCBI (Table 3) were used to make five mammalian expression vectors with step-wise expression levels. The three progressive deletions of the CMV promoter and enhancer regions designed by Promega in HaloTag® 7 Flexi Vectors (20) showed decreasing gene expression. Variable gene expression is useful for different research applications as over-expression of proteins can lead to misfolding (24) and cell death (25). Through introducing similar CMV deletions in the

pUdO1a-C1S and pUdO1a-C1B vectors, we aimed to create a set of mammalian expression vectors with tunable gene expression, allowing for the desired expression level to be achieved without sacrificing transfection efficiency.

Table 3. pUdO Mammalian Expression Vectors. Selection Markers, Length of CMV Promoter and Terminators corresponding to each pUdO mammalian expression vector assembled.

Vector Name	**Length of CMV Promoter/Enhancer**	**Terminator**	**Selection Marker**
pUdO1a-C1B	584 bp	BGH	Ampicillin
pUdO1a-C2B	508 bp	BGH	Ampicillin
pUdO1a-C3B	100 bp	BGH	Ampicillin
pUdO1a-C4B	52 bp	BGH	Ampicillin
pUdO1a-C5B	45 bp	BGH	Ampicillin
pUdO1c-C1B	584 bp	BGH	Chloramphenicol
pUdO1t-C1B	584 bp	BGH	Trimethoprim
pUdO1z-C1B	584 bp	BGH	Zeocin
pUdO1a-C1S	584 bp	SV40	Ampicillin
pUdO1a-C2S	508 bp	SV40	Ampicillin
pUdO1a-C3S	100 bp	SV40	Ampicillin
pUdO1a-C4S	52 bp	SV40	Ampicillin
pUdO1a-C5S	45 bp	SV40	Ampicillin
pUdO1c-C1S	584 bp	SV40	Chloramphenicol
pUdO1t-C1S	584 bp	SV40	Trimethoprim
pUdO1z-C1S	584 bp	SV40	Zeocin

*The pUdO mammalian expression vectors vary solely in CMV length and selection markers. The C1B and C1S vectors contain the full-length CMV promoter. C2B and C2S contain a version of the CMV identified on NCBI. The C3B, C3S, C4B, C4S, C5B and C5S vectors contain minimal CMV promoters identified in HaloTag® Flexi vectors. Vector sequences available in Appendix I. II.

1.2.0. Methods

1.2.1. General

pUdO1a, pUdO1c, pUdO1t, and pUdO1z plasmids were provided by the lab of Dr. Campbell-Valois (University of Ottawa, ON, Canada). All restriction enzymes were purchased from New England Biolabs (NEB) Ltd. (Whitby, ON, Canada). Oligonucleotide primers were purchase from ThermoFisher Scientific (Mississauga, ON, Canada). Omega Bio-tek E.Z.N.A.® miniprep Kits and gel extraction kits used to isolate DNA were purchased from VWR (Mississauga, ON, Canada).

1.2.2. Bacterial Cell Culture

Transformations of DH5α or DH10β *Escherichia coli* (*E. coli*) were performed as follows: competent cells were incubated with approximately 1-50 ng of isolated plasmid DNA on ice for 30 minutes, incubated at 42°C for 45 seconds, then placed back on ice for two to five minutes before addition of 225 µL lysogeny broth (LB) and incubated at 37°C in a shaking incubator for one hour. Following this incubation, 50-100-µL of transformation reaction was applied to an LB agar plate, supplemented with the appropriate antibiotic, and the plate was incubated either overnight at 35 °C or at room temperature for two to three days. Bacterial colonies grown on LB agar plates were used to inoculate 4 ml of antibiotic-containing LB culture and incubated overnight before plasmid DNA was isolated using an Omega Bio-Tek E.Z.N.A. Plasmid Miniprep kit purchase from VWR (Mississauga, ON, Canada). All plasmid isolates were then sequenced with Sanger sequencing by Genome Quebec (Montreal, QC, Canada).

1.2.3. Mammalian Cell Culture

For cell electrophysiology experiments, BOSC23 cells, derived from HEK293 cells, were maintained in Dulbecco's Modified Eagle's Medium (supplemented with 10% v/v fetal bovine serum, 100 units/mL penicillin and 100 μg/mL streptomycin) from VWR (Mississauga, ON, Canada). Transfections were performed using calcium phosphate precipitation. In brief, 200 ng of each vector was incubated with calcium phosphate for 4 hours before addition to mammalian cells. Functional cell electrophysiology assays were performed approximately 12-24 hrs post transfection. For Luciferase assay experiments, HEK293T cells were transfected with 1600 ng of DNA. Transfections were performed with Promega ViaFect Transfection Reagent (Fitchburg, WI, USA).

1.2.4. Standard Molecular Biology Techniques

All DNA digests were performed using 1x NEB CutSmart buffer and the indicated NEB restriction enzymes (New England Biolabs Ltd, Whitby, ON, Canada). Restriction digests were incubated at 37°C for at least thirty min. All digested DNA was visualized by 1% agarose gel electrophoresis with 0.05 μg/μl ethidium bromide purchased from Sigma Aldrich (Oakville, ON, Canada). Ligation reactions were performed using a 3:1 (insert:vector) stoichiometric ratio and NEB T4 DNA ligase in 1x NEB T4 DNA ligase. Ligation reactions were incubated at room temperature for at least one hour prior to transformation of *E. coli* cells.

1.2.5. Cell Electrophysiology

The coding DNA sequences of human α1-, β1-, δ-, γ-, and ε -nicotinic acetylcholine receptor (nAChR) subunits were individually cloned into the multiple cloning site of the pUdO1a-C1S vector. For patch clamp electrophysiology experiments, BOSC 23 cells were transfected with pUdO1a-C1S encoding the human α1-, β1-, δ-, and ε- (adult) or γ-subunits (fetal). Vectors for adult and fetal nAChR experiments were co-transfected with a pUdO1a-C1S vector encoding a green fluorescent protein (sequence available in appendix) to identify cells that had incorporated transfected vector DNA. As a positive control, BOSC 23 cells were also transfected with either the adult or fetal subunit combinations in pRBG4 vectors. Transfections with pRBG4 and pUdO1a-C1S vectors varied only in the transfected vector backbones, not the nAChR genes encoded in their respective multiple cloning sites. All patch clamp data was recorded on an Axopatch 200B amplifier (Molecular Devices) at –120 mV and 21˚C using 30 μM acetylcholine. The bath solution contained (in mM): 142 KCl, 5.4 NaCl, 0.2 $CaCl_2$, 1.7 $MgCl_2$, and 10 HEPES, while electrode solution contained (again in mM): 80 KF, 20 KCl, 40 K-aspartate, 2 $MgCl_2$, 1 EGTA, and 10 HEPES. Both bath and electrode solution were adjusted to pH 7.4 with KOH (28).

1.2.6. Luciferase Assays

HEK293T cells were transfected with 1600 ng of DNA. Cells were lysed with Promega Passive Lysis 5X Buffer (Fitchburg, WI, USA). Luciferase assays were performed using Promega Nano-Glo® Reporter Assay System (Fitchburg, WI, USA). 80 uL of Promega Nano-Glo® Reporter Assay Buffer of 40 uL of lysate. Bioluminescence

was measured on a Molecular Devices SpectraMax L Luminometer (San Jose, CA, USA).

1.2.7. Statistical Analysis

All statistical was analyses were performed in GraphPad PRISM 9.3.1 for windows. Statistical significance was determined with Mixed-effects model (REML), assuming unequal standard deviations, corrected with Geisser-Greenhouse correction. The confidence interval was set at 95% and statistical significance is defined as $p<0.05$.

1.3.0. Results

1.3.1. Overview

All pUdO mammalian expression vectors were sequence verified by Sanger sequencing and will be deposited on Addgene to be made universally accessible (Table 4).

Table 4. Mammalian pUdO Expression Vectors to be deposited on Addgene.

Vector Name	Length of CMV Promoter/Enhancer	Terminator	Selection Marker	Vector Size
pUdO1a-C1B	584 bp	BGH	Ampicillin	2852
pUdO1a-C2B	508 bp	BGH	Ampicillin	2776
pUdO1a-C3B	100 bp	BGH	Ampicillin	2368
pUdO1a-C4B	52 bp	BGH	Ampicillin	2320
pUdO1a-C5B	45 bp	BGH	Ampicillin	2313
pUdO1c-C1B	584 bp	BGH	Chloramphenicol	2676
pUdO1t-C1B	584 bp	BGH	Trimethoprim	2226
pUdO1z-C1B	584 bp	BGH	Zeocin	2364
pUdO1a-C1S	584 bp	SV40	Ampicillin	2762
pUdO1a-C2S	508 bp	SV40	Ampicillin	2685
pUdO1a-C3S	100 bp	SV40	Ampicillin	2278
pUdO1a-C4S	52 bp	SV40	Ampicillin	2230
pUdO1a-C5S	45 bp	SV40	Ampicillin	2223
pUdO1c-C1S	584 bp	SV40	Chloramphenicol	2586
pUdO1t-C1S	584 bp	SV40	Trimethoprim	2136
pUdO1z-C1S	584 bp	SV40	Zeocin	2274

1.3.2. Characterization of pUdO Vector's Expression Levels in Mammalian Cells

The relative expression levels of the set of mammalian vectors were characterized with luciferase assays (Figure 3). The set of mammalian expression vectors, with the luciferase gene inserted into the MCS, were transfected in HEK293T cells. As shown by

the decreasing luciferase signal, successive deletions to the CMV promoter and enhancer in pUdO1a-C1S and pUdO1a-C1B resulted in decreasing levels of luciferase activity, and thus presumably luciferase expression. Four successive deletions were engineered by deleting the 5' end of the CMV promoter the resulting promoter lengths ranged from 584 to 45 bp (Table 4.). In the BGH terminator containing vectors, a 30-fold difference in expression is shown between pUdO1a-C1B and pUdO1a-C5B. In the SV40 containing vectors a 36-fold difference in expression is shown between pUdO1a-C1S and pUdO1a-C5S. Most differences in expression levels in the CMV deletion series (BGH or SV40) are significant ($p>0.05$). However, in both the SV40 and BGH terminator series, pUdO plasmids containing C4 and C5 are not significantly different ($p>0.05$). Interestingly, the difference in expression between pUdO1a-C1S and pUdO1a-C2S is significant ($p<0.05$), while differences between pUdO1a-C1B and pUdO1a-C2B are not ($p>0.05$).

The pUdO1a-C1S and pUdO1a-C1B were compared to both pcDNA 3.1 and pRBG4, both established vectors with the same promoters and terminators. pUdO1a-C1S and pRBG4 vectors both have the SV40 terminator and full-length CMV promoter; the pUdO1a-C1B and pcDNA3.1 contain the same full-length CMV promoter and BGH terminator. The expression of the luciferase reporter gene in the pUdO1a-C1B displays similar levels of expression to pcDNA3.1 respectively ($p<0.05$) (Figure 3). The difference in expression levels for pRBG4 and pUdO1a-C1S are significant ($p>0.05$), with the expression under pUdO1a-C1S that demonstrated a lower level of expression. In addition, the mammalian pUdOs with different selection markers were assessed (Figure 5). The selection marker does not alter the level of luciferase expression for both the C1S and C1B series plasmids ($p>0.05$) when the same mass of DNA is transfected.

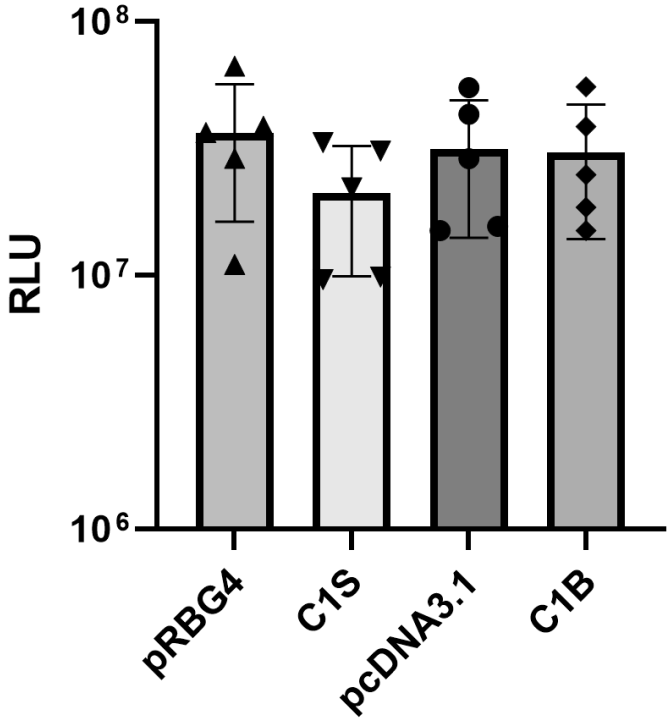


Figure 3. Comparison of mammalian pUdOs with commonly used Vectors for mammalian expression. The pRBG4 vector contains the same CMV promoter and enhancer (full-length) and SV40 terminator as pUdO1a-C1S (denoted C1S here). The pcDNA3.1 vector contains the same CMV promoter and enhancer (584-bp) and BGH terminator as pUdO1a-C1B. All vectors were transfected into HEK293T cells using the same mass of DNA. Mean of 5 biological replicates with technical replicates done in triplicate for each biological replicate.

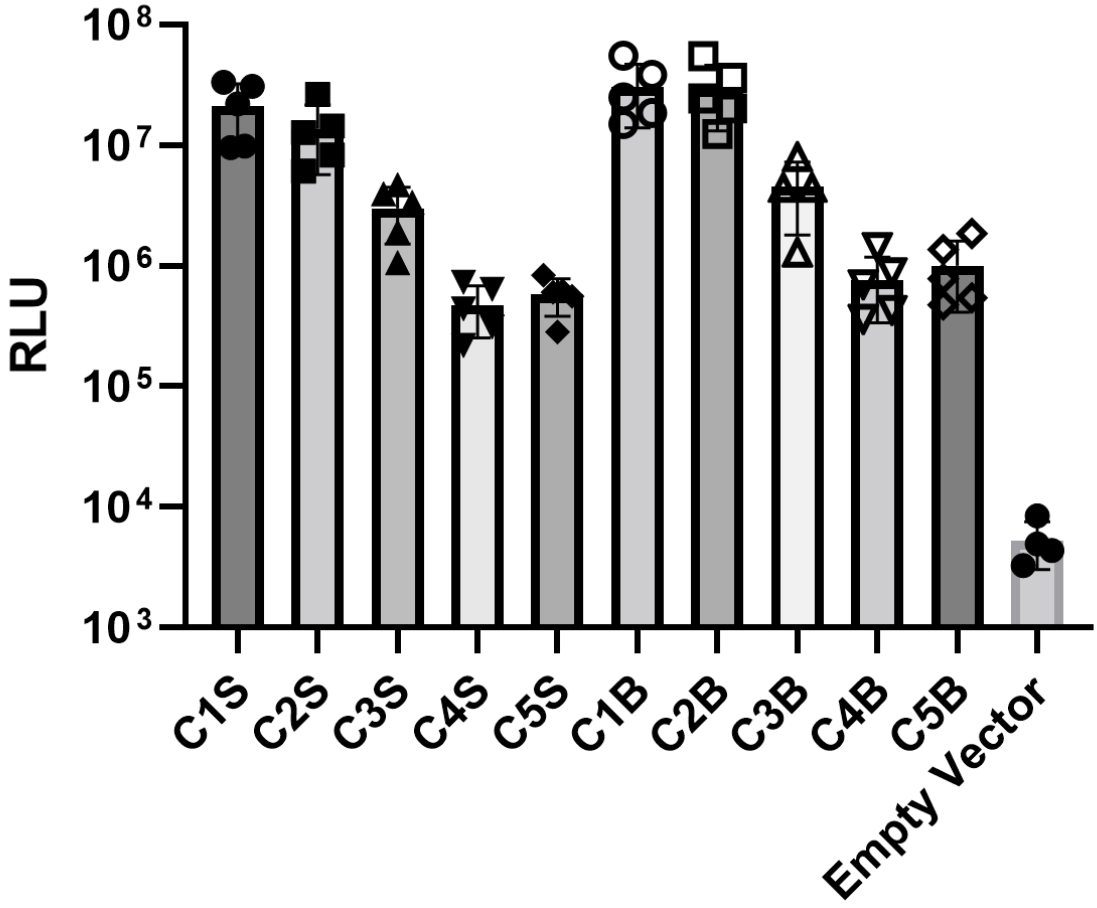


Figure 4. The Effect of CMV promoter and enhancer length on gene expression. All plasmids contain the luciferase reporter gene in the same position in the MCS. All vectors were transfected into HEK293T cells. Bar graphs represent the mean of 5 biological replicates with each data point shown. Each data point represents a mean of three technical replicates. In the empty vector data set one of the individual measurements appeared to be an outlier with a value of 3.38×10^6. This single data point is not shown and was removed prior to statistical analysis. Inclusion of this data point led to an overestimation of the mean bioluminescence (RLU) for this negative control. Error bars represent plus or minus one standard deviation from the mean. Tissue culture work and luciferase assays were performed by Louise Connell.

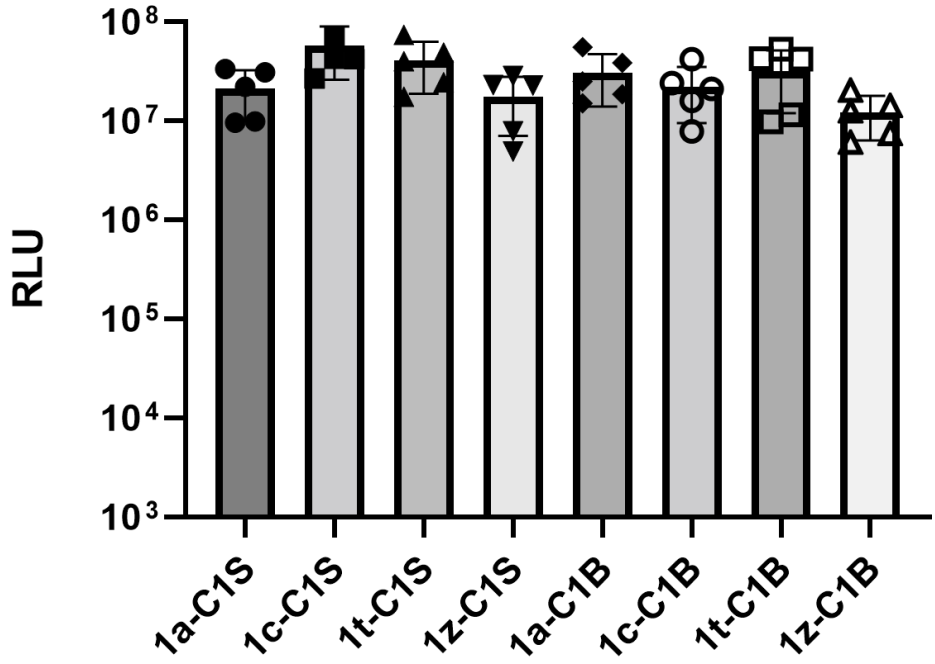


Figure 5. Mammalian pUdO Vectors with Different Selection Markers. All vectors contain the luciferase reporter gene in the same position in the MCS. All vectors were transfected into HEK293T cells. Bar graphs represent the mean of 5 biological replicates with each data point shown. Error bars represent one standard deviation from the mean. Tissue culture work and luciferase assays were performed by Louise Connell.

1.3.3. Cell Electrophysiology

Genes encoding human α1-, β1-, δ-, γ- and ε -nicotinic acetylcholine receptor (nAChR) subunits were inserted into the multiple cloning sites of pUdO1a-C1S vectors. For nAChR patch clamp electrophysiology experiments, BOSC 23 cells were transfected with pUdO1a-C1S with the coding sequence of the human α1-, β1-, δ-, and either ε-(adult) or γ-(fetal) subunits. Cells were also co-transfected with a pUdO1a-C1S vector encoding a green fluorescent protein to facilitate identification of cells that had incorporated vector DNA. As a positive control, BOSC 23 cells were also transfected with either the adult or fetal subunit combinations in their original pRBG4 vectors. Transfections with pRBG4 and pUdO1a-C1S vectors varied only in the transfected vector backbones, not the nAChR genes encoded in their respective multiple cloning sites. The fetal nAChR channels expressed by both the pRBG4 vectors or pUdO1a-C1S vectors both display a amplitudes of ~ 10 pA, typical for fetal nAChR channels. Similarly, both the adult nAChR channels expressed by pRBG4 or pUdO1a-C1S vectors displayed amplitudes of ~13 pA, characteristic of adult nAChR channels (Figure 6) (26).

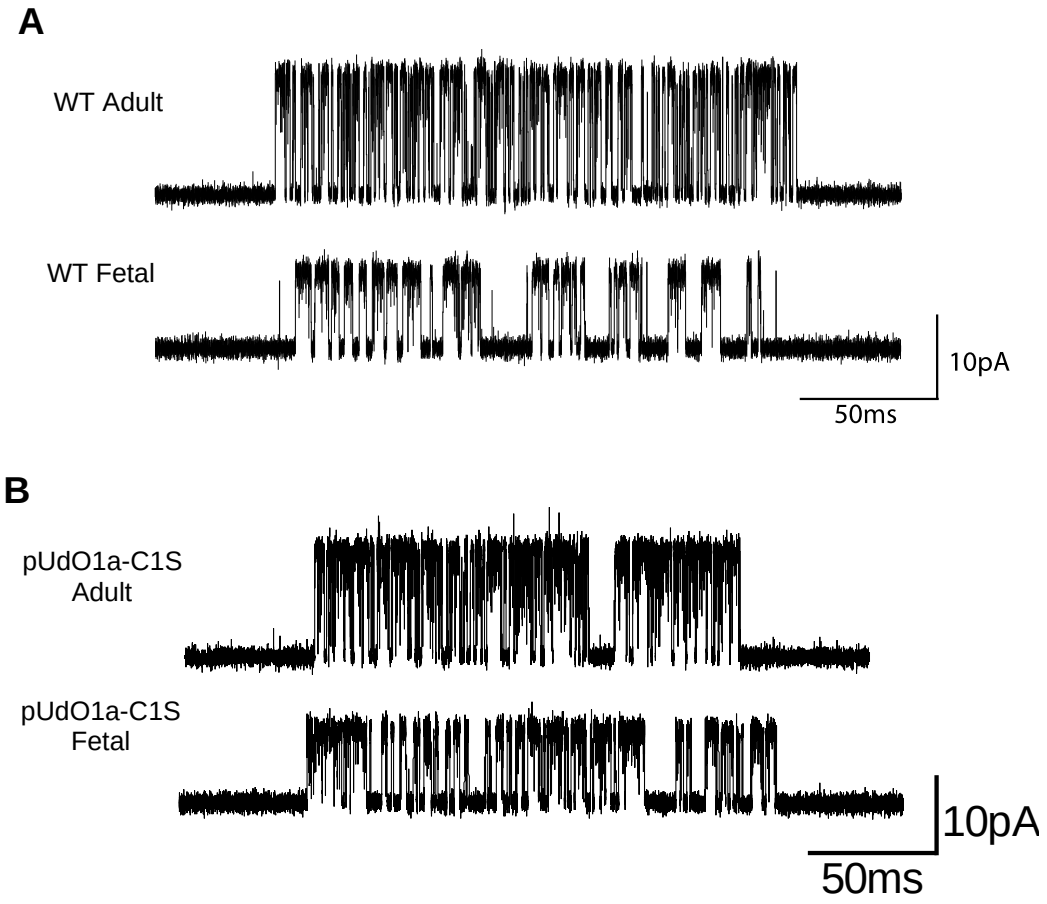


Figure 6. Single-Channel Patch Clamp Recording of adult and fetal nAChR expressed through transfection of pUdO1a-C1S or pRBG4 vectors. Patch-clamp recordings of wild-type adult and fetal nAChRs expressed from (A) control pRBG4 vectors and (B) pUdO1a-C1S vectors, both in BOSC23 mammalian cells. Cells were transfected with vector DNA and assayed by patch clamp electrophysiology in the presence of 30 μM acetylcholine 12-24 hrs post-transfection. All recordings were obtained with an applied voltage of -120 mV, where openings are upward deflections and represent inward cationic currents. Tissue culture and electrophysiology performed by Christian Tessier and Johnathon Emlaw.

1.4.0. DISCUSSION

1.4.1. Overview of Assembled pUdO Mammalian Expression Vectors

A set of pUdO vectors for protein expression in mammalian cells was engineered from the pUdO vector. The original pUdO plasmids, an engineered set of next-generation synthetic plasmids, are minimal with respect to size, maximised in terms of functionality, and designed specifically for use in modern molecular biology techniques. The new pUdO mammalian expression vectors, built on a subset of pUdO bacterial expression plasmids, expand the utility of the family of pUdO plasmids to gene expression in mammalian cells. The pUdO mammalian expression vectors were constructed by incorporating the well-established CMV promoter and SV40 terminator onto three pUdO vectors containing different selection markers, pUdO1a, pUdO1c, pUdO1t, and pUdO1z (Table 4). The CMV promoter, and SV40 and BGH terminators, are commonly used and were chosen for their strength and small size (17,18,19). To expand further the capabilities of the pUdO, we engineered an additional series based on the new pUdO1a. These contained successive deletions of the CMV promoter region, with the goal of creating a set of vectors with stepwise decreasing expression levels. Finally, although the pUdO mammalian expression vectors range in size, from 2136 bp to 2852 bp, the largest pUdO is substantially smaller than commonly mammalian expression vectors, being almost half the size of pcDNA3 (5446 bp). Smaller vectors should allow for better transformation and transfection efficiencies (1, 27).

1.4.2. Benefits of pUdO Mammalian Expression Vectors

Plasmid size is a key determinant in transformation efficiency (1). An early study of transformations in *E. coli* demonstrated that transformation efficiencies was inversely proportional to plasmid size (1). Without an inserted gene of interest, the pUdO1a, pUdO1c, pUdO1t, and pUdO1z vectors range in size from 1417 bp to 2043 bp. The minimalist design of pUdO plasmids makes them substantially smaller than conventional bacterial expression vectors. Extending the bacterial pUdO design to mammalian expression systems creates a set of novel vectors that although unproven, are expected to have increased transformation and transfection efficiencies. As observed for transformations, transfection efficiencies diminish with increasing vector size (27). The pUdO mammalian expression vectors vary in size from 2136 bp to 2852 bp (Table 4), considerably smaller than conventional mammalian expression vectors such as pcDNA (5446 bp) (28) and pCI (4006 bp) (29). Another characteristic of smaller vectors is their stoichiometric vector to mass ratio, a set mass of DNA will contain a higher number of smaller vectors compared to the same mass of larger vectors. This conveniently allows for stoichiometrically more pUdO mammalian vectors to be transfected into cells all the while transfecting the same mass of DNA as with larger vectors.

1.4.3. Relative Expression Levels of Mammalian pUdO Plasmids

Expression levels of Mammalian pUdO C1B and C1S Plasmids

The pUdO1a-C1B vector contains the same CMV promoter and enhancer, and BGH terminator, as pcDNA3, and also displays similar levels of gene expression (Figure 4). Similarly, all other pUdO C1B vectors (pUdO1c-C1B, pUdO1t-C1B and pUdO1z-C1B) display similar levels of gene expression in comparison to that of pcDNA3 (Figure

5). The C1S series of pUdO vectors (pUdO1a-C1S, pUdO1c-C1S, pUdO1t-C1S, and pUdO1z-C1S) contain the same full-length CMV promoter and enhancer and SV40 terminator as pRBG4. pRBG4, pUdO1a-C1S, pUdO1c-C1S, pUdO1t-C1S and pUdO1z-C1S all display similar levels of luciferase expression (Figure 4 and 5).

Tunable Expression Levels by Successive Deletions to the CMV Promoter/Enhancer

The pUdO1a-C1S and pUdO1a-C1B vectors contain the full-length CMV commonly used to express genes in mammalian cells, but a set of four additional CMV promoter and enhancer deletion vectors were engineered with increasingly large deletions in their CMV region. These vectors with truncated CMV promoter/enhancer regions were designed to allow for step-wise control of the level of gene expression (Table 4). The five different CMV promoter lengths achieve over a 30-fold difference in expression levels in the series of BGH terminator pUdO1a vectors and a 36-fold in the SV40 terminator series of pUdO1a vectors (Figure 3 and Figure 4). The smallest CMV promoter, 45 bp long, is sufficient for protein expression in HEK cells, despite the removal of 539 bp of the CMV promoter and enhancer. Incorporation of such a small promoter generates pUdO1a mammalian vectors only 2136 bp in size, less than half the size of commonly used cloning vectors, like pcDNA.

1.4.4. pUdO1a-C1S Expresses Genes in Mammalian Cells

Electrophysiology experiments were performed to confirm genes inserted into the pUdO1a-C1S MCS express functional proteins in mammalian cells. Muscle-type nicotinic acetylcholine receptors (nAChR) are pentameric but composed of 4 different subunits.

Expression of all four individual protein subunits is required to form functional nicotinic acetylcholine receptors. Single channel recordings of adult and fetal muscle-type nAChR shows pUdO1a-C1S vectors encoding nAChR subunits express functional proteins capable of forming the heterologously-expressed nAChR. In this experiment, both adult and fetal nicotinic acetylcholine receptors were analyzed to validate results. Both fetal and adult nicotinic acetylcholine receptors in the pRBG4 vector were used as positive controls (Figure 6). Functional nicotinic acetylcholine receptors were expressed in BOSC23 cells following transfection with nAChR coding DNAs (coding DNA) within either pRBG4 or pUdO1a-C1S vectors, and furthermore the activity of the resulting nAChRs was indistinguishable at the single-channel level. There is a noticeable difference in the amplitude measured of the adult and fetal nicotinic acetylcholine receptors in the wild type experiments. This has been observed in other studies comparing fetal and adult nicotinic acetylcholine receptors and is characteristic of the different subunit compositions, α1-, β1-, δ-, and ε-subunits in adult nAChR or α1-, β1-, δ-, and γ-subunits in fetal nAChR, and as such the fetal nAChR was expected to display a lower amplitude (34). The nAChR assembled with subunits expressed in pUdO1a-C1S compared to subunits expressed in pRBG4 showed no discernable differences, with currents of approximately 13 pA and 10 pA for adult and fetal nAChR respectively. Given the functionality of proteins should not be affected by the vectors that encode them, the amplitudes measured should not be affected, and this was indeed the case.

Since the pUdO1a-C1S vectors carry cDNA encoding nAChR subunits, the presence of functional nAChR provides direct evidence that the pUdO1a-C1S mammalian expression vectors express genes in mammalian cells. The presence of ‘bursts’ in the

conductance seen in the presence of acetylcholine indicates there are functional channels present that can facilitate the permeation of ions across the membrane in response to acetylcholine (Figure 6). The channel activity in the presence of acetylcholine confirms pUdO1a-C1S can express functional proteins in mammalian cells. Comparison of nicotinic acetylcholine receptors whose subunit genes are expressed in pRBG4 to those expressed in pUdO1a-C1S confirmed the vector used has no effect on the functionality of the channels formed. In addition, the pUdO1a-C1S vector is smaller than pRBG4, and although unproven, it is anticipated that less DNA (by mass) is required to transfect the same number of cells, and greater transfection efficiencies are also expected. This could allow for lower masses of DNA to be used for transfection, minimizing DNA toxicity (27).

1.4.5. Summary

The minimized pUdO mammalian expression vectors contain the elements necessary and sufficient for robust expression in mammalian cells. This toolbox of vectors contains vectors with different selection markers and tunable expression levels, and since the backbone structure is identical between vectors in the same series, genes can easily be transferred within the set as needed. This set of pUdO mammalian expression vectors constitute a novel synthetic biology tool that, in principle, should be useful for a broad range of applications.

Chapter 2: Linking the Presence of Calcium to Cell Survival

2.1.0. Introduction

2.1.1. General

The goal of this project is to regulate bacterial cell survival with a eukaryotic calcium-dependent transcription factor, with the aim of exploiting it for future directed evolution strategies aimed at engineering calcium-selective ion channel proteins in bacterial systems.

2.1.2. Benefits of Synthetic Biology Approaches in Bacteria

Synthetic approaches to biology can be applied to both mammalian and bacterial cells, with bacterial systems having some distinct advantages. Bacterial systems offer ease of cloning, protein expression, and clonal selection. Bacterial systems are also easier and cheaper to maintain than mammalian cell lines. In a directed evolution system, clonal selection allows for easy isolation of protein variants that display desired phenotypes; easily maintaining the genotype-phenotype linkage. Bacterial systems are ideal as they allow for clonal selection, isolating variants of interest. Transforming a library of plasmids containing different mutations in a gene, leads to each colony containing a unique protein mutation(s). This allows for individual phenotypes to be screened or selected for by simply inoculating different colonies/mutants.

2.1.3. Transcription Factors

Transcription factors regulate the transcription of a gene, thereby regulating its expression. Most transcription factors recognize specific DNA sequences, called operators,

and through their interactions with operators, control gene expression (30). Discovered in 1961, the *lac operon* system is a well-established example of control of gene expression through repression of transcription. The lac Inhibitor (lacI) binds to the lac operator (a 25 bp DNA sequence) preventing RNA polymerase from binding to the lac operator, which in turn prevents the transcription of subsequent downstream genes. In the *lac operon* system, gene expression is inducible, as the sugar lactose can bind to LacI causing an allosteric change that reduces the affinity of LacI for the lac operator sequence. Dissociation of LacI from the lac operator permits the expression of downstream genes (30). In synthetic biology the LacI and lac operator are widely used to impart inducible gene expression to a target gene. This system has been particularly successful for protein purification strategies aimed at extracting large quantities of toxic proteins that when constitutively expressed can inhibit cell growth. Transcription factors, like lacI, link a signal, like lactose, to the expression of a gene. Thereby the expression of a gene can be toggled "on" or "off" by the presence of a signal.

2.1.4. Downstream Regulatory Element Antagonist Modulator (DREAM)

2.1.4.1. Background

Dynorphins are a class of neuropeptides, encoded by the human prodynorphin gene, which are involved in pain management. The human prodynorphin gene is regulated in part, by a transcription factor called DREAM (31). Prior to the human prodynorphin gene is a DNA sequence denoted Downstream Regulatory Element (DRE). The binding of the Downstream Regulatory Element Antagonist Modulator protein (DREAM), at the DRE represses the expression of the prodynorphin gene in a manner reminiscent of the classic

lac operon system with regards to an inhibitor binding an operator until an external signal binds the inhibitor preventing the inhibitor from binding the operator. De-repression of the prodynorphin gene occurs in the presence of elevated intracellular calcium concentrations (31). Identified in 1999, by Naranjo *et al.*, the DREAM protein is 256 amino acids long, comprised of four EF hand motifs, and two functional calcium binding sites (31, 32, 33). In the absence of calcium, DREAM forms a tetramer capable of binding DRE around a central 'GTCA' sequence (34). Elevated intracellular calcium causes DREAM to dissociate into dimers, which exhibit lower affinity for the DRE. Dissociation of DREAM from the DRE derepresses gene expression, allowing for the expression of subsequent/downstream genes (34). The ability of DREAM to repress gene expression in the absence of calcium has been established (31, 33). In mammalian HEK cell lines, DREAM has been shown to repress gene expression when the DRE sequence is present upstream of a reporter gene in an expression vector (31), unlike in the lacO system where the lac operator prevents RNA polymerases from binding a promoter sequence, it is also possible that DREAM prevents the progression of RNA polymerases along the DNA, thereby preventing transcription. In HEK cells, a plasmid expressing DREAM was co-transfected with the DRE reporter vector, leading to repression of a reporter gene. Co-transfection of a DREAM containing plasmid and a reporter plasmid containing a mutated DRE sequence showed recovery of gene expression (31, 33) consistent with DREAM binding specifically to the DRE to repress gene expression. Finally, electrophoretic mobility shift assays have shown DREAM associates with a double stranded DRE oligonucleotide in the absence of calcium but does not associate at saturating calcium levels, consistent with calcium dependent de-repression (35).

2.1.4.2. The Basis for Bacterial Expression

The ability of DREAM to repress gene expression has only been established in mammalian cells. DREAM-C, a truncated version of DREAM that appears to be more soluble in bacteria, retained key properties of DREAM essential for the regulation of gene expression. Electrophoretic mobility shift assays of DREAM-C extracted from bacterial cells (BL21gold DE3) showed the DREAM-C associates with double-stranded oligonucleotides of human DRE in the absence of calcium (40, 41). In the presence of calcium however, there is no evidence of DREAM-C binding the DRE oligonucleotides (40, 41). Although explicit control of gene expression has not been demonstrated in bacterial cells, the ability of DREAM-C to bind the DRE is calcium dependent suggesting that the DREAM system might provide an opportunity to link intracellular calcium levels to gene expression in bacteria.

2.1.5. Levansucrase

To assess the ability of DREAM to regulate gene expression in bacterial, the DREAM-DRE system was implemented to regulate the expression of the *sacB* gene. The *sacB* gene encodes an enzyme, levansucrase, originally identified in *B. bacillus* (36,37). Levansucrase converts sucrose to levan, a compound proposed to accumulate in the periplasm of bacteria leading to cell death (36) (Appendix II.I). Therefore, if the *sacB* gene is expressed in the presence of sucrose, cells are unable to grow and thrive. As such, cell growth in the presence of sucrose is a proxy for the expression of *sacB.* This allows for the

effects of transcription factors on gene expression to be assessed based on the growth (or lack thereof) in the presence and absence of sucrose.

2.5.6. Linking *sacB* Expression to Regulation by DREAM

In future directed evolution applications aimed at evolving calcium permeable ion channels, linking the presence of calcium to cell survival provides a method for selecting protein variants of interest. For this reason, the ability of DREAM to repress gene expression by binding the DRE sequence was assessed using the *sacB gene*. After the development of a robust binary growth assay, the ability of DREAM to effect transcription was assessed by inserting DREAM and the DRE sequence into the pSacB vector. The DRE sequence is inserted directly following the T7 promoter (and +1 transcription start site) prior to the *sacB* gene. As its name suggests the downstream regulatory element (DRE) must be downstream of the promoter, and upstream of a gene, to prevent gene expression in mammalian cells upon DREAM binding. Repression of gene expression by the DREAM-DRE system in bacterial cells was characterized by cell growth in the presence of sucrose. Repression of gene expression should be dependent on both the presence of the DRE sequence upstream of the *sacB* gene and the expression of DREAM. The characterization of DREAM as a binary switch would provide the basis for future development of an integrated system to evolve calcium-selective ion channels.

2.2.0. Methods

2.2.1. Assembly of Vectors

The pACYCDUET1 vector was a gift from the lab of Dr. Chica at the University of Ottawa. Geneblocks containing either the DREAM gene or the hDRE sequence were purchased from IDT and inserted into the pACYCDUET1 vector via Gibson assembly (*New England Biolabs* NEBuilder® HiFi DNA Assembly Master Mix). All genes were expressed in a plasmid variant of the pAYCYDUET1 plasmid, denoted pACYCDUET1$_X$. The pACYCDUET1$_X$ vector was created by site-directed mutagenesis to remove the lacI gene and promoter from pACYCDUET1. The DREAM gene was previously inserted into the pAYCYDUET1 vector by Abdel Rahman Lepabic via Gibson assembly. Site-directed mutagenesis of pACYCDUET1_DREAM created pACYCDUET1_DREAM-C by removing the first 64 amino acids of the DREAM gene. The DREAM and DREAM-C genes were moved into the pACYCDUET1$_X$ vector by restriction cloning from pACYCDUET1. The *sacB* gene was inserted from pRMMB69 (Addgene #48116) to the second MCS of pAYCYDUET1$_X$, hereafter referred to as pSacB. The DRE sequence was inserted into the empty pACYCvector via Gibson assembly using a gene block from IDT (Appendix III) then further cloned into the pSacB vectors to create pSacB_DRE where the *sacB* gene in under control of the DRE. Restriction cloning was used to add the DREAM or DREAM-C genes in the alternative multiple cloning sites of the pSacB_DRE or pSacB plasmids.

2.2.3. Bacterial Cultures

Transformation of DH5α or DH10β *Escherichia Coli* (*E. Coli*) competent cells was performed with plasmids using the following standard protocol: Cells were incubated with approximately 1-50 ng plasmid DNA on ice for 30 min, incubated at 42 °C for 45 sec, then placed back on ice for 2-5 min before addition of 225 µL lysogeny broth (LB) and incubation at 37 °C in a shaking incubator for 1 hr. Following this incubation, 50-100 µL of transformation reaction was applied to an LB agar plate, supplemented with antibiotic, and the plate was incubated either overnight at 35 °C or at room temperature for two to three days. Bacterial colonies grown on LB agar plates were used to inoculate 4 mL of antibiotic-containing LB culture and incubated overnight before plasmid DNA was isolated using an Omega Bio-Tek E.Z.N.A. Plasmid Miniprep kit purchase from VWR (Mississauga, ON, Canada). All plasmid isolates were then Sanger sequenced by Genome Quebec (Montreal, QC, Canada) or the DNA Sequencing Facility at the Hospital for Sick Children (Toronto, ON, Canada).

2.2.4. Restriction cloning

All DNA digests were performed using 1x NEB (New England Biolabs Ltd) Cutsmart buffer and the indicated NEB restriction enzymes (Whitby, ON, Canada). Restriction digests were incubated at 37 °C for at least 30 min. All digested DNA was visualized by 1%-agarose gel electrophoresis with 0.05 µg/µl ethidium bromide purchased from Sigma Aldrich (Oakville, ON, Canada). Ligation reactions were performed using a

3:1 (insert:vector) ratio and NEB T4 DNA ligase (Whitby, ON, Canada) in 1x NEB T4 DNA ligase buffer (Whitby, ON, Canada). Ligation reactions were incubated at room temperature for at least 1 hr prior to transformation of *E. coli* cells.

2.2.5. Cell Growth Assays

All vectors were transformed in BL21gold (DE3) cells. An aliquot of BL21gold (DE3) cells were provided by the lab of Dr. Campell-Vallois at the University of Ottawa. Gene expression was induced by IPTG. 0.1 mM) Transformed cells were inoculated overnight in LB supplemented with 25 ug/ml chloramphenicol, 0.1 mM IPTG and varying volumes of syringe filtered sucrose (0.5 g/mL) or the equivalent volume of syringe-filtered distilled water. Cultures were grown at 37 °C in an Eppendorf New Brunswick E24 Orbital Shaker (Hamburg, Germany) for 16 hrs, prior to measuring the optical density at 600 nm on an Eppendorf 3P Biospectrometer basic (Hamburg, Germany) (Figure 7).

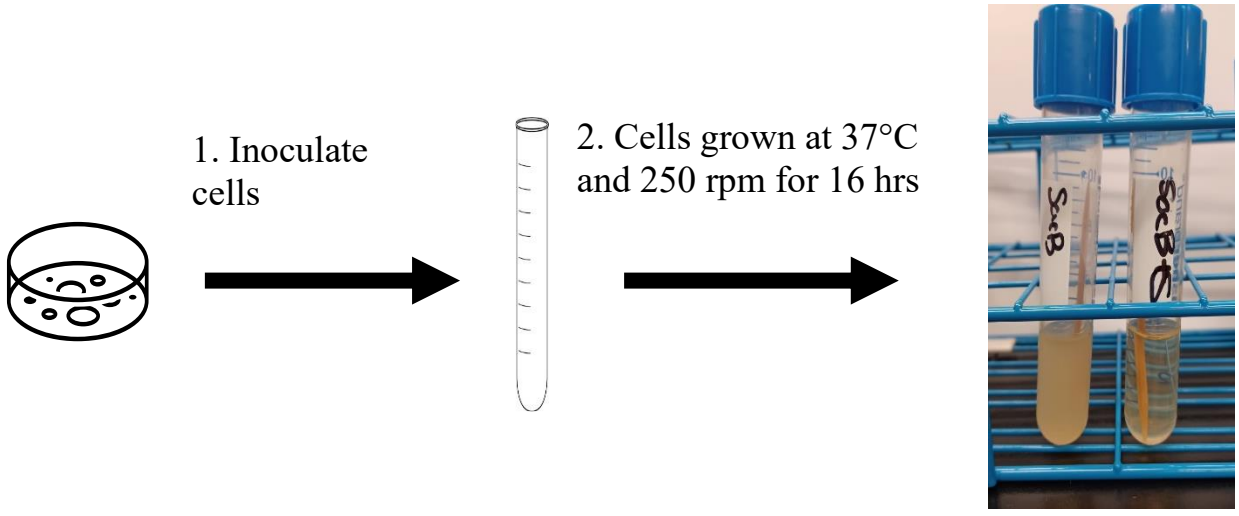


Figure 7. Schematic Overview of Cell Growth Assays. Cells containing pSacB were inoculated into liquid media containing sucrose or and equivalent volume of water. Cells were grown overnight at 37 °C and 250 rpm. Expression of *sacB* is characterized by a lack of cell growth in the presence of sucrose.

2.2.6. Statistical Analysis

All statistical was analyses were performed in GraphPad PRISM 9.3.1 for windows. Statistical significance was determined with Brown-Forsythe and Welch ANOVA tests, assuming unequal standard deviations, corrected for multiple comparisons using Dunnett's T3 multiple comparisons test. The confidence interval was set at 95% and statistical significance is defined as $p<0.05$.

2.3.0. Results

2.3.1. Optimization of Sucrose Concentration

To characterize cell growth as it relates to the expression of *sacB*, a simple growth assay was set up. Cells containing a control vector (empty pACYC DUET1$_X$) or the pSacB plasmid were transformed in BL21gold DE3 cells and plated on LB-agar (supplemented with chloramphenicol). Colonies were inoculated directly into LB with 0.1 mM IPTG, 25 ug/ml chloramphenicol, with either sucrose (0.5 g/mL) or the same volume of water.

To determine the minimum sucrose concentration required for cell death, cell growth assays were performed at various sucrose concentrations to assess cell growth after 16 hours post inoculation. 0.5% sucrose was insufficient to prevent cell growth in cells expressing levansucrase, with cells growth reaching an average OD600 value of 1.541 (Figure 8A). Although 1% sucrose prevent cell growth in a fraction of replications, it was not sufficient to consistently prevent cell growth in the the pSacB condition (Figure 8B). Both 2.5% and 5% sucrose conditions reproducibly prevented cell growth in the pSacB conditions (Figure 8B, 8C). However, in the presence of 5% sucrose, the control also demonstrated a decrease in cell growth (Figure 8D). Given that 2.5 % sucrose demonstrated a smaller difference between the control conditions, while at the same time maintaining a significant difference in the pSacB conditions, 2.5% sucrose was used for all subsequent experiments.

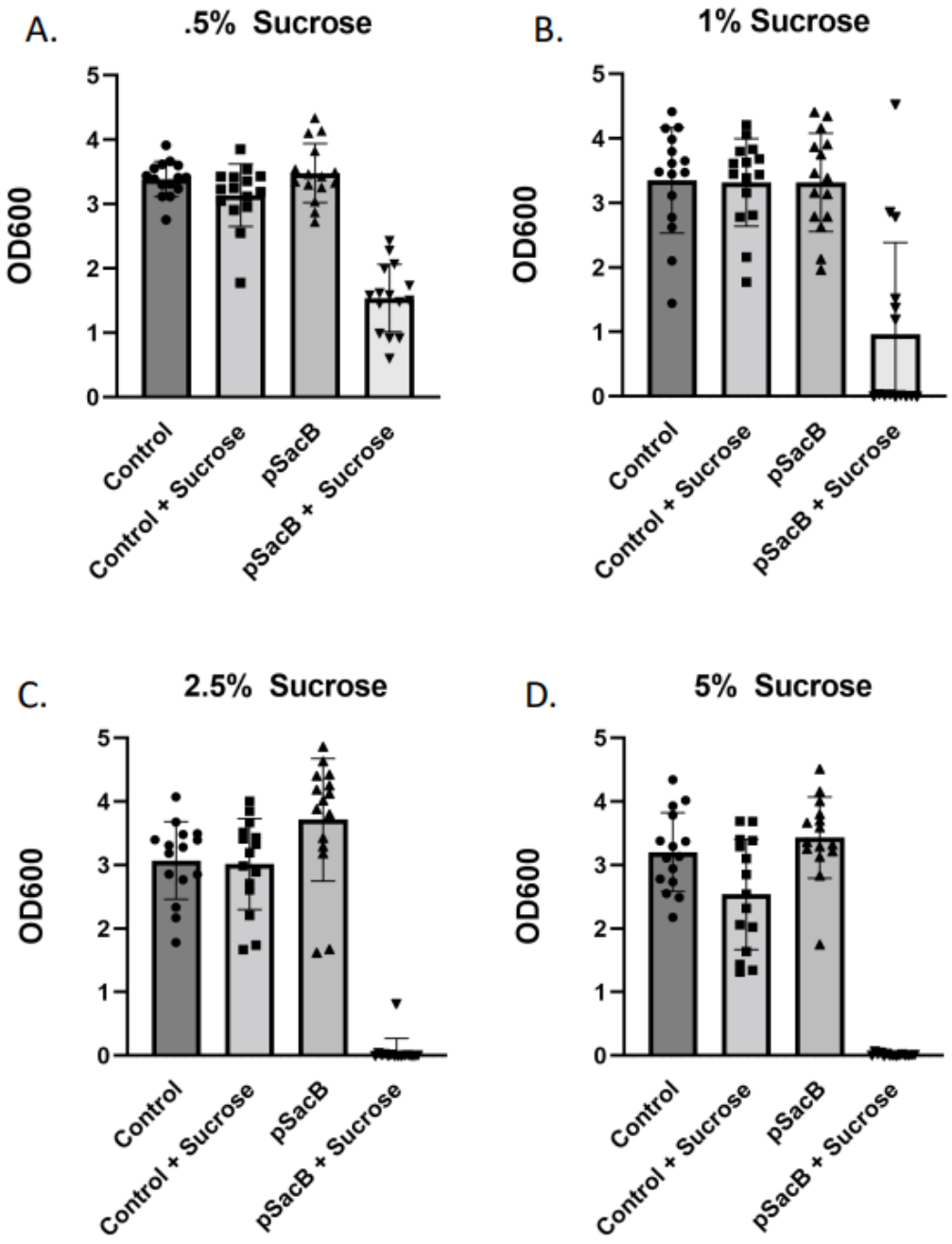


Figure 8. Optimization of Sucrose Concentration in Cell Growth Assays. BL21gold DE3 transformed colonies were inoculated in 8 mL LB (supplemented with 0.1 mM IPTG and 25 ug/mL chloramphenicol) with varying volumes of 0.5 g/ml sterile filtered sucrose to achieve sucrose concentrations of (A) 0.5% (w/v), (B) 1% (w/v), (C) 2.5% (w/v) or(D) 5% (w/v) or the equivalent volume of sterile filtered distilled water. Cultures were grown at 37 °C and 250 rpm for 16 hrs post-inoculation. Optical density (OD) was measured at a wavelength of 600 nm. Bars represent the mean of all replicates. Error bars represent plus or minus one standard deviation from the mean.

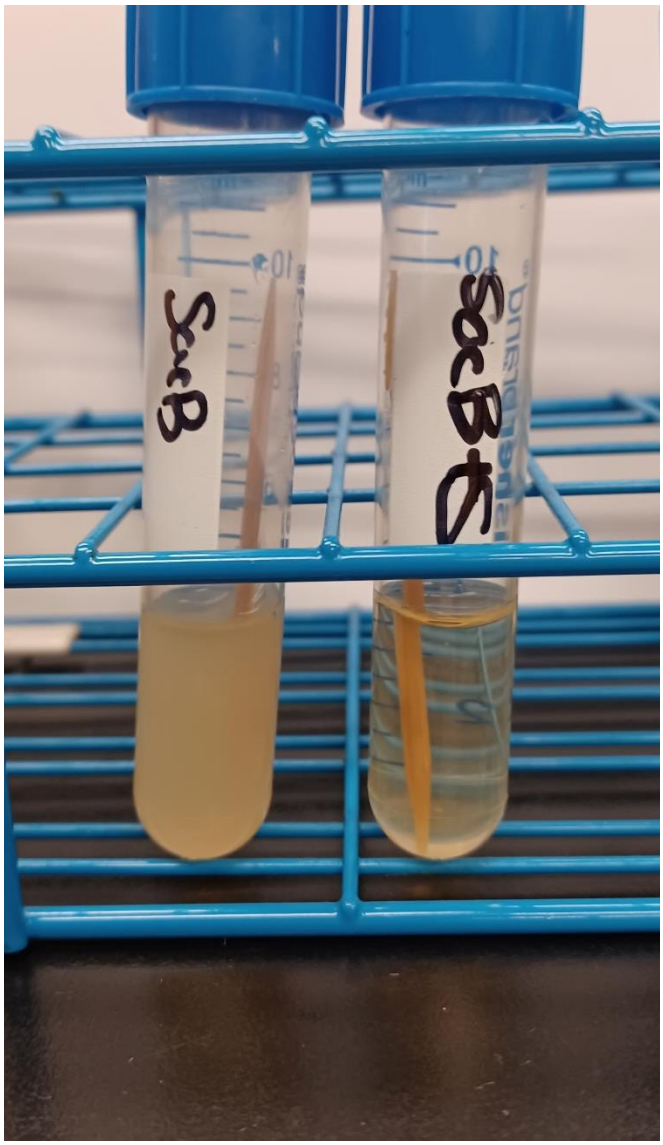


Figure 9. Cell Growth 16 hrs post-inoculation of BL21g DE3 cells containing pSacB in the absence or presence of 2.5% sucrose in a 4 ml Culture Volume. One representative culture from each condition is shown. 4ml of culture prepared with 0.1 mM IPTG, 200 uL of 0.5 g/ml sterile filtered sucrose or the equivalent volume of sterile filtered distilled water. Left: pSacB + water Right: pSacB + Sucrose

2.3.2. The Negative Effects of the DRE Sequence on *sacB* Expression

The 25 bp human prodynorphin DRE sequence was inserted immediately after the T7 promoter sequence (and +1 Transcription Start Site) in place of a pre-existing lacO sequence (also 25 bp) prior to the *sacB* gene (Figure 10A) (Table 5). The growth assay was repeated with this new vector denoted pSacB_DRE. The insertion of the DRE sequence was not expected to impact cell growth in the absence or presence of sucrose.

Cells transformed with pSacB_DRE showed an increased cell growth (mean OD600 of 1.096) in the presence of 2.5% sucrose where as pSacB transformed cells demonstrated little to no cell growth (mean OD600 of 0.064) (Figure 10B). The difference in cell growth of cells containing pSacB or pSacB_DRE in the presence of sucrose is significant ($p<0.05$), and thus swapping the LacO sequence for the DRE sequence decreases the expression of the *sacB* gene.

Table 5. Comparison of LacO and DRE sequence

LacO Sequence	GGAATTGTGAGCGGATAACAATTCC
DRE Sequence	GAAGCCGGAGTCAAGGAGGCCCCTG

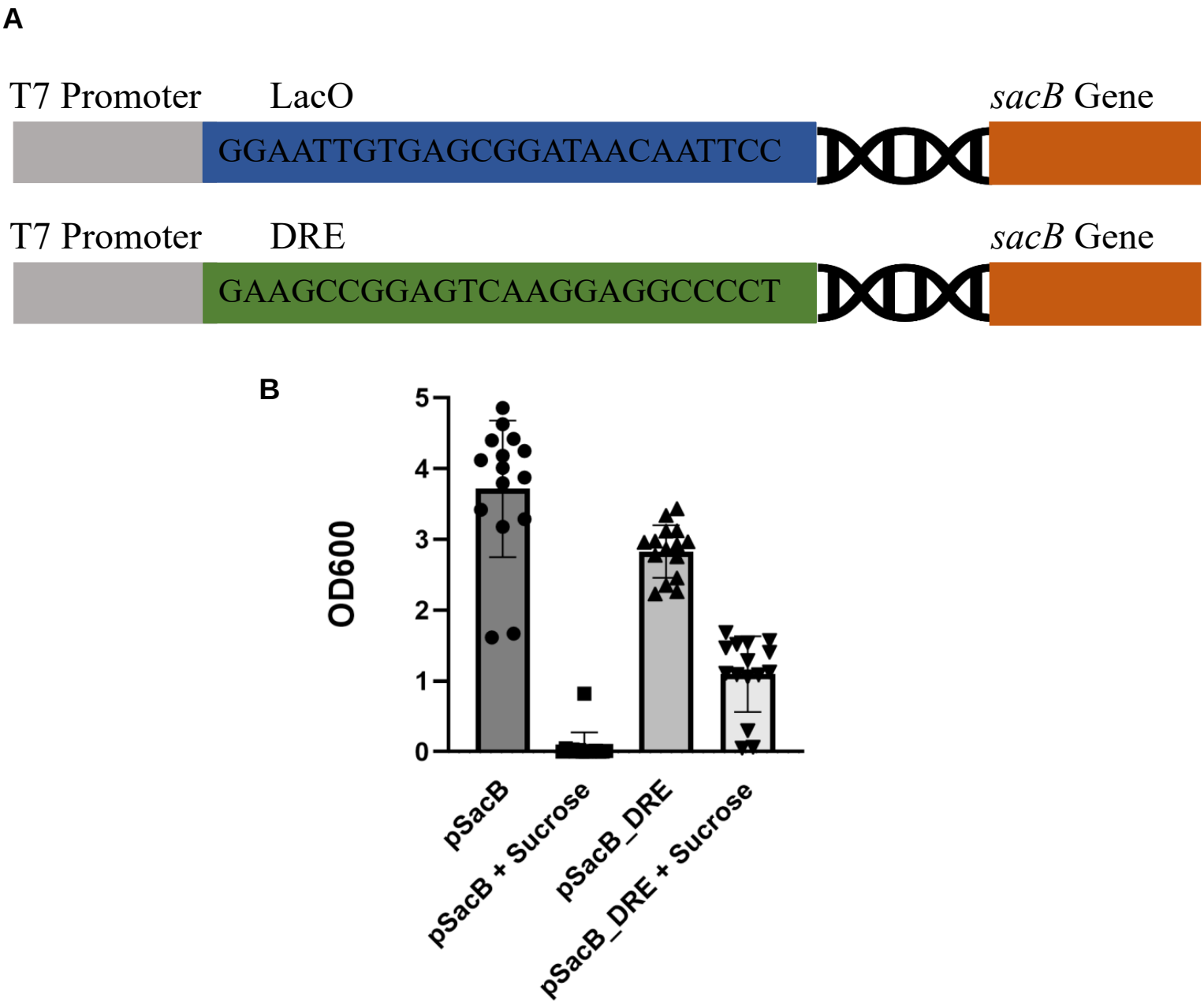


Figure 10. The Effect of DRE in place of LacO in Gene Expression (A) Schematic of pSacB (top) and pSacB_DRE (bottom) vectors. In pSacB_DRE the hDRE is inserted in place of a pre-existing lacO sequence. Both the lacO and DRE sequence are 25 bp. (B) Transformed BL21gold DE3 colonies inoculated in 8 mL LB (supplemented with 0.1 mM IPTG and 25 ug/mL chloramphenicol) with 400 uL of 0.5 g/ml sterile filtered sucrose or 400 uL sterile filtered distilled water grown at 37 °C and 250 rpm for 16 hrs post-inoculation. Error bars represent plus or minus one standard deviation of the mean. N=15, 3 colonies inoculated for each transformation and 5 transformations. p values and statistical analysis in Appendix II.

2.3.3. Recovery of Gene Expression by Insertion of DREAM in pSacB_DRE

The DRE sequence 'represses' *sacB* expression. Unexpectedly, the expression of DREAM recovers the expression of the *sacB* gene, preventing cell growth in the presence of sucrose. The DREAM gene was inserted into both pSacB and pSacB_DRE, and denoted as pSacB_DREAM, and pSacB_DREAM_DRE, respectively. Both pSacB_DREAM and pSacB_DREAM_DRE were transformed in BL21g DE3 cells. As with previous growth assays, expression of *SacB* was assessed by cell growth in the presence or absence of 2.5% sucrose. In the presence of sucrose, pSacB_DREAM and pSacB_DREAM_DRE demonstrated a lack of cell growth similar to that of pSacB in the presence of sucrose. In the presence of sucrose, the differences in cell growth between pSacB_DRE and both pSacB_DREAM and pSacB_DREAM_DRE were significant ($p<0.05$) (Figure 11). The addition of DREAM into the pSacB_DRE vector recovers the expression of the *sacB* gene sufficiently enough to prevent cell growth.

We also inserted the DREAM-C gene into pSacB and pSacB_DRE. The resulting plasmids were denoted as pSacB_DREAMC and pSacB DREAMC_DRE, respectively. Cell growth of the pSacB_DREAMC and pSacB DREAMC_DRE conditions were similar to that of pSacB_DREAM and pSacB_DREAM_DRE, respectively (Figure 11). Cells containing pSacB_DREAMC_DRE demonstrated minimal growth (mean OD600 of 0.020) in the presence of sucrose; significantly less than the growth of cells in the absence of sucrose ($p<0.05$) and significantly less than the growth of cells in the pSacB_DRE containing cells in the presence of sucrose ($p<0.05$).

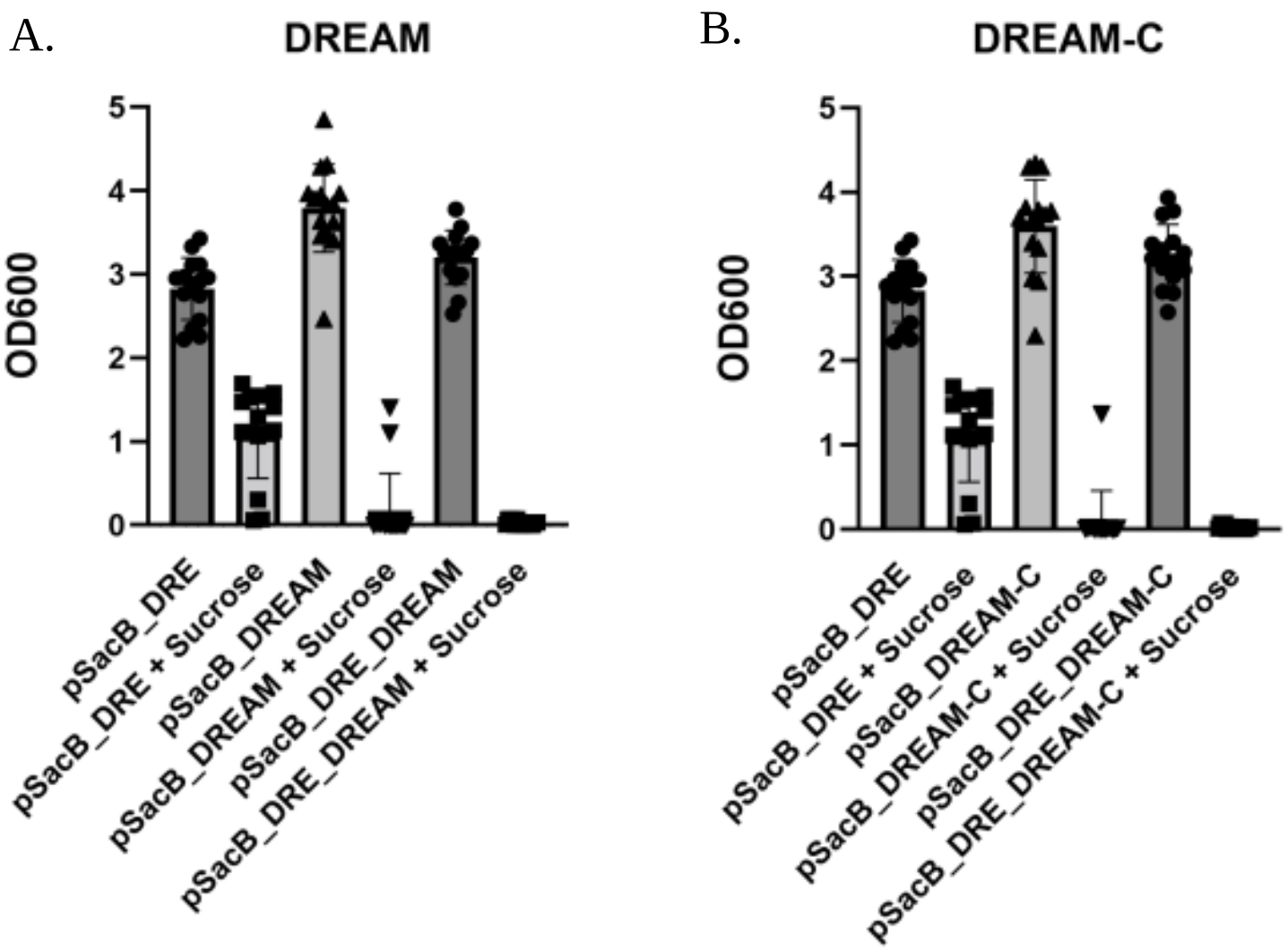


Figure 11. A. Cell Growth of BL21gold DE3 cells containing either pSacB_DRE, pSacB_DREAM, or pSacB_DRE_DREAM. (A) Cell growth of DREAM vectors (pSacB_DREAM, pSacB_DRE_DREAM) compared to cell growth of pSacB_DRE (left). (B). Cell growth of DREAMC vectors (pSacB_DREAM C, pSacB_DRE_DREAMC) compared to cell growth of pSacB_DRE (right). BL21g DE3 cells transformed with the respective plasmids were inoculated in 8 mL LB (supplemented with 0.1 mM IPTG and 25 ug/mL chloramphenicol) and either 2.5% (w/v) sucrose or the equivalent volume of sterile filtered distilled water grown at 37 °C and 250 rpm for 16 hrs post-inoculation. OD was measured at a wavelength of 600 nm. Bars represent the mean of all replicates. Error bars represent plus or minus one standard deviation from the mean. N=15, 3 colonies inoculated for each transformation and 5 transformations.

.

2.3.4. Optimization of Cell Growth Assay

To increase cell growth of pSacB_DRE in the presence of sucrose, the volume of the assay was decreased to 4 ml from 8 ml. Cells containing pSacB_DRE grew to a mean OD600 of 2.762 in the 4 mL growth assay compared to an OD600 value of 1.096 in the 8ml growth assay. Cell growth in the presence of sucrose for the vectors pSacB_DREAM, pSacB_DREAMC, pSacB_DREAM_DRE, and pSacB_DREAMC_DRE, displayed a similar lack of growth in the 4 mL growth assay as seen in the 8 mL growth assay, with all OD600 values below 0.070 OD600 (Figure 12). All proceeding assays were performed with 4 mL culture volume to maximize differences in cell growth.

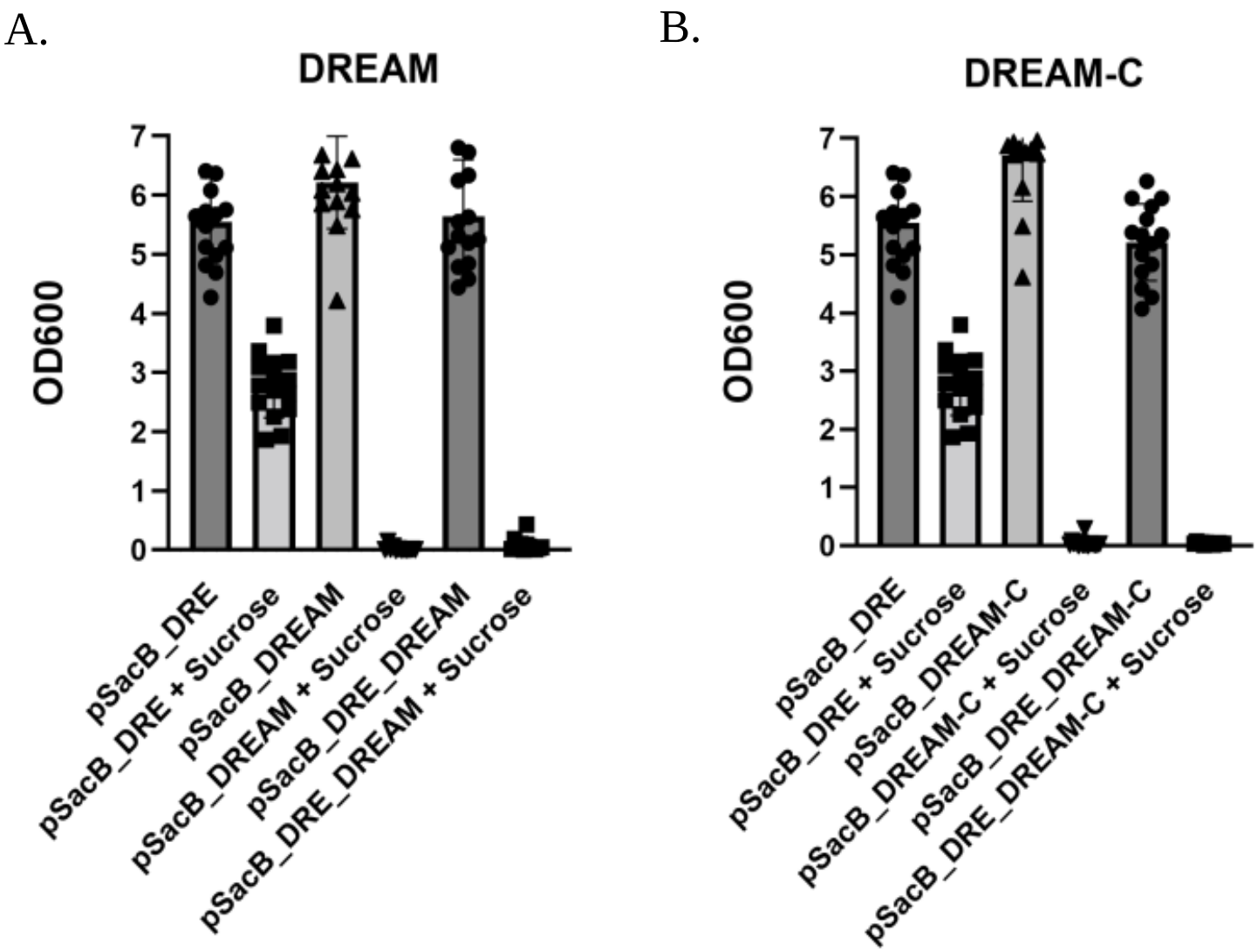


Figure 12. The Effects of the DRE and DREAM on Cell Growth as a Function of SacB Expression. Growth of BL21g DE3 cells transformed with respective plasmids containing either the DRE sequence, DREAM gene or both the DRE sequence and the DREAM gene. (A) Cell growth of DREAM vectors (pSacB_DREAM, pSacB_DRE_DREAM) compared to cell growth of pSacB_DRE (left). (B). Cell growth of DREAMC vectors (pSacB_DREAMC, pSacB_DRE_DREAMC) compared to cell growth of pSacB_DRE (right). Individual colonies were inoculated in 4 ml LB (supplemented with 0.1 mM IPTG and 25 ug/mL chloramphenicol) and either 2.5% (w/v) sucrose or the equivalent volume of sterile filtered distilled water grown at 37 °C and 250 rpm for 16 hrs post-inoculation. OD was measured at a wavelength of 600 nm. Bars represent the mean of all replicates. Error bars represent plus or minus one standard deviation from the mean. N=15, 3 colonies inoculated for each transformation and 5 transformations.

2.3.5. The Effect of a Histag Prior to DREAM

The DREAM protein expressed in pSacB_DREAM and pSacB_DREAM_DRE contains a histidine tag(his-tag), a feature of the pSacB vector. Given the unexpected repression by the DRE and subsequent unexpected recovery by DREAM, the effect of the his-tag on the ability of DREAM to recover *sacB* expression was assessed to determine if it impacted the DRE-DREAM interaction affecting cell growth. The his-tag present in pSacB vectors was removed using site-directed mutagenesis to assess the impact of the his-tag on the DREAM-DRE repression system. These new vectors were denoted pSacB_DREAM$_{hisdel}$ and pSacB_DRE_DREAM$_{hisdel}$. In the presence of sucrose, cells containing pSacB_DRE grew to a mean of 2.762 OD600, whereas cell containing pSacB_DRE_DREAM$_{hisdel}$, have a mean of 0.490. This difference in cell growth is statistically significant ($p<0.05$) Although not significant, cell growth in the presence of sucrose pSacB_DRE_DREAM$_{hisdel}$ displays a larger mean OD600 value of 0.490 compared to the pSacB_DRE_DREAM of 0.061 (Figure 12 and 13), largely due to three replicates where higher cell growth was seen. Both DREAM and DREAMC recover expression in a more reproducible manner (Figure 12). Overall, the same trends in cell growth are demonstrated regardless of the presence or absence of the his-tag in DREAM or DREAM$_{hisdel}$, respectively. However, DREAM$_{hisdel}$ recovers *sacB* expression (reduced by the DRE), in a less reproducible manner.

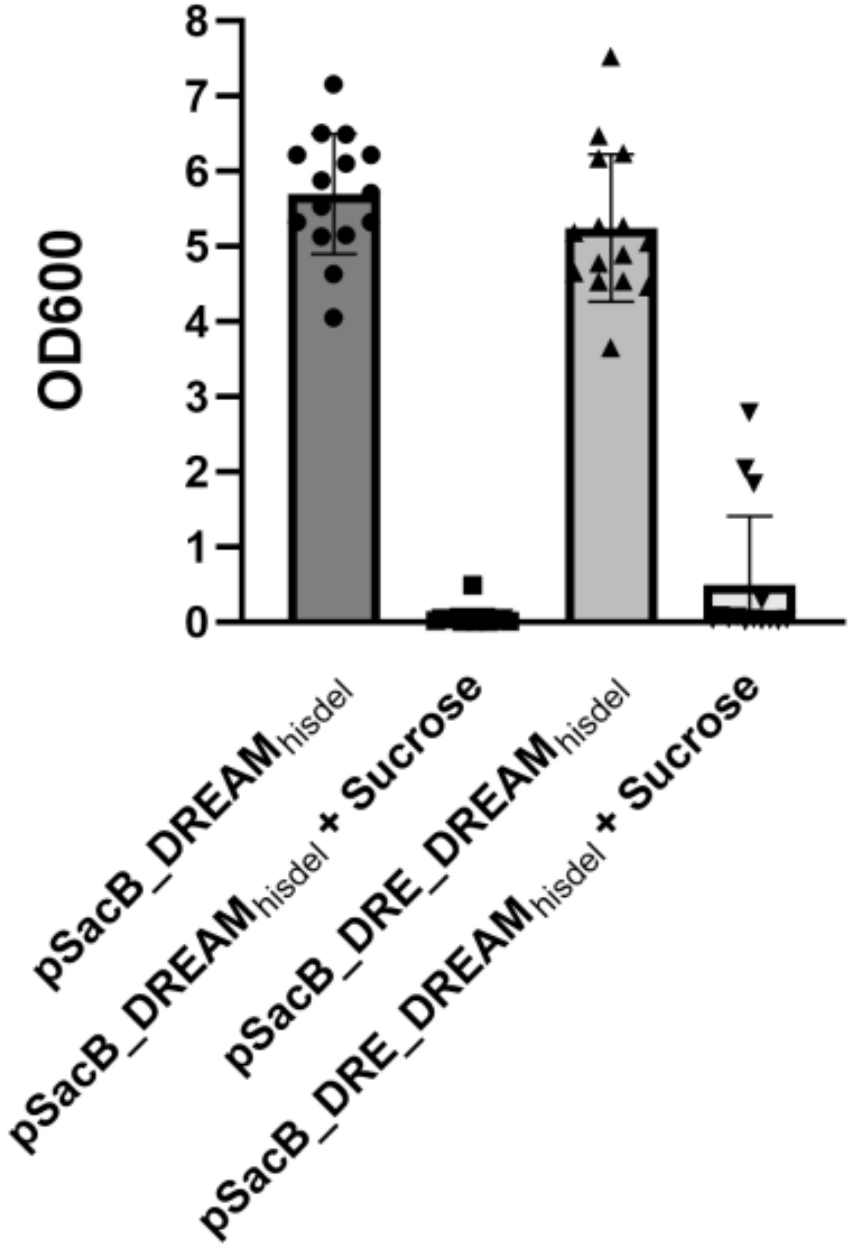


Figure 13. Cell Growth of Vectors expressing DREAM$_{hisdel}$. Growth of BL21g DE3 cells transformed with respective plasmids containing either the pSacB_DREAM$_{hisdel}$ or pSacB_DRE_DREAM$_{hisdel.}$ Individual colonies were inoculated in 4ml LB (supplemented with 0.1 mM IPTG and 25 ug/mL chloramphenicol) and either 2.5% (w/v) sucrose or the equivalent volume of sterile filtered distilled water grown at 37 °C and 250 rpm for 16 hrs post-inoculation. Optical density was measured at a wavelength of 600 nm. Bars represent the mean of all replicates. Error bars represent plus or minus one standard deviation from the mean. N=15, 3 colonies inoculated for each transformation and 5 transformations.

2.3.6. Further Assessment of the hDRE Sequence on Gene Expression

The DRE sequence caused an unexpected repression of *sacB* expression. The origin of this repression was unclear. To probe the unexpected impact on *saccB* gene expression by the insertion of the DRE sequence, a lacO/hDRE chimera was engineered, denoted lacODRE. The lacODRE chimera contains the first five base pairs of lacO and the last 20 bp of the DRE sequence (Figure 14A). The first five bp of lacO and the DRE sequence share some homology (two bp), and only three bp differ between the lacODRE and DRE sequences (Figure 14A). Interestingly, these three bp are necessary for the repression of *sacB* by the DRE sequence. In the presence of sucrose, the pSacB_DRE condition reaches a mean OD of 2.762 (Figure 12), where as pSacB_lacODRE only reached mean OD of 0.015 (Figure 14). In the presence of sucrose, pSacB_lacODRE demonstrates a lack of cell growth similar to that seen when in pSacB where the DRE sequence is not present in the vector at all. The central 'GTCA' essential for the DREAM-DRE binding is maintained in the lacODRE sequence. To assess if DREAM had a different affect on gene expression in the absence of the unexpected 'repression' by the DRE sequence, DREAM and DREAM variants DREAM-C and $DREAM_{hisdel}$, were inserted into the pSacB_lacODRE vector, denoted pSacB_lacODRE_DREAM, pSacB_lacODRE_DREAMC, and pSacB_lacODRE_$DREAM_{hisdel}$. In the presence of sucrose, little to no cell growth occurred, and as such no effect on *sacB* expression was observed (Figure 15).

A.

```
CLUSTAL O(1.2.4) multiple sequence alignment

DRE             GAAGCCGGAGTCAAGGAGGCCCCTG   25
lacODRE         GGAATCGGAGTCAAGGAGGCCCCTG   25
                * *  **********************
```

B.

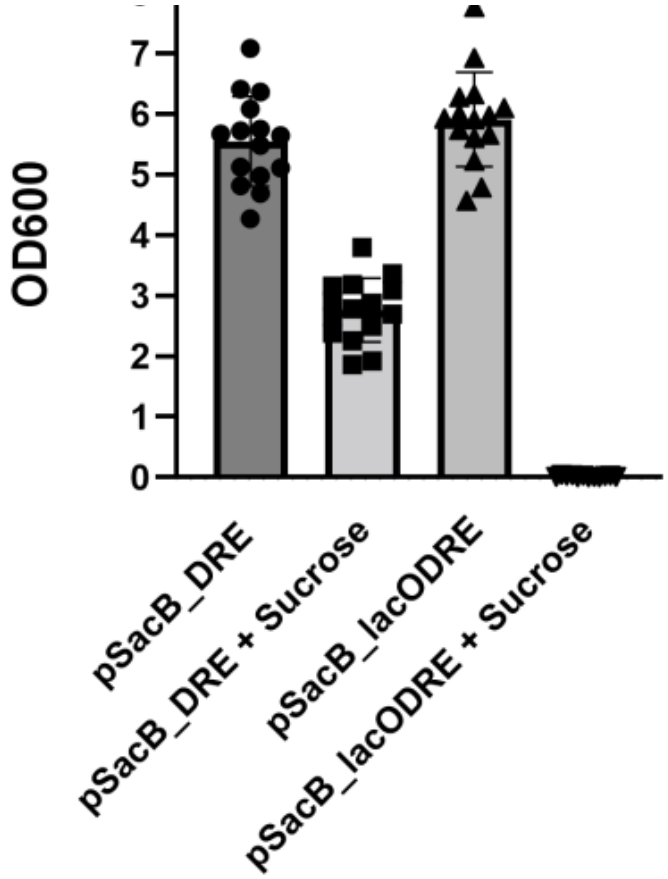


Figure 14. Growth of BL21g DE3 cells transformed with respective plasmids containing either the pSacB, pSacB_DRE or pSacB_lacODRE. (A) Alignment of the hDRE sequence in pSacB_DRE and the lacODRE chimera in pSacB_lacODRE (top). (B) Cell Growth of pSacB_DRE and pSacB_lacODRE (bottom). Individual colonies were inoculated in 4 ml LB (supplemented with 0.1 mM IPTG and 25 ug/mL chloramphenicol) and either 2.5% (w/v) sucrose or the equivalent volume of sterile filtered distilled water grown at 37 °C and 250 rpm for 16 hrs post-inoculation. Optical density was measured at a wavelength of 600 nm. Bars represent the mean of all replicates. Error bars represent deviation of the mean. N=15, 3 colonies inoculated for each transformation and 5 transformations.

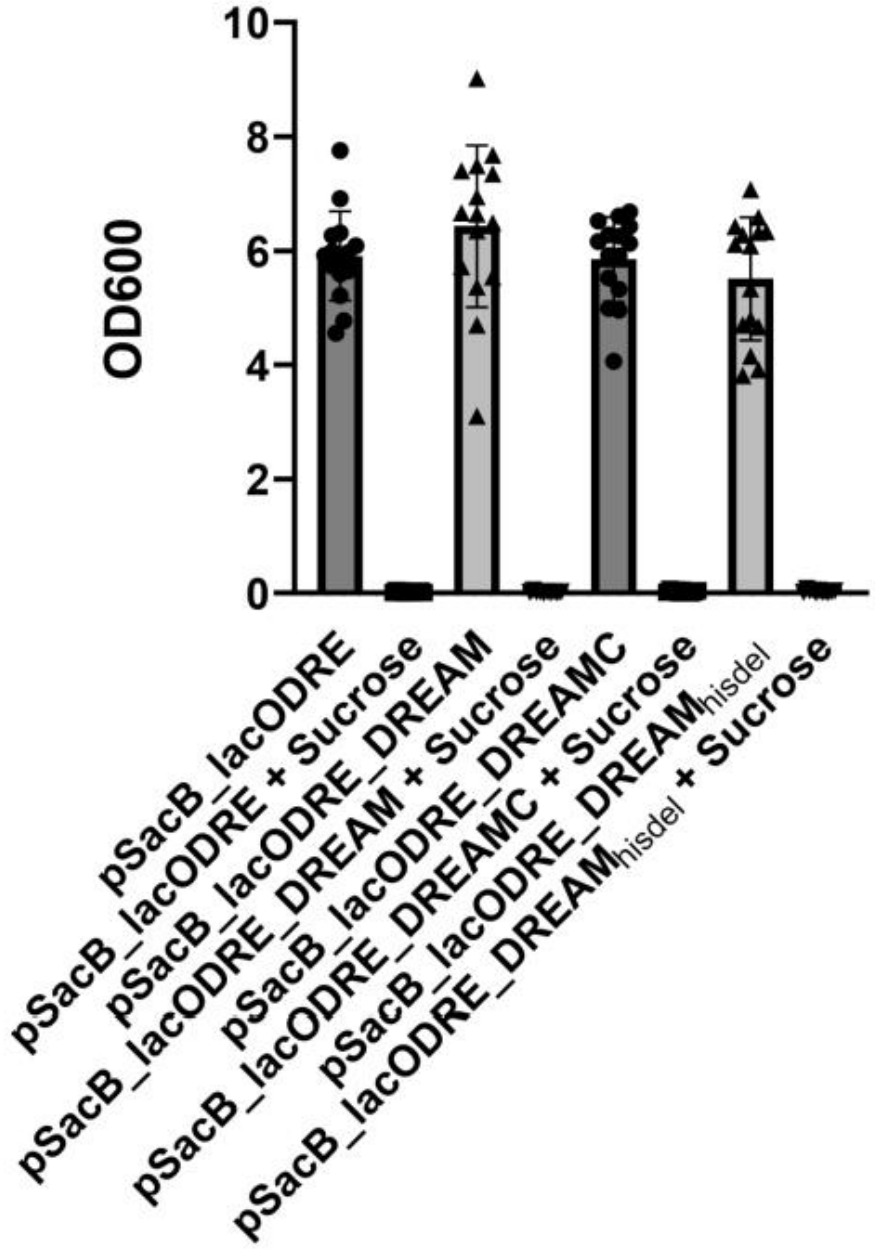


Figure 15. The Effect of DREAM (and DREAM variants) on the Expression of *sacB* Downstream the lacODRE Chimera. Individual colonies were inoculated in 4ml LB (supplemented with 0.1mM IPTG and 25ug/mL chloramphenicol) and either 2.5% (w/v) sucrose or the equivalent volume of sterile filtered distilled water grown at 37 °C and 250 rpm for 16 hrs post-inoculation. Optical density was measured at a wavelength of 600 nm. Bars represent the mean of all replicates. Error bars represent plus or minus the standard deviation of the mean. N=15, 3 colonies inoculated for each transformation and 5 transformations.

2.3.7. Single Nucleotide Substitution in DRE

To assess if the recovery of expression by DREAM is dependent on DRE-DREAM binding, a single nucleotide mutation shown to decrease the occupancy of DRE by DREAM (38) was installed in the DRE (Figure 16A). This adenine to thymine mutation in position 13 is the 25 bp DRE sequence occurs is the central 'GTCA' and is reported to reduce the occupancy of DRE and DREAM to 14% of the native DRE sequence (50). The DRE mutation, denoted DREmut, was installed in pSacB_DRE to create pSacB_DREmut. The pSacB_DREmut displayed similar cell growth in the presence of sucrose to that of pSacB_DRE (Figure 16) with a mean OD value of 2.252 compared to the 2.762 of pSacB_DRE. The DREmut demonstrated repression of the *sacB* gene similar to DRE, with no statistically significant differences. Given the reduced occupancy of the DREmut with DREAM, DREAM was not expected to recover the expression of *sacB* as it does with DRE. However, in a similar manner to the pSacB_DRE_DREAM plasmid, the addition of DREAM into the empty MCS of pSacB_DREmut recovered the expression of the *sacB* gene as demonstrated by the lack of cell growth in the presence of sucrose (Figure 17). This same trend is demonstrated for pSacB_DRE_DREAMC and pSacB_DRE_DREAM$_{hisdel}$, where the repression of the *sacB* gene by DREmut is recovered as reflected by the lack of cell growth in the presence of sucrose.

A.

```
DRE         GAAGCCGGAGTCAAGGAGGCCCCTG   25
DREmut      GAAGCCGGAGTCTAGGAGGCCCCTG   25
            ************ ************
```

B.

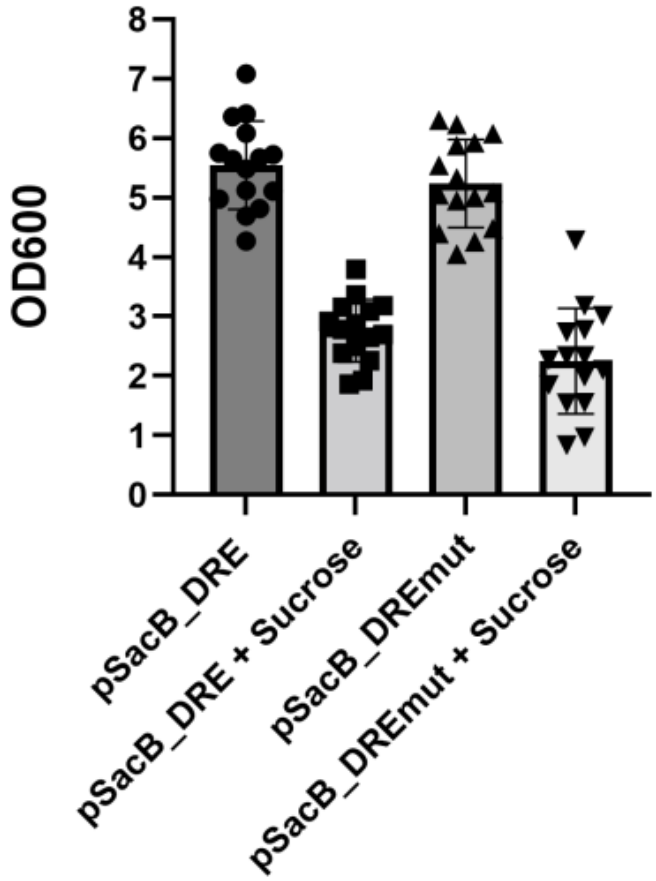


Figure 16. **Single nucleotide mutation in the hDRE sequence. (**A) Schematic of DRE sequence and a mutated DRE sequence (top). An Adenine to Thymine mutation installed in DRE at position 13 in the 3' to 5' direction. The A13T mutation is reported to reduce the occupancy of DRE by DREAM (38). (B) Growth of BL21g DE3 cells transformed with respective plasmids containing either the pSacB, pSacB_DRE or pSacB_lacODRE (bottom). Individual colonies were inoculated in 4 ml LB (supplemented with 0.1 mM IPTG and 25ug/mL chloramphenicol) and either 2.5% (w/v) sucrose or the equivalent volume of sterile filtered distilled water grown at 37 °C and 250 rpm for 16 hrs post-inoculation. OD was measured at a wavelength of 600 nm. Bars represent the mean of all replicates. Error bars represent plus or minus the standard deviation of the mean. N=15, 3 colonies inoculated for each transformation and 5 transformations.

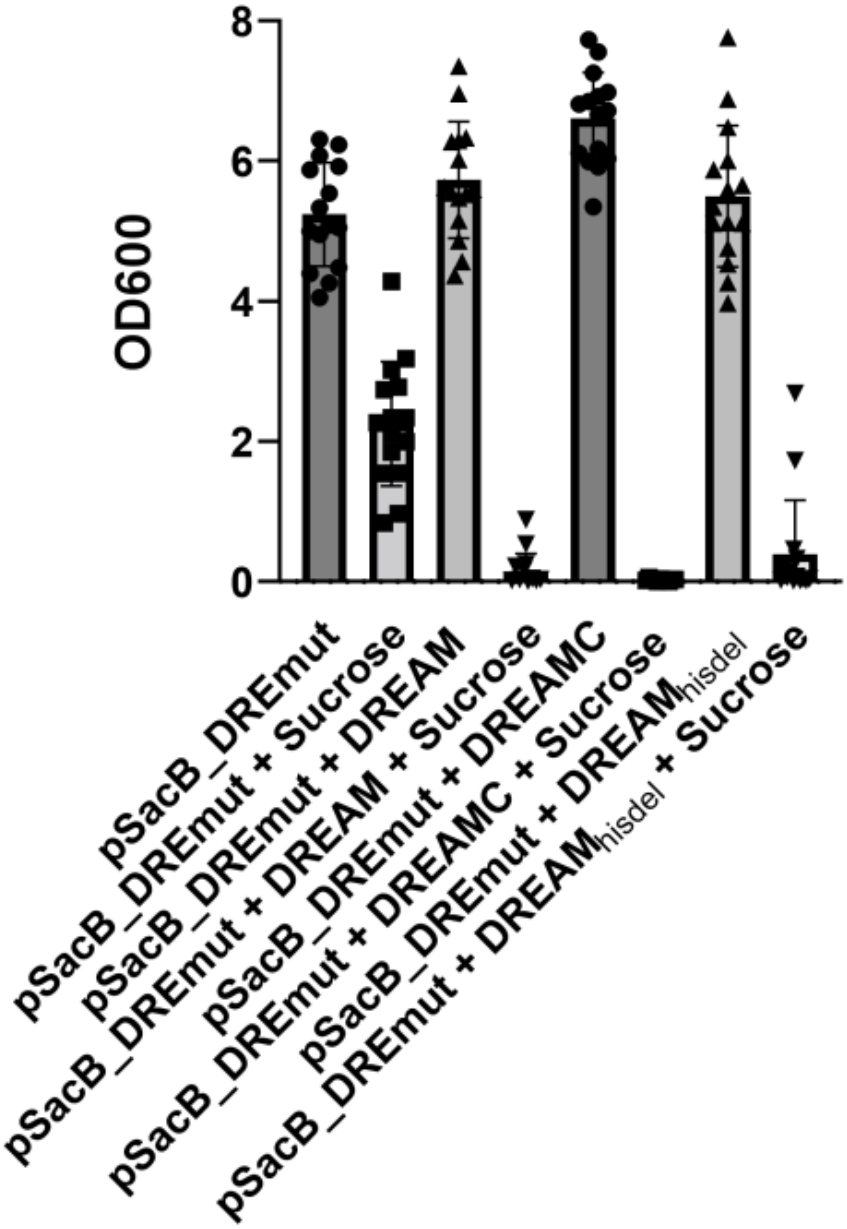


Figure 17. Growth of BL21g DE3 cells transformed with respective plasmids containing either the pSacB, pSacB_DRE or pSacB_lacODRE. Individual colonies were inoculated in 4ml LB (supplemented with 0.1 mM IPTG and 25 ug/mL chloramphenicol) and either 2.5% (w/v) sucrose or the equivalent volume of sterile filtered distilled water grown at 37 °C and 250 rpm for 16 hrs post-inoculation. Optical density was measured at a wavelength of 600 nm. Bars represent the mean of all replicates. Error bars represent plus or minus one standard deviation of the mean. N=15, 3 colonies inoculated for each transformation and 5 transformations.

2.4.0. Discussion

2.4.1. Cell Growth Assay

To assess the ability of DREAM to regulate gene expression in the absence of calcium, a cell growth assay system was set up linking the expression of *SacB* to cell growth. As DREAM, a calcium-dependent transcription factor, would be used downstream to screen for the presence of intracellular calcium, the reporter gene levansucrase was chosen to create a system in which the presence of calcium is linked to the life or death of the cells. The expression of the *sacB* gene, encoding levansucrase, directly impacts to the ability of the bacterial cells to grow in the presence of sucrose (39). A cell growth assay was set up such that in the presence of sucrose, cell growth is prevented by *sacB* expression. BL21g DE3 cells were transformed with pSacB or the empty control vector (pACYC DUET1$_X$). Cell growth, as a function of OD measurements, was measured in the presence or absence of sucrose 16 hrs after inoculation (Figure 17). Four different sucrose concentrations were assessed to find the minimum sucrose concentration required to prevent cell growth in levansucrase expressing cells. 0.5% (w/v) and 1% sucrose are both insufficient to reproducibly prevent cell growth in cells containing pSacB 16 hrs post-inoculation. (Figure 6A, 6B). Presumably, these sucrose concentrations are too low for sufficient accumulation of levan in the periplasm of cells that is necessary to prevent cell growth (57). While all cells containing pSacB grew in the absence of sucrose, both 2.5% and 5% sucrose consistently prevent cell growth in cells containing pSacB (Figure 6C, 6D). 2.5% sucrose was used for further experiments as the minimal concentration required to inhibit cell growth at 16 hrs. Cells transformed with the control vector grew in the presence and absence of sucrose. In the presence of sucrose, cell growth, or lack thereof, is

dependent on the expression of the *sacB* gene. As such, this cell growth assay can be used to assess the effect of the human prodynorphin DRE sequence and DREAM on gene expression in an easy, high throughput manner.

2.4.2. Repression of Gene Expression by DRE Sequence and Recovery by DREAM

To assess the ability of the DRE and DREAM system to repress gene expression, additional pSacB vectors were constructed to incorporate (1) the DRE sequence, (2) DREAM, or (3) both the DRE sequence and DREAM. Additional vectors were constructed with a C-terminal truncated version of DREAM, called DREAM-C. In mammalian cells, DREAM-C retains the ability to bind the DRE sequence and de-repress in a calcium dependent manner (31, 33). DREAM-C appears to be soluble when expressed in bacteria and has been extracted from bacterial cells (40). To construct these vectors, firstly, the 25 bp human prodynorphin DRE sequence was inserted immediately following the T7 RNAP promoter sequence in place of an existing 25 bp lacO sequence in pSacB to create pSacB_DRE; the only difference between these two vectors being the 25 bp DRE or lacO sequence (Figure 7). Interestingly, the presence of the human prodynorphin DRE sequence appears to 'repress' levansucrase expression as demonstrated by cell growth in the presence of 2.5% sucrose in cells containing pSacB _DRE (Figure 7). Cells containing pSacB _DRE grew to an average OD600 of 1.096 in the presence of sucrose compared to an average OD600 of 0.064. Even more unexpectedly, incorporation of DREAM or DREAM-C in the other multiple cloning site of pSacB_DRE recovers expression of levansucrase as demonstrated by the lack of cell growth in the presence of 2.5% sucrose at 16 hrs post-inoculation, with an average OD600 value of 0.029 (pSacB_DREAM_DRE) or 0.020

(pSacB _DREAMC_DRE). (Figure. 8A and 8B). Thus, the presence of DREAM or DREAM-C appeared to activate or recover expression of s*acB.* The difference in cell growth in the presence or absence of sucrose for the pSacB_DRE is statistically significant ($p<0.05$).

To further assess the ability of the DRE to repress levansucrase expression, culture volume was decreased from 8 ml to 4 ml to increase cell growth (61). Decreasing culture volume alone showed an increase in the cell growth at 16 hrs in the pSacB_DRE condition (Figure 9). In the 8 ml culture volume the mean OD600 for the DRE repressed cultures was 1.096 in the presence of 2.5% sucrose, whereas the OD600 of the cultures where levansucrase was expressed (DREAM alone or DRE and DREAM) was less than 0.018. Decreasing the culture volume to 4 mL increased the average OD600 for the DRE 'repressed" cultures to 2.762, while maintaining a low average OD600' under 0.070 OD600 for all levansucrase expressing bacterial cell cultures in the presence of 2.5% sucrose. Cells containing pSacB _DRE still showed a decrease in cell growth in the presence of sucrose compared to the absence of sucrose, indicating there may be some 'leaky' expression of levansucrase, and the DRE sequence alone may not completely prevent gene expression. However, the difference between the pSacB_DRE and pSacB cell growth is statistically significant ($p < 0.05$) in the presence of 2.5% sucrose. DREAM, unexpectedly, recovers the expression of *sacB* sufficiently enough to prevent cell growth in the presence of sucrose and these differences in OD are significant enough to differentiate visually (and statistically). In downstream directed evolution applications, this allows for easy visual determination of colonies containing a desired phenotype, allowing for easy, high throughput screening for plasmids containing genotypes of interest.

2.3.3. Interrogating Repression of Gene Expression by DRE

The insertion of the DRE was not expected to affect the expression of the *sacB* gene. To further investigate this repression, a lacO DRE chimera was created comprising of the first five bp of the lacO sequence and the last 20 bp of the DRE sequence (Figure 13A). The first five bp of lacO and DRE share two identical bases, only differing in three bases. These three bases are necessary for repression of *sacB* expression. The replacement of the first five bp of the DRE with five bp of lacO recovers the expression of *sacB*, and cells fail to grow in the presence of sucrose (Figure 13B). The first five bp of the DRE sequence play a critical role in the repression of gene expression. Different nucleotides in the +2 to +8 positions following the transcription start site of the T7 promoter have been shown to impact T7 promoter activity to over a five-fold degree (42). It is possible the three bp difference in the DRE decreases the T7 promoter activity. This could occur as a result of decreasing the affinity of T7 RNA polymerase for the sequence, or by interfering with the transcription bubble formed early in transcription that allows transcriptional machinery to interact with single stranded DNA. However, these three bp are not essential for DREAM and DRE binding, as DREAM largely interacts with the 'GTCA' sequence central to the DRE sequence (44, 50). However, in the absence of DRE-based repression; there is no apparent change in cell growth by the addition of DREAM (or DREAMC or $DREAM_{hisdel}$) in conjunction with the lacODRE chimera (Figure 14). There are a number of possible reasons that DREAM is not preventing gene expression. The T7 RNA

polymerase is a strong promoter, DREAM may simply not bind strongly enough to the DRE sequence to prevent the progression of the T7 RNA polymerase along the DNA.

2.3.4. A Single Nucleotide Mutation in DRE to Probe DREAM-DRE Interactions

DREAM appears to recover the expression of *sacB* repressed by the insertion of the DRE sequence. To determine if this increase in *sacB* expression is dependent on the DRE-DREAM interaction, a single nucleotide mutation known to affect the affinity of DREAM for the DRE was installed. In Electrophoretic Mobility Shift Assays, DRE molecules containing the A13T mutation demonstrated approximately a 14% binding occupancy with DREAM compared to the original DRE sequence (50). If an interaction between DRE and DREAM was necessary for the recovery of *sacB* expression (repressed by the DRE), we expected that the A13T mutation would prevent the recovery of expression by DREAM allowing cells to grow in the presence of sucrose. The mutant DRE maintains the DRE's ability to repress *sacB* expression (Figure. 15). However, this mutation does not appear to impact the ability of DREAM (or DREAM variants) to recover the expression of *sacB* as demonstrated by the minimal cell growth (Figure 16), suggesting that the recovery of *sacB* expression is either not dependent on the DRE-DREAM interaction, or that the minimal DREmut-DREAM interactions (14% binding occupancy (38)) are sufficient to allow enough expression of *sacB* to prevent cell growth.

2.3.5. Summary and Future Directions

There is still much to be learned about the control of gene expression by DREAM (and the DRE sequence) in bacterial systems. At this point, it is still unclear whether the DREAM DRE system has the potential to control gene expression in a calcium-dependent manner in bacteria. However, a number of future experiments may be performed to assess the DREAM DRE system in bacteria.

To assess if the decrease in gene expression is due to the proximity of the T7 promoter sequence, the DRE element could be moved farther downstream. In addition, an alternative lacO-DRE chimera could be assembled containing the first 5bp of the DRE and the last 20 bp of the lacO, should there be a decrease in expression here we could conclude that the decrease is not dependent on the whole DRE sequence but simply dependent on the nucleotides immediately following the T7 promoter. To explore how DREAM recovers the expression of levansucrase when the DRE is present, DREAM and levansucrase should not be transcriptionally linked. This would require the use of two different vectors, on expression levasucrase (with the DRE present) and one expression DREAM. This would provide a better understanding of the recovery of expression and indicate if the recovery is related to the transcriptional linkage between DREAM and levansucrase. In addition, a gene similar in size to DREAM should be inserted into the empty MCS in pSacB_DRE to ensure the recovery of expression is specifically dependent upon the expression of DREAM.

To further explore the potential for DREAM in a directed evolution strategy, an alternative expression system should be considered that uses a weaker promoter. It is possible that the T7 promoter system is too strong, and DREAM does not bind the DRE

strongly enough to prevent the progression of the T7 RNA polymerase along the DNA, If this is the case, a weaker RNA polymerase and promoter system may yield better results.

Thus far, the DREAM DRE system has been assessed with the final goal of linking calcium to cell growth, thus using the *sacB* gene as a reporter of the impacts of DREAM and the DRE sequence. However, this only provides a course output with relation to the level of *sacB* gene expression, an alternative reporter gene with a more quantitative output should be used assess the DREAM DRE system.

Lastly, should the DREAM DRE system be shown to repress gene expression, future experiments should assess the calcium dependent nature of this repression in bacteria. Should the DREAM DRE system prove to regulate gene expression in a calcium-dependent manner inn bacteria, the DREAM DRE system may be exploitable in bacterial directed evolution strategies aimed at engineering novel calcium-conducting ligand gated ion channels.

REFERENCES

1. Hanahan D. Studies on transformation of Escherichia coli with plasmids. J Mol Biol. 1983 Jun 5;166(4):557-80. doi: 10.1016/s0022-2836(83)80284-8. PMID: 6345791.

2. Smithsonian Institution Archives, 2020. *The History Of PCR (RU9577)*. Smithsonian Videohistory Collection. [online] Available at: <http://siarchives.si.edu/research/videohistory_catalog9577.html> [Accessed 18 April 2020].
3. Agarwal KL, Büchi H, Caruthers MH, Gupta N, Khorana HG, Kleppe K, Kumar A, Ohtsuka E, Rajbhandary UL, Van de Sande JH, Sgaramella V, Weber H, Yamada T. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. Nature. 1970 Jul 4;227(5253):27-34. doi: 10.1038/227027a0. PMID: 5422620.

4. Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Isolating, Cloning, and Sequencing DNA. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26837/
5. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front Microbiol. 2015 Mar 31;6:242. doi: 10.3389/fmicb.2015.00242. PMID: 25873913; PMCID: PMC4379921.
6. Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. *Nature Education* 1(1):100
7. Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. The Initiation and Completion of DNA Replication in Chromosomes. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26826/
8. Ghooi RB, Thatte SM. Inhibition of cell wall synthesis--is this the mechanism of action of penicillins? Med Hypotheses. 1995 Feb;44(2):127-31. doi: 10.1016/0306-9877(95)90085-3. PMID: 7596307.
9. Kaushik D, Mohan M, Borade DM, Swami OC. Ampicillin: rise fall and resurgence. J Clin Diagn Res. 2014 May;8(5):ME01-3. doi: 10.7860/JCDR/2014/8777.4356. Epub 2014 May 15. PMID: 24995206; PMCID: PMC4080027.
10. Deaner M, Alper HS. Promoter and Terminator Discovery and Engineering. Adv Biochem Eng Biotechnol. 2018;162:21-44. doi: 10.1007/10_2016_8. PMID: 27277391.
11. Makrides SC. Vectors for gene expression in mammalian cells. New Comprehensive Biochemistry. 2003;38:9–26. doi: 10.1016/S0167-7306(03)38002-0. Epub 2004 Jan 7. PMCID: PMC7147855.

12. Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Transcription in Prokaryotes. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9850/
13. Fs
14. Willemsen, A., & Zwart, M. P. (2019). On the stability of sequences inserted into viral genomes. *Virus evolution, 5*(2), vez045. https://doi.org/10.1093/ve/vez045
15. Hiszczyńska-Sawicka E, Kur J. Effect of Escherichia coli IHF mutations on plasmid p15A copy number. Plasmid. 1997;38(3):174-9. doi: 10.1006/plas.1997.1307. PMID: 9435019.

16. Thompson MG, Sedaghatian N, Barajas JF, Wehrs M, Bailey CB, Kaplan N, Hillson NJ, Mukhopadhyay A, Keasling JD. Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. Sci Rep. 2018 Jan 25;8(1):1590. doi: 10.1038/s41598-018-20016-w. PMID: 29371642; PMCID: PMC5785507.
17. Belshaw R, Pybus OG, Rambaut A. The evolution of genome compression and genomic novelty in RNA viruses. Genome Res. 2007 Oct;17(10):1496-504. doi: 10.1101/gr.6305707. Epub 2007 Sep 4. PMID: 17785537; PMCID: PMC1987338.
18. Goodwin EC, Rottman FM. The 3'-flanking sequence of the bovine growth hormone gene contains novel elements required for efficient and accurate polyadenylation. J Biol Chem. 1992 Aug 15;267(23):16330-4. PMID: 1644817.
19. Kessler MM, Beckendorf RC, Westhafer MA, Nordstrom JL. Requirement of A-A-U-A-A-A and adjacent downstream sequences for SV40 early polyadenylation. Nucleic Acids Res. 1986 Jun 25;14(12):4939-52. doi: 10.1093/nar/14.12.4939. PMID: 3014439; PMCID: PMC311502.
20. Slater, M., Hartzell, D., Hartnett, J., Wheeler, S., Stecha, P., Karissima, N. and Promega Corporation, 2008. Achieve the Protein Expression Level You Need with the Mammalian HaloTag 7 Flexi Vectors. *Promega Notes*, 100, pp.16-18.
21. snapgene.com n.d. *pFC17K HaloTag® CMVd3* [online] Available at: < https://www.snapgene.com/resources/plasmid-files/?set=mammalian_expression_vectors&plasmid=pFC17K_HaloTag_CMVd3 > [Accessed 18 April 2020]
22. snapgene.com n.d. *pFN23K HaloTag® CMVd2* [online] Available at: < https://www.snapgene.com/resources/plasmid-files/?set=mammalian_expression_vectors&plasmid=pFN23K_HaloTag_CMVd2 > [Accessed 18 April 2020]
23. snapgene.com n.d. *pFC15A HaloTag® CMVd1* [online] Available at: < https://www.snapgene.com/resources/plasmid-files/?set=mammalian_expression_vectors&plasmid=pFC15A_HaloTag_CMVd1 > [Accessed 18 April 2020]
24. Rubaiy, H., 2017. A Short Guide to Electrophysiology and Ion Channels. *Journal of Pharmacy & Pharmaceutical Sciences*, 20, p.48.
25. Gray D. Overview of protein expression by mammalian cells. Curr Protoc Protein Sci. 2001 May;Chapter 5(1):Unit5.9. doi: 10.1002/0471140864.ps0509s10. PMID: 18429190; PMCID: PMC7162368.
26. C Kopta, JH Steinbach *Comparison of mammalian adult and fetal nicotinic acetylcholine receptors stably expressed in fibroblasts*, Journal of Neuroscience 1 June 1994, 14 (6) 3922-3933; **DOI:** 10.1523/JNEUROSCI.14-06-03922.1994
27. Lesueur LL, Mir LM, André FM. Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro. Mol Ther Nucleic Acids. 2016 Mar 8;5(3):e291. doi: 10.1038/mtna.2016.4. PMID: 27111417; PMCID: PMC5014460.
28. snapgene.com n.d. *pcDNA3.1(+)* [online] Available at: < https://www.snapgene.com/resources/plasmid-

files/?set=basic_cloning_vectors&plasmid=pcDNA3.1(%2B)> [Accessed 30 September 2022]
29. snapgene.com n.d. *pCI* [online] Available at: < https://www.snapgene.com/resources/plasmid-files/?set=mammalian_expression_vectors&plasmid=pCI > [Accessed 18 April 2020]
30. Becker NA, Peters JP, Maher LJ 3rd, Lionberger TA. Mechanism of promoter repression by Lac repressor-DNA loops. Nucleic Acids Res. 2013 Jan 7;41(1):156-66. doi: 10.1093/nar/gks1011. Epub 2012 Nov 9. Erratum in: Nucleic Acids Res. 2013 Apr 1;41(6):3962. Lionberger, Troy A [added]. PMID: 23143103; PMCID: PMC3592455.
31. Carrión, A., Link, W., Ledo, F. *et al.* DREAM is a Ca^{2+}-regulated transcriptional repressor. *Nature* **398**, 80–84 (1999). https://doi.org/10.1038/18044
32. Lusin JD, Vanarotti M, Li C, Valiveti A, Ames JB. NMR structure of DREAM: Implications for Ca(2+)-dependent DNA binding and protein dimerization. Biochemistry. 2008 Feb 26;47(8):2252-64. doi: 10.1021/bi7017267. Epub 2008 Jan 18. PMID: 18201103.
33. Cheng HY, Pitcher GM, Laviolette SR, Whishaw IQ, Tong KI, Kockeritz LK, Wada T, Joza NA, Crackower M, Goncalves J, Sarosi I, Woodgett JR, Oliveira-dos-Santos AJ, Ikura M, van der Kooy D, Salter MW, Penninger JM. DREAM is a critical transcriptional repressor for pain modulation. Cell. 2002 Jan 11;108(1):31-43. doi: 10.1016/s0092-8674(01)00629-8. PMID: 11792319.
34. Carrión AM, Mellström B, Naranjo JR. Protein kinase A-dependent derepression of the human prodynorphin gene via differential binding to an intragenic silencer element. Mol Cell Biol. 1998 Dec;18(12):6921-9. doi: 10.1128/MCB.18.12.6921. PMID: 9819380; PMCID: PMC109275.
35. Mellström B, Naranjo JR. Ca2+-dependent transcriptional repression and derepression: DREAM, a direct effector. Semin Cell Dev Biol. 2001 Feb;12(1):59-63. doi: 10.1006/scdb.2000.0218. PMID: 11162748.
36. Li H, Ullrich MS. Characterization and mutational analysis of three allelic lsc genes encoding levansucrase in Pseudomonas syringae. J Bacteriol. 2001 Jun;183(11):3282-92. doi: 10.1128/JB.183.11.3282-3292.2001. PMID: 11344135; PMCID: PMC99625.
37. Hiszczyńska-Sawicka E, Kur J. Effect of Escherichia coli IHF mutations on plasmid p15A copy number. Plasmid. 1997;38(3):174-9. doi: 10.1006/plas.1997.1307. PMID: 9435019.
38. Ledo F, Link WA, Carrión AM, Echeverria V, Mellström B, Naranjo JR. The DREAM-DRE interaction: key nucleotides and dominant negative mutants. Biochim Biophys Acta. 2000 Dec 20;1498(2-3):162-8. doi: 10.1016/s0167-4889(00)00092-6. PMID: 11108959.
39. Li H, Ullrich MS. Characterization and mutational analysis of three allelic lsc genes encoding levansucrase in Pseudomonas syringae. J Bacteriol. 2001 Jun;183(11):3282-92. doi: 10.1128/JB.183.11.3282-3292.2001. PMID: 11344135; PMCID: PMC99625.
40. Osawa M, Tong KI, Lilliehook C, Wasco W, Buxbaum JD, Cheng HY, Penninger JM, Ikura M, Ames JB. Calcium-regulated DNA binding and oligomerization of the neuronal calcium-sensing protein, calsenilin/DREAM/KChIP3. J Biol Chem. 2001 Nov 2;276(44):41005-13. doi: 10.1074/jbc.M105842200. Epub 2001 Sep 4. PMID: 11535596.
41. Osawa M, Dace A, Tong KI, Valiveti A, Ikura M, Ames JB. Mg2+ and Ca2+ differentially regulate DNA binding and dimerization of DREAM. J Biol Chem. 2005 May 6;280(18):18008-14. doi: 10.1074/jbc.M500338200. Epub 2005 Mar 3. PMID: 15746104

42. Conrad, T., Plumbom, I., Alcobendas, M. *et al*. Maximizing transcription of nucleic acids with efficient T7 promoters. *Commun Biol* **3**, 439 (2020). https://doi.org/10.1038/s42003-020-01167-x

APPENDIX I

I.I Alignment of CMV Promoter and Enhancer Variants

C1/1-184 GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC GGGGTCATTA

C2/1-508 ---------- ---------- ---------- ---------- ----------

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-584 GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG

C2/1-508 ---------- ---------- ------CGTT ACATAACTTA CGGTAAATGG

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA

C2/1-584 CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 CGTATGTTCC CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG

C2/1-584 CGTATGTTCC CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 GTGGAGTATT TACGGTAAAC TGCCCACTTG GCAGTACATC AAGTGTATCA

C2/1-584 GTGGAGTATT TACGGTAAAC TGCCCACTTG GCAGTACATC AAGTGTATCA

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 TATGCCAAGT ACGCCCCCTA TTGACGTCAA TGACGGTAAA TGGCCCGCCT

C2/1-584 TATGCCAAGT ACGCCCCCTA TTGACGTCAA TGACGGTAAA TGGCCCGCCT

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC TTGGCAGTAC

C2/1-584 GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC TTGGCAGTAC

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA

C2/1-584 ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC

C2/1-584 CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC

C3/1-100 ---------- ---------- ---------- ---------- ----------

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 ACCCCATTGA CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC

C2/1-584 ACCCCATTGA CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC

C3/1-100 ---------- ---------- ---------- ----CCAAAA TCAACGGGAC

C4/1-52 ---------- ---------- ---------- ---------- ----------

C5/1-45 ---------- ---------- ---------- ---------- ----------

C1/1-508 TTTCCAAAAT GTCGTAACAA CTCCGCCCCA TTGACGCAAA TGGGCGGTAG

C2/1-584 TTTCCAAAAT GTCGTAACAA CTCCGCCCCA TTGACGCAAA TGGGCGGTAG

C3/1-100 TTTCCAAAAT GTCGTAACAA CTCCGCCCCA TTGACGCAAA TGGGCGGTAG

C4/1-52 ---------- ---------- ---------- -- GACGCAAA TGGGCGGTAG

C5/1-45 ---------- ---------- ---------- --------- A TGGGCGGTAG

C1/1-508 GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCT

C2/1-584 GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCT

C3/1-100 GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCT

C4/1-52 GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCT

C51-45 GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCT

Appendix Figure 1. **Alignments of the five versions of the cytomegalovirus promoters implemented in pUdO mammalian plasmids.** Alignment shows only the CMV region of each mammalian plasmid. Alignments were performed with Jalview software. The TATA box is highlighted for all promoters.

APPENDIX I.II Vectors Assembled Prior to and During my MSc.

Table 6. Vectors Assembled Prior to MSc.

Vector	**Insert (If applicable)**
pUdO1a-C1S	N/A
pUdO1a-C2S	N/A
pUdO1a-C3S	N/A
pUdO1a-C4S	N/A
pUdO1a-C5S	N/A
pUdO1c-C1S	N/A
pUdO1t-C1S	N/A
pUdO1a-C1S + GFP	GFP
pUdO1a-C1S + alpha	nAchR alpha subunit
pUdO1a-C1S + beta	nAchR bets subunit
pUdO1a-C1S + delta	nAchR delta subunit
pUdO1a-C1S + gamma	nAchR gamma subunit
pUdO1a-C1S + epsilon	nAchR epsilon subunit

Table 7. Vectors assembled during MSc.

Vector	**Insert (If applicable)**
pUdO1z-C1S	N/A
pUdO1a-C1B	N/A
pUdO1a-C2B	N/A
pUdO1a-C3B	N/A
pUdO1a-C4B	N/A
pUdO1a-C5B	N/A
pUdO1c-C1B	N/A
pUdO1t-C1B	N/A
pUdO1z-C1B	N/A
pUdO1a-C1S + Luciferase	Luciferase
pUdO1a-C2S + Luciferase	Luciferase
pUdO1a-C3S + Luciferase	Luciferase
pUdO1a-C4S + Luciferase	Luciferase
pUdO1a-C5S + Luciferase	Luciferase
pUdO1c-C1S + Luciferase	Luciferase
pUdO1t-C1S + Luciferase	Luciferase
pUdO1z-C1S + Luciferase	Luciferase
pUdO1a-C1B + Luciferase	Luciferase
pUdO1a-C2B + Luciferase	Luciferase
pUdO1a-C3B + Luciferase	Luciferase
pUdO1a-C4B + Luciferase	Luciferase
pUdO1a-C5B + Luciferase	Luciferase
pUdO1c-C1B + Luciferase	Luciferase
pUdO1t-C1B + Luciferase	Luciferase
pUdO1z-C1B + Luciferase	Luciferase

Table 8. Non-pUdO Vectors Assembled

Vector	Insert
pcDNA3.1 + Luciferase	Luciferase
pcDNA.03 + Luciferase	Luciferase
pRBG4 + Luciferase	Luciferase

APPENDIX I.III Empty pUdO Expression Vector Sequences

All sequences shown were verified by Sanger Sequencing

> pUdO1a

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGCTCCGAAGAGCCGAGTGTCTTC
GGTTCGAGACCGAATTCGCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATC
GATTTGGATCCACCCTGCAGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTT
CGACGTAAGTAACTGACTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCG
TTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACATTCAAATATGTATCCGCTCATGAGACAATAACCC
TGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATT
CCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCT
GAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAG
TTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATC
CCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGT
ACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCG
CTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCC
ATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAA
CTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCA
GGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGT
GGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAC
GACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATT
AAGCATTGGTAATAATACGACTCACTATAGGGTAACTAATTGAGGCGCGCCTACCACCGTCAAAAAAA
ACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1c

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGCTCCGAAGAGCCGAGTGTCTTC
GGTTCGAGACCGAATTCGCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATC
GATTTGGATCCACCCTGCAGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTT
CGACGTAAGTAACTGACTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCG
TTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACAACGTAAGAGGTTCCAACTTTCACCATAATGAAAT
AAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGA
AAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTT
CAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTA
AAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCG
GAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGT
TTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCT
ACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGA
GAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATAT
GGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGC
CGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTAC
AACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATAATACGACTCACTATAGGGTAACTAATTGAG
GCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTA
AGTGACCCTTTAGTGAGGGTTAATT

>pUdO1t

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGCTCCGAAGAGCCGAGTGTCTTC
GGTTCGAGACCGAATTCGCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATC
GATTTGGATCCACCCTGCAGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTT
CGACGTAAGTAACTGACTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCG
TTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACACTGTTGACAATTAATCATCGGCATAGTATATCGG
CATAGTATAATACGACAAGGTGAGGAACTAAACCATGGGTCAAAGTAGCGATGAAGCCAACGCTCCC
GTTGCAGGGCAGTTTGCGCTTCCCCTGAGTGCCACCTTTGGCTTAGGGGATCGCGTACGCAAGAAATC
TGGTGCCGCTTGGCAGGGTCAAGTCGTCGGTTGGTATTGCACAAAACTCACTCCTGAAGGCTATGCGG
TCGAGTCCGAATCCCACCCAGGCTCAGTGCAAATTTATCCTGTGGCTGCACTTGAACGTGTGGCCTAAT
AATACGACTCACTATAGGGTAACTAATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTA
GCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTGACCCTTTAGTGAGGGTTAAT

>pUdO1z

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGCTCCGAAGAGCCGAGTGTCTTC
GGTTCGAGACCGAATTCGCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATC
GATTTGGATCCACCCTGCAGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTT
CGACGTAAGTAACTGACTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCG
TTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACAGTGTTGACAATTAATCATCGGCATAGTATATCGG
CATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCA
CCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTG
GAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGT
GGTGCCGGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCG
GAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGT
GGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGA
CTGATAATACGACTCACTATAGGGTAACTAATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCT
TTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1a-C1S

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGACATTGATTATTGACTAGTTAT
TAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC
GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACG
GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT
TCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAG
CCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACC
GTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGAAACTTGTTTATTGCAG
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATT
CTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCTCGGTACCAAATTCCAGA
AAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACAT
TCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAG
TATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTC
ACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGA
ACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCA
CTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCC
GCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGC
ATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCT
GACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCC
TTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGT
AGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAAT
TAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGG
TTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGA
TGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATA
GACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTATAGGGTAAC
TAATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTT
CCTAGGTAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1a-C2S

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGCGTTACATAACTTACGGTAAAT
GGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAG
TACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGG
CATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCT
ATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT
TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA
ATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAA
GCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAGCCAGTT
TAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACCGTCGAC
ACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGAAACTTGTTTATTGCAGCTTATA
ATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTT
GTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCTCGGTACCAAATTCCAGAAAAGAG
GCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACATTCAAAT
ATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAG
TATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG
AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGA
TCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAA
AGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATAC
ACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACA
GTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAAC
GATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATC
GTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAAT
GGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAG
ACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATT
GCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAA
GCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGA
TCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTATAGGGTAACTAATTG
AGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGG
TAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1a-C3S

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGATGGGCGGTAGGCGTGTACGGT
GGGAGGTCTATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGC
TCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAG
TCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGAAACTTGTT
TATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTC
ACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCTCGGTACCAA
ATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCG
TTTTACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAA
GGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGT
TTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATG
ATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACT
CGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTA
CGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAA
CTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATG
TAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCAC
GATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCC
GGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCG
GCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACT
GGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATG
AACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTA
TAGGGTAACTAATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTT
TTCAACCTTCCTAGGTAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1a-C4S

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGACGCAAATGGGCGGTAGGCGT
GTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATT
CGCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTG
CAGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGAA
ACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCA
TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCTCG
GTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACT
GGCCGTCGTTTTACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATA
TTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTG
CCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCAC
GAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGT
TTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAA
GAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAA
GCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTG
CGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGG
GATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGA
CACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAG
CTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCC
CTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCA
GCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTAT
GGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACT
CACTATAGGGTAACTAATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTT
TTATTTTTCAACCTTCCTAGGTAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1a-C5S

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGCCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATAT
AAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAGCCAG
TTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACCGTCG
ACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGAAACTTGTTTATTGCAGCTTAT
AATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGT
TGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCTCGGTACCAAATTCCAGAAAAGA
GGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACATTCAAA
TATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGA
GTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCA
GAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGG
ATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTA
AAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATA
CACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGAC
AGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAA
CGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGAT
CGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAA
TGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATA
GACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTAT
TGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTA
AGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAG
ATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTATAGGGTAACTAATT
GAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAG
GTAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1c-C1S

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGACATTGATTATTGACTAGTTAT
TAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC
GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACG
GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT
TCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAG
CCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACC
GTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGAAACTTGTTTATTGCAG
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATT
CTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCTCGGTACCAAATTCCAGA
AAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACAA
CGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAG
ATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCC
AATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTC
AGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTC
ACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTG
ATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGG
AGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGA
AAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAG
TTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAA
ATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATG
GCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAA
TAATACGACTCACTATAGGGTAACTAATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTT
AGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1t-C1S

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAA
AAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT
GGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
GAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGA
TAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAA
CGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCG
TGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGG
GTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCG
GGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAA
AACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAG
AAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGACATTGATTATTGACTAGTT
ATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTA
CGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTT
CCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCA
CTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGC
CCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG
TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAC
GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGAC
TTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT
CTATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCT
AGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGA
CCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGAAACTTGTTTATTGC
AGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGC
ATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCTCGGTACCAAATTCCA
GAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTAC
ACTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAAC
CATGGGTCAAAGTAGCGATGAAGCCAACGCTCCCGTTGCAGGGCAGTTTGCGCTTCCCCTGAGTGCCA
CCTTTGGCTTAGGGGATCGCGTACGCAAGAAATCTGGTGCCGCTTGGCAGGGTCAAGTCGTCGGTTGG
TATTGCACAAAACTCACTCCTGAAGGCTATGCGGTCGAGTCCGAATCCCACCCAGGCTCAGTGCAAAT
TTATCCTGTGGCTGCACTTGAACGTGTGGCCTAATAATACGACTCACTATAGGGTAACTAATTGAGGC
GCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAG
TGACCCTTTAGTGAGGGTTAATT

>pUdO1z-C1S
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGACATTGATTATTGACTAGTTAT
TAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC
GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACG
GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT
TCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAG
CCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACC
GTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGAAACTTGTTTATTGCAG
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATT
CTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCCTCGGTACCAAATTCCAGA
AAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACAG
TGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCA
TGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGG
ACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGT
GACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTGC
GCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGG
CCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCA
ACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGATAATACGACTCACTATAGGGTAACTAATTGAGG
CGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAA
GTGACCCTTTAGTGAGGGTTAATT

>pUdO1a-C1B
GATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGT
TTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAA
TACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCT
CGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTC
AAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGC
TTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTTTGAGAAAGCGCCACGCTTCC
CGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGA
GCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG
ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCCAGGAAACA
GCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGG
CTTTTTTTTTCGACCAAAGGGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGC
CCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC
CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT
GCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGA
CCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGT
TTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCC
CCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCGAAGAG
CCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCG
TACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGA
CACGTCGCTCTTCGACGTAAGTAACTGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCC
CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGA
TTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTCGGTACCAAATTCCAGAA
AAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACATT
CAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGT
ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCA
CCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAA
CTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCAC
TTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCG
CATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCA
TGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTG
ACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCT
TGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTA
GCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATT
AATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGT
TTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGAT
GGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAG
ACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTATAGGGTAACT
AATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTC
CTAGGTAAGTGACCCTTTAGTGAGGGTTAATTTAGAAAAGATCAAAGGATCTTCTTGA

>pUdO1a-C2B
CTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACA
TTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTG
GTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACA
GCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGC
TATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAG
AATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGA
GGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGA
ACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACA
ACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT
GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATA
AATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTG
AGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTATAGGGTAACTAATTGAGGCGC
GCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTG
ACCCTTTAGTGAGGGTTAATTTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA
TCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCA
ACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCG
TAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA
GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACC
GAACTGAGATACCTACAGCGTGAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACA
GGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAG
TGACTAGCGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGCGT
TACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAA
TGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGAC
GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATC
TACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCG
GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAA
TCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTAC
GGTGGGAGGTCTATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCG
AGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGG
AAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGActgtgcctt
ctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgag
taggtgtcattctattctgggggggtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggCTCG
GTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACT
GGCCGTCGTTTTACATTCAAATATGTATCCG

>pUdO1a-C3B
CTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACA
TTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTG
GTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACA
GCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGC
TATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAG
AATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGA
GGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGA
ACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACA
ACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT
GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATA
AATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTG
AGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTATAGGGTAACTAATTGAGGCGC
GCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTG
ACCCTTTAGTGAGGGTTAATTTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA
TCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCA
ACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCG
TAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA
GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACC
GAACTGAGATACCTACAGCGTGAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACA
GGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAG
TGACTAGCGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGCCA
AAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTG
TACGGTGGGAGGTCTATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTC
GCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGC
AGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGACT
GTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCA
CTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTC
TGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGG
ATGCGGTGGGCTCTATGGCTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTT
CGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACATTCAAATATGTATCCG

>pUdO1a-C4B
CTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACA
TTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTG
GTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACA
GCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGC
TATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAG
AATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGA
GGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGA
ACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACA
ACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT
GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATA
AATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTG
AGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTATAGGGTAACTAATTGAGGCGC
GCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTG
ACCCTTTAGTGAGGGTTAATTTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA
TCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCA
ACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCG
TAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA
GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACC
GAACTGAGATACCTACAGCGTGAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACA
GGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAG
TGACTAGCGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGAC
GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCGAAGAGCCGAGTGT
CTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCT
ATCGATTTGGATCCACCCTGCAGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCT
CTTCGACGTAAGTAACTGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTT
CCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGA
CAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTCGGTACCAAATTCCAGAAAAGAGGCCTC
CCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACATTCAAATATGTA
TCCG

>pUdO1a-C5B
CTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACA
TTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTG
GTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACA
GCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGC
TATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCT
CAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAG
AATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGA
GGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGA
ACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACA
ACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT
GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATA
AATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCC
CGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTG
AGATAGGTGCCTCACTGATTAAGCATTGGTAATAATACGACTCACTATAGGGTAACTAATTGAGGCGC
GCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTG
ACCCTTTAGTGAGGGTTAATTTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA
TCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCA
ACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCG
TAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA
GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACC
GAACTGAGATACCTACAGCGTGAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACA
GGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAG
TGACTAGCGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGATG
GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTT
CGAGACCGAATTCGCGAGCTCTCGCTAGCCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTT
GGATCCACCCTGCAGGAAGTCTAGACCGTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACG
TAAGTAACTGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC
CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGATGCGGTGGGCTCTATGGCTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAA
GGGGGGCCTTTTTTCGTTTTGGTCCTAAGCTTAACTGGCCGTCGTTTTACATTCAAATATGTATCCG

>pUdO1c-C1B

TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGACATTGATTATTGACTAGTTAT
TAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC
GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACG
GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT
TCCAAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAG
CCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACC
GTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTT
TCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGT
GGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTA
AGCTTAACTGGCCGTCGTTTTACAACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTA
CCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCAC
TGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGC
TCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATA
AGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTTCGTA
TGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAG
CAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATAT
TCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTT
TTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTC
TTCGCCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGAT
TCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTG
CGATGAGTGGCAGGGCGGGGCGTAATAATACGACTCACTATAGGGTAACTAATTGAGGCGCGCCTAC
CACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTATTTTTCAACCTTCCTAGGTAAGTGACCCTT
TAGTGAGGGTTAATT

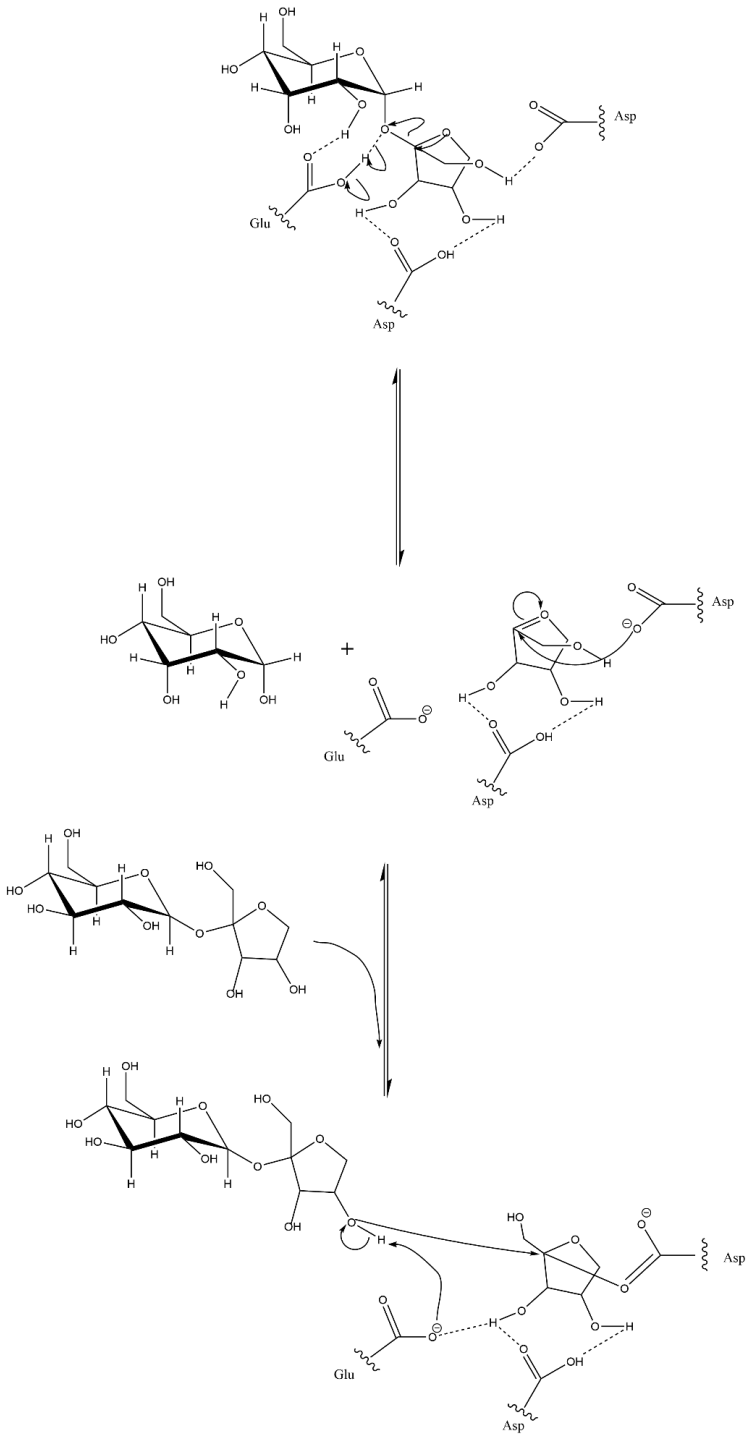
>pUdO1t-C1B
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGACATTGATTATTGACTAGTTAT
TAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC
GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACG
GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT
TCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAG
CCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACC
GTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTT
TCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGT
GGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTA
AGCTTAACTGGCCGTCGTTTTACACTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAAT
ACGACAAGGTGAGGAACTAAACCATGGGTCAAAGTAGCGATGAAGCCAACGCTCCCGTTGCAGGGCA
GTTTGCGCTTCCCCTGAGTGCCACCTTTGGCTTAGGGGATCGCGTACGCAAGAAATCTGGTGCCGCTTG
GCAGGGTCAAGTCGTCGGTTGGTATTGCACAAAACTCACTCCTGAAGGCTATGCGGTCGAGTCCGAAT
CCCACCCAGGCTCAGTGCAAATTTATCCTGTGGCTGCACTTGAACGTGTGGCCTAATAATACGACTCAC
TATAGGGTAACTAATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCGCCGTTTTTA
TTTTTCAACCTTCCTAGGTAAGTGACCCTTTAGTGAGGGTTAATT

>pUdO1z-C1B
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAA
AACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTG
GCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT

AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTTTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
TCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGG
GTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCCAGGAAACAGCTATGACCATGCCCGGGTAAGTGACTAGCGGAAACACAGA
AAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGGGACATTGATTATTGACTAGTTAT
TAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC
GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACG
GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT
TCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTCTCCGAAGAGCCGAGTGTCTTCGGTTCGAGACCGAATTCGCGAGCTCTCGCTAG
CCAGTTTAAACAATTGCGGCCGCGTACTAGTCTATCGATTTGGATCCACCCTGCAGGAAGTCTAGACC
GTCGACACGGTCTCTGAGAGAAGACACGTCGCTCTTCGACGTAAGTAACTGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTT
TCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGT
GGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTC
TATGGCTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTTGGTCCTA
AGCTTAACTGGCCGTCGTTTTACAGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAAT
ACGACAAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGA
CGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACT
TCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGAC
AACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTC
CACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCG
GGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGATAAT
ACGACTCACTATAGGGTAACTAATTGAGGCGCGCCTACCACCGTCAAAAAAAACGGCGCTTTTTAGCG
CCGTTTTTATTTTTCAACCTTCCTAGGTAAGTGACCCTTTAGTGAGGGTTAATT

Appendix II

II.I Levansucrase mechanism



Appendix Figure 2. Proposed Mechanism of Levansucrase Conversion of Sucrose to Levans by Chunsrivirot et al. 3 amino acids (two arginines and one glutamine) play key roles in the catalytic mechanism. In this proposed mechanism, levansucrase binds sucrose and forms a fructosyl-arginine intermediate while releasing glucose. Additional sucrose binds in the active site and hydrolysis occurs again releasing more glucose and expanding a growing chain of levans.