# Development of Novel Metabolite Biosensors


## Alexander Carpenter – 45007675

**Department of Chemistry and Biomolecular Sciences**

**Macquarie University**

**Supervision: Professor Ian Paulsen,**



Submitted: 9/10/2017

Office of the Gene Technology Regulatory (OGTR) Reference number: 5201401059

## Declaration of Authorship and Of Thesis Submission History

This is to certify that I, Alexander Carpenter, being a candidate for the degree of Master of Research, have not submitted this work to any other university or educational institution for the purposes of attaining a higher degree. I also certify that this work is my own, and was not taken, in whole or in part, from un-attributed sources.

.....

Signature of Candidate
Date: 9-10-2017

Elizabeth Daniel
Daniel
.....

Full Name & Signature of Witness
Date: 9/10/2017

# 1: Abstract

Adaptive laboratory evolution (ALE) offers a potentially powerful method of genetic engineering for *S. cerevisiae* cell factories. One of the major limitations in ALE is the limited genetic diversity generation currently capable with traditional mutagenesis methods. The SCRaMbLE technique, outlined as a part of the Synthetic Yeast Genome Project (Sc 2.0) has the capacity to generate significant genetic diversity for use in ALE. Discussed here is the partial construction of synthetic chromosome XIV as a part of Sc 2.0, with successful integration of approximately 80 kilobases of synthetic DNA. Also presented here is the development of two biosensors for use in *S. cerevisiae* ALE experiments for the metabolites butanol and methionine, both of which are industrially significant chemicals. The butanol biosensor was based on the BmoRp transcription factor from *Thauera butanovorans*. Treatment of cells containing this biosensor with 5mM of butanol resulted in a 1.7 fold increase in GFP expression. The methionine biosensor was based on the *S. cerevisiae MET17* promoter and has a dynamic range of 8.6 fold with the addition of 5mM methionine. Thorough characterisation of the components that comprise these biosensors suggest promising modifications that could improve their function and ultimately enable their use for ALE.

# 2: Introduction

## 2.1: Advanced chemical manufacturing in the 21st century

### 2.1.1: Cell Factories

Cell factories are a promising technique utilizing cells for the production of complex chemicals in commercial quantities. This generally involves the optimization of enzymes used in a metabolic pathway, which culminates in the production of a compound of interest [1]. Cells can be grown on a simple carbon source, and through engineered metabolism produce complex and valuable chemical compounds. This could allow inexpensive precursors such as sucrose to be converted into high value compounds, all synthesised in a self-replicating host [1]. The potential applications and implications of genetic engineering and cell factories are promising for the future.

### 2.1.2: Renewable compounds

Cell factories offer a unique ability to convert simple substrates into highly complex materials at commercial levels [1]. This can be used to replace previously unrenewable resources with a more environmentally friendly and renewable version [2]. For example, fuels and plastics are both petrochemical derivatives, and as such are a finite resource. However, terpenoids, alcohols, and alkanes are all promising alternative biofuels which could be produced by cell factories [2]. If these cells are grown on a sustainably sourced feedstock, this could turn liquid fuels and plastics into carbon neutral renewable resources [2].

### 2.1.3: Adaptable manufacturing processes

Cell factories can also offer a financial advantage over traditional chemical synthesis. The infrastructure needed to grow cells is generally less expensive than accommodating an equivalent series of chemical processes [3]. That is, starting from a simple carbon source, metabolism can catalyse the creation of complex chemicals through 10's to 100's of enzymatic steps and can be done primarily in a single bio reactor. To accomplish the same process *ex vivo* would require significantly more time, reaction vessels, and separation techniques etc. [3]. Overall, this results in a lower initial cost to building a cell factory based manufacturing plant, and therefore a smaller barrier for businesses to begin production [3]. Furthermore, as the growth vessel would remain fairly similar from one strain of cell factory

to the next, if a business wanted to change their compound of interest, this transition would involve less physical reconfiguring of a production plant [3]. This would make the process of manufacturing considerably more adaptable to consumer demand/market trends [3].

### 2.1.4: Astronautics

Cell factories have significant applications in the field of astronautics. They can be used for converting inorganic substrates into usable materials on other planets. As described previously, they can be utilized as an adaptable manufacturing processes for chemical compounds. Furthermore, cell factories could act as life support systems recycling carbon dioxide and waste material into oxygen and to produce critical medicines [4]. Additionally, cell factories could eventually be used for terraforming of potential habitats [4]. Finally, a significant concern in astronautics is reducing the amount of weight/space required for each tool included on a mission. A major advantage of cell factories in this regard is that they can be frozen as a small stock, and then self-replicate to usable quantities when needed [4].

### 2.1.5: *Saccharomyces cerevisiae* as a cell factory

There are numerous species currently under investigation as potential industrial cell factories. For example *Escherichia coli, Bacillus subtilis,* and *Corynebacterium glutamicum* [1]. However, one of the most promising species for a general cell factory platform is *Saccharomyces cerevisiae*. This yeast has "generally regarded as safe" status, can utilize a diverse range of carbon sources, can grow in a wide range of pH, and can grow both anaerobically and aerobically [1]. Additionally, as a model organism, a wealth of genetic, transcriptomic, and metabolic data already exists to aid in engineering. Furthermore, its historical use in the production of alcohol means there is a large wealth of knowledge in scaling up/use at industrial volumes [1]. It has been shown that *S. cerevisiae* is capable of producing a wide range of valuable chemicals. For example malate [5], lactate [6], and naringenin [7] to name a few. Most impressively has been the production of the anti-malarial drug Artemisinin at commercially viable levels [8]. These advantageous characteristics and examples of previous engineering have demonstrated that *S. cerevisiae* is a viable platform for cell factories and genetic engineering.

## 2.2: Traditional genetic engineering

### 2.2.1: Rationale of traditional genetic engineering

Traditional genetic engineering is the dominant method for creating cell factories which can produce commercial volumes of a desired compound. It involves the rational genetic manipulation of metabolic pathways to increase flux towards a compound of interest [9, 10]. The process has two major steps. The first is the identification/introduction of the production pathway. This involves either identifying which enzymatic processes are present/needed which would form the compound of interest [9, 10]. If a suitable enzymatic pathway is not natively present then heterologous expression of pathway genes from another suitable species is necessary. The second step is the optimization of gene expression levels to maximise production pathway flux. This involves overexpressing enzymes involved in the synthesis of the compound of interest, and the down regulation of genes involved in the consumption of the compound or its precursors [9, 10]. This step also requires the balancing of regulatory mechanisms to avoid negative feedback into the production pathway.

These processes are repeated in a cyclical process of “design, build, test, learn” in which each iteration of a genetic design is synthesized, its productivity assayed, and then outcomes used to inform a new round of metabolic designs [9]. This strategy has been employed successfully for a range of compounds which have achieved commercial viability. For example lysine [11, 12], 1,3-Propanediol [13], 7-ADCA [14], 1,4-Butanediol [15], Artemisinic acid[16], and Isobutanol [9, 17-19].

### 2.2.2: Limitations

As successful as traditional genetic engineering has been, there are several significant limitations to the technique. Traditional genetic engineering requires a significant knowledge of the metabolic pathways to be exploited [9]. That is, all enzymatic steps involved, and the role of each intermediate compound in pathway regulation are both critically important pieces of information needed to make traditional engineering work [9]. Furthermore, unexpected inhibition, or unexpected flux into or out of the production pathway from peripheral metabolism will cause havoc in a design [20]. Additionally, to achieve full production potential requires altering expression of enzymes peripheral to the core metabolic pathway involved in target compound creation [20]. That is, non-obvious

upstream, downstream, and peripheral metabolic processes impact on target compound production in ways that are not always obvious for rational engineering. These limitations mean that the traditional engineering of a strain from conception to commercial production can take 6-8 years of research and cost approximately $50 million [9].

### 2.2.3: Adaptive laboratory evolution

One alternative method for metabolic engineering which has been gaining interest in recent years is adaptive laboratory evolution (ALE). ALE involves the passaging of a microbial community whilst selecting for production of a compound of interest [20]. Mutations which are congruent with the artificial selection pressure applied will infrequently occur within the population. Cells which possess these mutations have a selective advantage and become enriched in the population over time [20]. This allows for a non-biased screening of genetic variants which increase the production of a target compound [21]. The main advantages of this technique over traditional engineering is that ALE can reveal non-obvious beneficial genetic changes, can take less time/money, and require less understanding of the production pathway of interest [20].

ALE was originally used to adapt cells to alternative growth medium. Examples of this include passaging cells repeatedly on minimal growth medium until increased growth rate was achieved [22], or repeated passaging to establish a glycine prototrophy [23]. One of the best examples demonstrating the power of ALE was a study in which an *E. coli* strain was rationally engineered to over-produce lactic acid at a concentration of 1.25 g/L. After engineering, the strain was subjected to ALE selecting for fast growing sub-populations. Since growth rate and lactic acid production are intrinsically linked metabolically, this resulted in a 35% increase in lactic acid production above what could be rationally designed [24].

Until quite recently, ALE experiments have been limited to phenotypes that are directly related to cell survival. For example, selecting for the fastest growing cells, selecting for ability to grow on a specific media, selecting for resistance to a target compound. In those instances ALE is possible because the desired phenotype is innately linked to a selective advantage. However, for ALE to become a platform technology for metabolic engineering it must be able to select for compounds that imbue no inherent selective advantage.

### 2.2.4: Biosensors

Biosensors are a new strategy currently being investigated which would allow the production of any compound of interest to result in a selective advantage. In their simplest form, biosensors are tools used to detect a compound of interest and then produce a signal [25]. Their mechanisms of actions are somewhat diverse but can largely be delineated into two broad classes (Figure 1). The first are biosensors that detect a target compound and give a report signal directly (e.g. a protein with a compound binding domain and an inducible fluorophore domain); the second are biosensors which detect a target compound and then induce a secondary signal (e.g. a transcription factor or g-protein coupled receptor which binds the compound of interest and then induces transcription of a reporter gene) [25]. A classic example of this second design is the LysG based lysine biosensor [26]. This biosensor uses the lysine activated transcriptional regulator (LysGp) to induce transcription of an enhanced yellow fluorescent protein (*eYFP*) gene [26]. When expressed in a population of *E. coli* cells, this resulted in lysine concentration dependent production of *eYFP*.

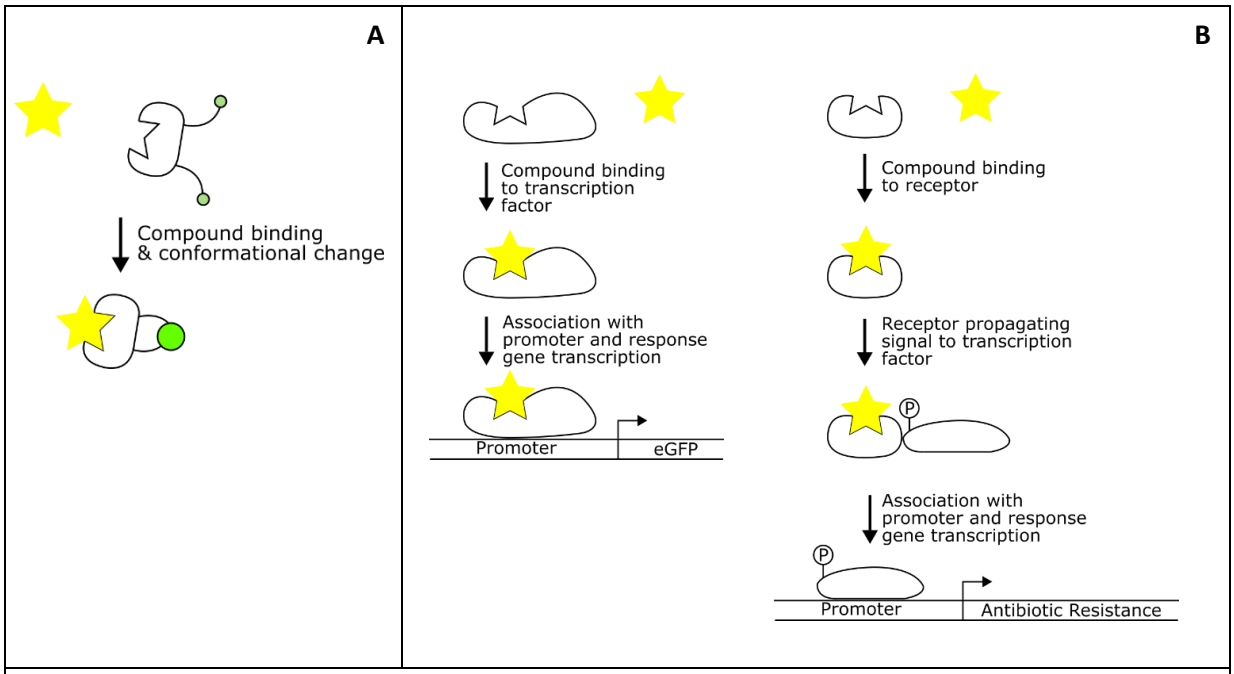


Figure 1: Stylistic representation of class 1 and class 2 biosensors. **A:** depicts a protein based biosensor with two tethered fluorescent protein response elements. The binding of the target compound (star) induces a conformation change bringing the two fluorophores in close proximity, changing the fluorescence profile. **B:** Depicts two examples of biosensors. On the left is a transcription factor that is activated by the target compound and induces transcription of an enhanced green fluorescent protein. On the right is a target compound receptor which binds to the target compound and propagates its signal via phosphorylation of a transcription factor. This transcription factor then activates the transcription of an antibiotic resistance gene.

Biosensors allow the cell specific monitoring of the concentration of a target compound. When employed across a population this reveals a concentration distribution, allowing the identification of highly productive cells [25]. These cells can then be sorted from the population and a new higher producing population is established. This process can be iterated, with subsequent populations producing increased concentrations of a target compound (Figure 2) [25]. The implementation of biosensors for use in ALE has been somewhat stilted. Whilst some valuable compounds have readily applicable receptors/transcription regulators, most do not, and thus the engineering of novel sensory machinery is necessary. Additionally, the difficulties in transferring transcriptional regulators between domains of life is also hindering the creation of biosensors. Thus, new methods for biosensor creation are necessary.

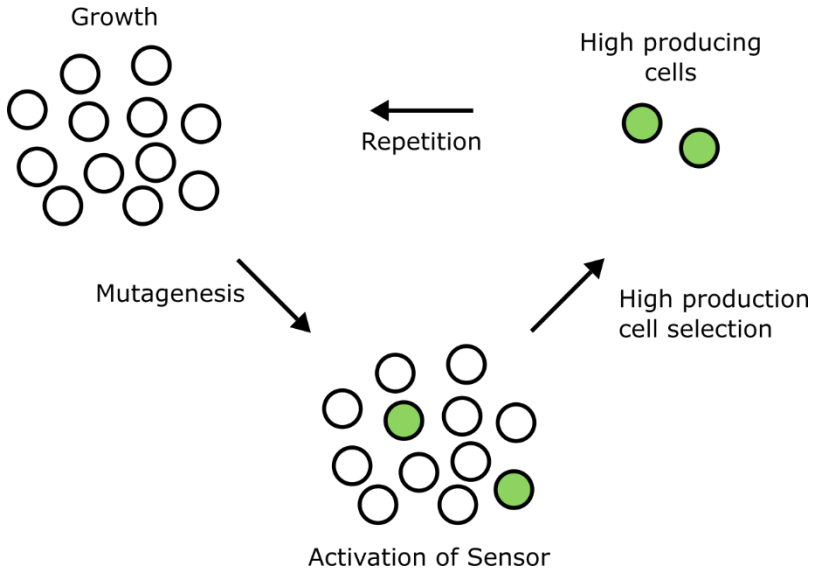


Figure 2: Representation of the process of ALE using a biosensor mediated approach. With each iteration the highest producing cells are selected and used to form a new population. This enriches for increased production phenotypes over time.

### 2.2.5: Cheating

One limitation to ALE as a technique for metabolic engineering is the issue of cheaters. Cheaters are cells in ALE experiments which obtain mutations which allow them to trigger a sensor, without producing more of the target compound [25]. That is a mutation that results in constitutive activation of the signal output. For example, in the above mentioned LysG based lysine biosensor, lysine induces the LysGp transcriptional regulator to drive

expression of an e*YFP* gene under the control of the *LYSE* promoter [26]. A mutation which constitutively activates the *LYSE* promoter would produce more *eYFP* but would not have an increase in lysine production. Cheaters like this pose serious problems to ALE experiments if they run for an extended period. Cheaters have all the selective advantage of overproducers but with none of the metabolic costs of compound production, resulting in enrichment of false positives in the population.

Several attempts have been made in developing methods to mitigate cheater propagation in ALE populations. Most cheater mitigation techniques are specific to the biosensor being assayed, exploiting some artefact of the regulatory pathway/reactive intermediate. However, a general method for cheater mitigation has been developed in *E.coli* based on the *TOLC* gene [27]. Expression of the TolC protein imbues resistance to SDS but a susceptibility to colicin E1. The generic cheater mitigation technique works by placing the *TOLC* gene under the control of a sensing transcription factor. Under positive selection the cells are grown in SDS and high producers activate the transcription factor/the *TOLC* gene, imparting selective advantage [27]. Under negative selection, one of the genes involved in the compounds production is down-regulated so that no cell should be producing the compound of interest. Then they are treated with colicin E1. Only cells which are constitutively over-expressing the *TOLC* gene will be susceptible, and be removed from the population.

As effective as the TolC based cheater mitigation method is, it still has several limitations. The first issue is that this methodology doesn't remove all cheaters in the population, just kills a percentage of them with colicin E1. Whilst this is acceptable for short runs of ALE, over longer timeframes this will allow a small subpopulation of cheaters to persist. This could pose substantial problems if any of these cheaters obtain the ability to resist colicin E1. More importantly, this methodology also selects for any strain which develops highly sensitive (but not constitutively active) sensor activation. This could be alterations in the transcription factor binding affinity for the compound of interest, transcription factor affinity for the promoter binding sequence and/or mutations in the promoter sequence making it more sensitive to the transcription factor. This type of mutation would generate a cheater which is in essence over-reacting to the presence of the target compound. The TolC based cheater mitigation technique also requires placing a key enzyme involved in target

compound synthesis under the control of an inducible promoter. This means that ALE experiments could not affect the expression levels of whichever enzyme is selected. Finally, this methodology has only been shown to work in *E. coli* and it is unclear if a similar system is possible in *S. cerevisiae*.

### 2.2.6: Generation of diversity

An important consideration in ALE experiments is generating genetic diversity. For ALE to work efficiently, diverse libraries of genetic variants must be generated in a timely and cost efficient manner. Classic methods of generating diversity are ultraviolet (UV) light and ethyl methanesulfonate (EMS). These techniques have been used for decades to induce point mutations randomly in the genome.

However, point mutations alone are insufficient to generate the wide range of genetic changes represented in natural evolution, and are insufficient to realise the full potential of ALE. A range of alternative methods for generating genetic diversity have been investigated [21]. Unfortunately, the majority of these techniques cause genetic changes on a scale too small for high throughput ALE. For example techniques like DNA shuffling [28], random chimeragenesis on transient templates (RACHITT) [29], and sequence homology independent protein recombination (SHIPREC) have their origins in protein engineering and as such are only applicable for generating diversity within a handful of selected proteins at a time [21, 30]. Techniques that can re-arrange multiple genes in a non-biased, high throughput fashion are critical for ALE experiments.

A promising new genetic diversification technique which could be implemented in ALE experiments is synthetic chromosome rearrangement and modification by *loxP*-mediated evolution (SCRaMbLE) [31]. This technique has been developed as part of the Synthetic Yeast Genome Project (Sc 2.0), which has the goal to generate a yeast strain with an entirely synthetic genome [32]. One design feature of this synthetic genome involves the flanking of every non-essential open reading frame with l*oxP* recombination sites [32]. SCRaMbLE works via inducing l*oxP* based recombination through expression of the Cre recombinase [32]. The open reading frames flanked by *loxP* sites are then capable of undergoing duplication, deletion, or inversions [31]. The advantage of this technique over other genetic diversification techniques is that its functional unit is one open reading frame. That is, whole genes can increase/decrease in expression level without a high probability of inducing non-

sense/mis-sense mutations, or catastrophic chromosomal rearrangements. Furthermore, as every non-essential gene is flanked by *loxP* sites, this allows for a high throughput, fairly non-biased method of generating diversity. The Sc 2.0 project is a major collaboration between many laboratories around the world and requires vast amounts of synthetic genome construction. It is imperative that this construction is completed for the future success of ALE as a metabolic engineering technique.

## 2.3: Methionine as a target for metabolic engineering

### 2.3.1: Methionine use in the poultry industry

Methionine is an essential amino acid in poultry diets. Poultry lack the ability to synthesise methionine *de novo* and thus it is required in nutritional supplementation [33]. There are roughly 44.1 billion chickens produced by the poultry industry worldwide each year [34, 35]. Each of these chickens requires 30.38 g of methionine and cysteine for optimal growth within the first 40 days of their lives [36]. This makes the total global market for both amino acids 1.3 million tonnes per year. However, methionine is also the limiting amino acid in almost all poultry feed. That is, poultry must be over-fed to achieve the correct concentration of methionine in the diet, wasting feed/money [33]. Furthermore, if methionine is supplemented at levels above those minimally required for growth, it has other beneficial effects. Increased methionine concentrations in feed stocks improves the activity of the poultry immune responses [37]. This allows fowl to have an increased quality of life and reduce the chance of infection. This reduced infection rate would lower the need to use blanket antibiotic treatment in the poultry industry, decreasing its contribution to the emergence of multidrug resistant pathogens [37].

One solution the industry has been using to deal with methionine shortages is chemically derived synthetic methionine. Unfortunately, synthetic methionine exists in a racemic composition of both D- and L- enantiomers [37]. Whilst L-methionine are produced in nature and is readily metabolised by poultry, D-methionine is metabolised into toxic by-products [37]. This has caused concern for both poultry and human health, and its use in the industry is being questioned [37]. Avoiding the use of D-methionine has sparked interest in the production of methionine from natural sources. This is because the creation of enantiomerically pure methionine is very difficult using chemical processes, but is

significantly easier using enzymatic pathways. One way in which this could be done is through using *S. cerevisiae* as a cell factory for the production of methionine.

### 2.3.2: Methionine production and gene regulation in *S. cerevisiae*

Methionine is produced and regulated as a part of the sulfur metabolic network. This network is the pathway that takes exogenous sulfate, and through a series of metabolic steps, enters the methyl cycle and the reversible transsulfuration pathway (Figure 3) [38]. Extracellular sulfate is taken up by the cell using the high affinity sulfate transporters Sul1p and Sul2p [39]. From here Met3p, Met14p, Met16p, Met5p, Met10p, and Met17p catalyse the formation of the sulfur containing compound homocysteine (Figure 3)[40]**.** From here the pathway branches into two directions. The first is the methyl cycle which produces methione, S-adenosyl methione, and S-adenosyl homocysteine via the enzymes Met6p, Sam1p and Sam2p, and Sah1p respectivley. The second is the transsulfuration pathway which catalyzes the formation of cystathioine and cysteine via Cys4p and Cys3p respectively, and which also reverses this process through the enzymes Str2p and Str3p.

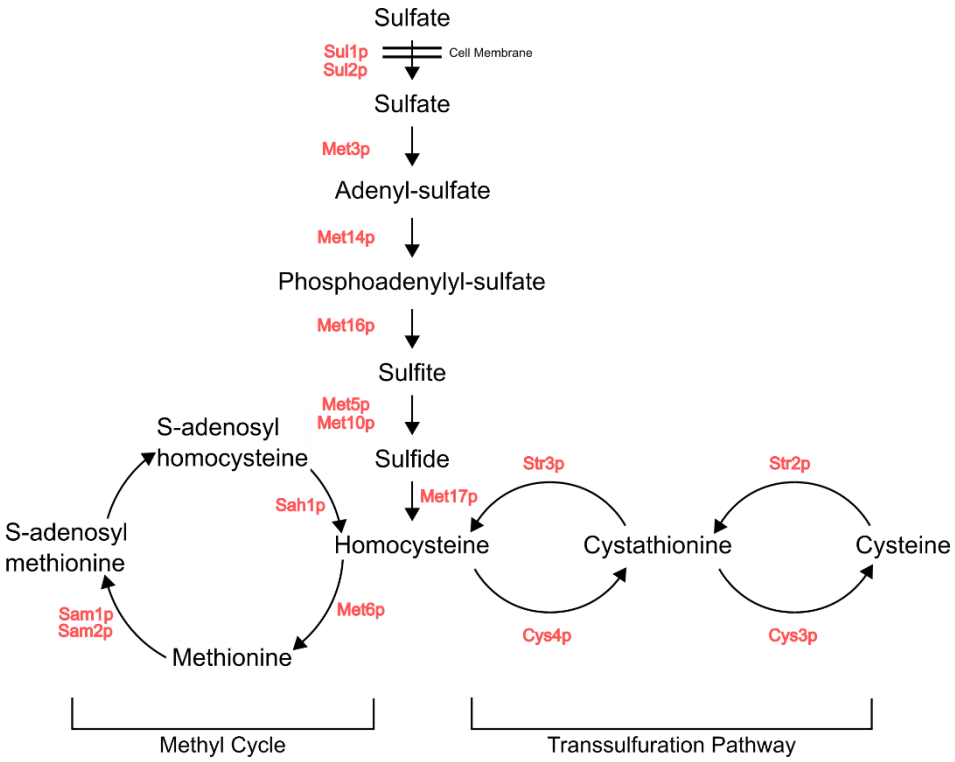


Figure 3: Diagram of the sulfur network, including the methyl cycle and transsulfuration pathway. Red text denotes catalytic enzymes involved in each step of the network. Figure shows the conversion of sulfate into the branch point compound homocysteine before entering into the interconnected methyl cycle and transsulfuration pathway. Not shown are associated co-factors. Figure was adapted from Ljungdahl PO and Daignan-Fornier B [40]

Both the methyl cycle and the transsulfuration pathway use homocysteine as starting substrates to produce methionine and cysteine respectively [40]. Due to the cyclical nature of the methyl cycle, and the reversible nature of the transsulfuration pathway, and their dependence on the metabolite homocysteine, the pools of methionine and cysteine are intrinsically linked [38]. That is, if the concentration of methionine increases, so too does the concentration of cysteine and *vice versa* (figure 3).

Most of the network is regulated by transcriptional activator Met4p. Met4p interacts with Met31p/Met32p/Cbf1p which allows the recruitment of transcriptional machinery and induction of transcription [41]. There are 45 known Met4p regulated promoters, and each one contains binding sites for Met31p or Met32p [41]. When cellular levels of cysteine are diminished these work together to activate the 45 known Met genes, including the well characterized *MET17* [42].This allows *de novo* synthesis of methionine and cysteine. However, this process is halted in the presence of high cysteine concentration through the activity of the ubiquitin ligase complex $SCF^{Met30}$ [40, 42].

### 2.3.3: Building a methionine biosensor for use in ALE

The sulfur network can be exploited to develop a basic sensor for the end products of the sulfur network, including methionine. By putting a responder under the control of the *MET17* promoter, a signal could be generated based on the activation of the sulfur network. Under conditions of high cysteine, the sensor would be turned off, and under conditions of low cysteine the sensor would be turned on. As the concentrations of the end products of the sulfur network exist in an interconnect pool, any genetic changes that increase flux into the sulfur network would affect a p*MET17* controlled sensor. That is, any increase in the production of S-adenosyl methionine, S-adenosyl homocysteine, homocysteine, cystathionine, cysteine, or methionine would activate the sensor by increasing the concentration of all other metabolites in the network.

## 2.4: Butanol as a target for metabolic engineering

### 2.4.1: Butanol as a biofuel

Climate change has increasingly been at the forefront of innovation in recent years as scientists and engineers attempt to find ways to halt or mitigate its effects. It poses a serious threat to life and has the potential to disrupt much of human activity [43]. Climate change is being caused by unfettered release of greenhouse gasses into the atmosphere [43]. The majority of which come from the burning of fossil fuels [43]. Additionally, even without the threat of climate change, world reserves of oil and gas are tipped to deplete as early as 2042 [44]. To help avoid the consequences of climate change, obtain energy independence, and to prepare for depletion, a substitute for fossil fuel derived liquid fuels is necessary.

Butanol is a four carbon alcohol that could feasibly substitute for liquid fuels. When compared to both gasoline and diesel, butanol shares very similar chemical properties when used in internal combustion engines [45, 46]. The similarities are so profound that butanol can be used at 100% concentration in a conventional internal combustion engine [47]. Furthermore, when compared to ethanol, another potential biofuel, butanol is more energy dense due to its additional two carbon atoms [47].This potential has been recognised by industry with two companies, Gevo and Butamax, beginning commercial synthesis of isobutanol [9]. If butanol could be produced cheaply and easily in *S. cerevisiae* from renewable carbon sources, mankind would no longer be dependent on fossil fuels for

transportation. Furthermore, assuming the carbon sources used are sourced in a carbon neutral manner, then the process would not have a net release of carbon into the atmosphere.

### 2.4.2: Engineering butanol production

The metabolic engineering of *S. cerevisiae* to produce butanol has had moderate success in the past 5 years. Initial attempts utilized a series of genes from *Clostridia* species known to be responsible for butanol production [48]. These enzymes were expressed in *S. cerevisiae* strains and utilizied acetyl-CoA as a precursor. This was able to generate cultures producing 16.3 mg/L of butanol [48]. After engineering increased flux to acetyl-CoA, this strategy resulted in strains producing butanol at concentrations of 130 mg/L [49]. A significant improvement came when it was discovered that 1-Butanol is produced endogenously by *S. cerevisiae* during the degradation of threonine [50]. Under normal circumstances produced 1-butanol is degraded by alcohol dehydrogenase 1. However, when this gene was deleted 1-butanol could accumulate in the growth media up to 120 mg/L. Furthermore, when the enzymes catalysing the butanol production pathway were over-expressed this accumulation increased to 242.8 mg/L. Finally, these two pathways were combined and further optimised, leading to a production of 835 mg/L [51]. Whilst this is a significant improvement, this titre is still far from a being a viable alternative to fossil fuels.

### 2.4.3: Creating a butanol biosensor for use in ALE

BmoRp is a transcriptional regulator from *Thauera butanovorans* which controls the expression of an alkane monooxygenase (BMO)[52]. BmoRp induces transcriptional response via binding to the *BMO* promoter sequence in response to short chain alcohols. In its native context, BmoRp responds to $C_2$-$C_8$ alcohols to allow *T. butanovorans* to utilize these compounds as carbon sources [52]. Transcriptional activity via BmoRp, and p*BMO* is most pronounced when activated by butanol [52].

The preferential butanol-mediated activation of the *T. butanovorans BMO* promoter via BmoRp has been exploited previously to generate a butanol biosensor in *E. coli* (figure 4) [53]. This sensor put a tetracycline resistance and *eGFP* genes under the control of the *BMO* promoter [53]. Activation of the sensor via increased butanol concentrations induced resistance to tetracycline and production of *eGFP*. In this system, the BmoRp transcriptional activator was also expressed under the *BMO* promoter, creating an amplifying signal [53].

Using a combination of tetracyline treatment and FACS, higher butanol producing cells could be selected from the population. This biosensor construction was used successfully for an ALE experiment increasing 1-butanol production from 345 mg/L to 493 mg/L [53].

It may be possible to create a butanol biosensor for use in *S. cerevisiae* based on the BmoRp transcription factor. As the previously developed sensor was for a prokaryotic host, utilizing the BmoRp transcription factor in *S. cerevisiae* would require significant sensor modification.

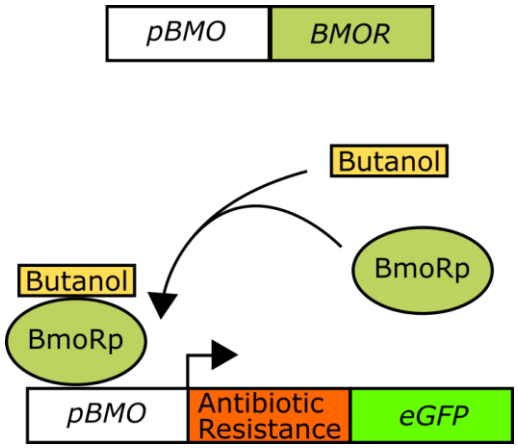


Figure 4: Simplistic representation of a butanol biosensor which has been developed for use in *E. coli*. Top and bottom bars represent DNA sequences at different sites on the same plasmid, whilst the BmoR protein is shown as an oval interacting with butanol to induce transcription. Bars are not to scale.

## 3: Aims

The aims of this project were

1. Aid in the progress of the Sc 2.0 project by contributing to the construction of synthetic chromosome 14.
2. Design, construct, and characterize a butanol biosensor for use in *S. cerevisiae* ALE experiments.
3. Design, construct, and characterize a methionine biosensor for use in *S. cerevisiae* ALE experiments.

# 4: Methods

## 4.1: *S. cerevisiae* and *E.coli* culture media

The *S. cerevisiae* strain in which all biosensors were constructed and tested was BY4742 [54]. Unless otherwise stated, *Saccharomyces cerevisiae* strains were grown at 30$^{o}$C in 6.7 g/L yeast nitrogen base without amino acids (Sigma-Aldrich) supplemented with 10g/L leucine, 10g/L lysine, 5g/L histidine and 1% glucose (Sd media).

The *E.coli* strain DH5α was used as an intermediary in biosensor construction. Unless otherwise stated *Escherichia coli* strains were grown at 37$^{o}$C in lysogeny broth supplemented with 0.1 mg/mL ampicillin.

## 4.2: DNA Purification

Plasmid DNA was purified using the Monarch Plasmid Miniprep Kit supplied by New England BioLabs. PCR products were purified using either the Zymoclean Gel DNA Recovery Kit supplied by Zymo Research or with the QIAquick PCR purification kit supplied by Qiagen. Primers were designed using the software Geneious Pro V10.2.2 [55]. A complete list of primers used is available in Supplementary Table 1. Primers were supplied by Integrated DNA Technologies.

## 4.3: Megachunk assembly

The methodology for construction of synthetic chromosomes for the Sc 2.0 project was outlined in Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CLV, Chandrasegaran S, Cai Y, Boeke JD and Bader JS [56]. The scope of work carried out in this project covered the integration of megachunks U, V, and W which comprise 98 kilobases of synthetic DNA, into an existing partially synthetic chromosome 14 containing strain (Sc 2.0 Syn 14 G-T). Megachunk U was comprised of 5 chunks averaging approximately 8 kilobases in size, V was comprised of 4 chunks averaging approximately 7.3 kilobases in size, W was comprised of 4 chunks averaging approximately 8 kilobases in size.

For each megachunk, its component chunks were PCR amplified and transformed simultaneously into a strain with the previous iteration of construction, starting with strain Sc 2.0 Syn 14 G-T (Figure 5). The chunks comprising each megachunk have 40 base pairs of

overlap between each other allowing yeast assembly to be used to connect each “chunk” into a “megachunk using yeast homologous recombination [57].

Additionally, each megachunk had approximately 500 base pairs of homology to the previous megachunk at the 5’ end and to the native yeast genome at the 3’ end. This allows chromosomal integration of the megachunk via homologous recombination, replacing the native genomic sequence at that site.

Towards the 3’ end of each megachunk either a *LEU2* or *URA3* gene was also encoded. As the destination strain is auxotrophic for both amino acids this allowed for the complementation of 1 auxotrophy per megachunk integration. Each megachunk was designed so that upon recombination, it would remove the previous auxotrophic rescue gene. This allowed 2 auxotrophic markers to be used in an alternating pattern during rounds of recombination. This resulted in a sequential construction of strains which contained megachunks G-U, G-V, and finally G-W (Figure 5).

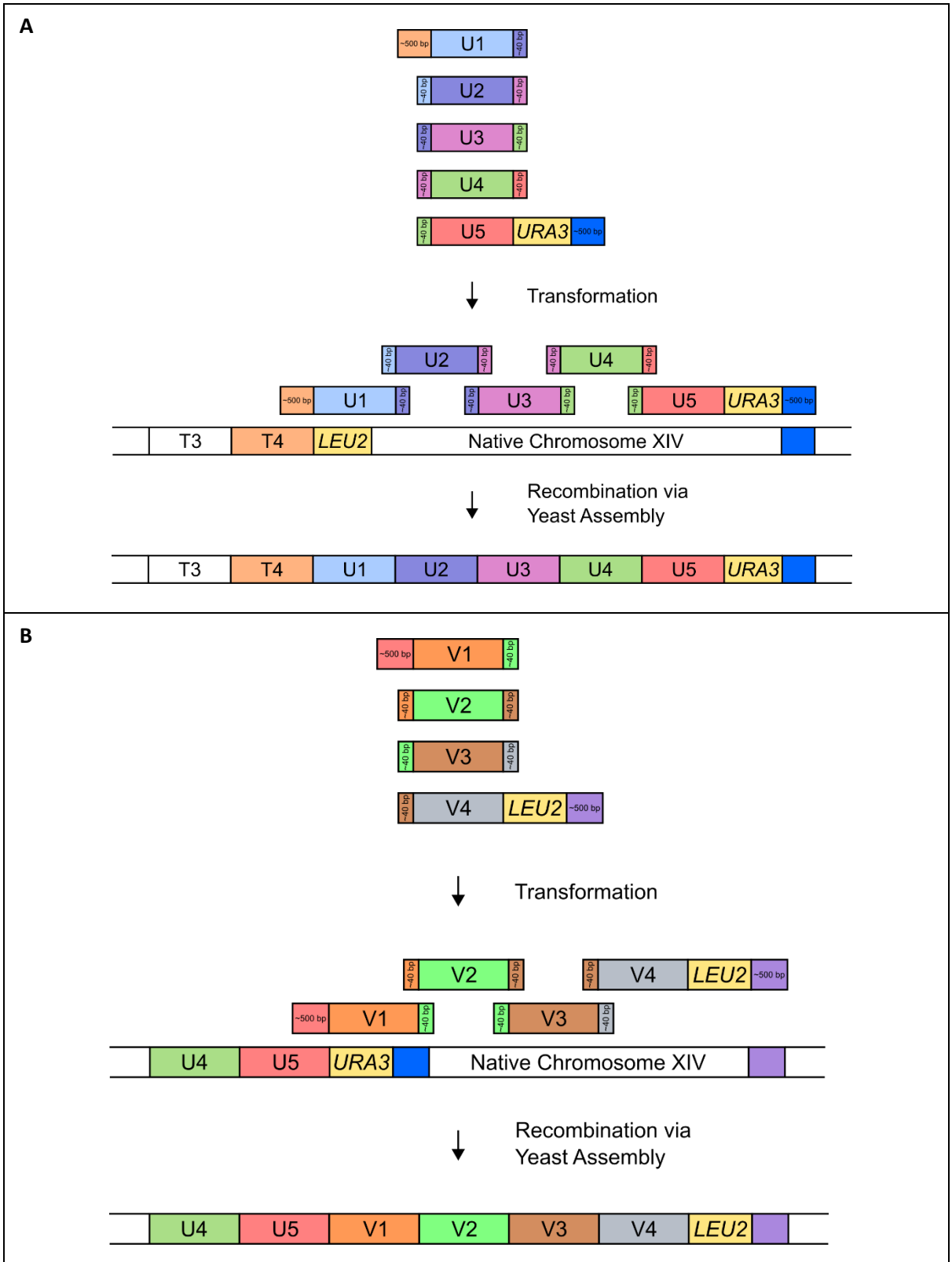


Figure 5: Simplified representation of the iterative construction process for synthetic chromosome 14. Bars are representative of DNA sequences. **A** shows chunks U1-U5 amplified by PCR, transformation into the strain Sc 2.0 Syn 14 G-T, and recombination with the native chromosome. **B** shows chunks V1-V4 amplified by PCR, transformation into strain Sc 2.0 Syn 14 G-U, and recombination with the native chromosome. Bars are not to scale.

## 4.4: Megachunk assessment

Screening of colonies for correct megachunk integration and native locus replacement was performed with qPCR using primer tags as were previously designed using the BioStudies software package [56]. Two sets of PCR tags were used in assessment of each megachunk. One set was designed to amplify only when a megachunk had been incorporated, whilst the second set was designed to only amplify the native sequence. Strains which passed qPCR assessment were then whole genome re-sequenced by Macrogen using True-Seq Nano library preparation with 470 bp inserts, and paired-end Illumina HiSeq 2500 sequencing. Read data was processed using the software Geneious v10.2.2. Paired-end reads were assembled to an edited version of the S288C reference genome where native chromosome 14 was replaced with synthetic chromosome 14 [58]. This was done with the "Map read to reference" function with the "Geneious" mapper, and "highest sensitivity" [55]. Analysis of resultant assembly was completed manually by assessing read coverage, and read disagreement with the reference sequence. The raw reads were of high-quality (Q30 =91%, Q20 = 95%), and were therefore not trimmed prior to assembly.

## 4.5: Biosensor design

Biosensors were designed under a schema of modular parts. Frequently used components include the *TEF1* promoter, *ADH1* and *CYC1* terminators, enhanced green fluorescent protein (eGFP), a red fluorescent protein (mCherry), a *S. cerevisiae* synthetic core promoter (SCP), SL7 linker, Vp16 activation domain, TetR protein, TetR binding site (*tetO*), BmoR protein, and BmoR binding site (p*BMO*). A summary of the parts used in this project, and their individual functions are shown in Table 1.

| Table 1: Overview of basic biosensor components | | |
|---|---|---|
| Component | Function of Component | Source |
| *TEF1* promoter | Constitutively active *S. cerevisiae* promoter [59] | Amplified from S288C gDNA |
| *ADH1* terminator | Commonly used *S. cerevisiae* transcriptional terminator [60] | Amplified from S288C gDNA |
| *CYC1* terminator | Commonly used *S. cerevisiae* transcriptional terminator[60] | Amplified from S288C gDNA |
| Enhanced green fluorescent protein (*eGFP*) | A commonly used green fluorescent protein [61] | Amplified from custom synthesis via IDT |
| mCherry | A commonly used red fluorescent protein [61] | Amplified from custom synthesis via IDT |
| *S. cerevisiae* synthetic core promoter (SCP) | A bioinformatically designed core promoter based on similarities identified between *S. cerevisiae* promoters, experimentally validated. [62] | Amplified from custom synthesis via IDT |
| SV40 nuclear localization sequence (*SV40 NLS*) | A peptide sequence which is used as a signal for active transport into the nucleus. [63] | Amplified from custom synthesis via IDT |
| *SL7* linker | A flexible linking domain used previously in protein engineering as a tether [64] | Amplified from custom synthesis via IDT |
| *VP16* activation domain | An activation domain sourced from the herpes simplex virus. It has been used previously as a generic activator of transcription. [65] | Amplified from custom synthesis via IDT |
| *TETR* | A transcriptional repressor sourced from *E. coli* codon optimised for expression in *S. cerevisiae* [66] | Amplified from custom synthesis via IDT |
| *tetO* | DNA sequence that is bound by TetR protein. [66] | Amplified from custom synthesis via IDT |
| *BmoR* | A transcription factor from *T. butanovorans* codon optimised for expression in *S. cerevsiae*. [52, 53] | Amplified from custom synthesis via IDT |
| p*BMO* | DNA sequence that is bound by BmoR in response to butanol.[52, 53] | Amplified from custom synthesis via IDT |
| Full sequences available in Supplementary Table 2 | | |

## 4.6: Biosensor creation

All biosensors in this project were constructed on the PRS416 plasmid backbone, which contains the uracil auxotroph complementing gene *URA3*. All biosensor plasmids were transformed into *S. cerevisiae* strain BY4742 genotype MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0. Biosensors were constructed primarily using Gibson Assembly [57]. PCR was used to amplify individual components to create homology to each other or to the destination plasmid necessary for Gibson Assembly. PRS416 plasmid was digested with either *Sma*I, or *Eco*53KI restriction enzymes to enable linearization prior to assembly. Gibson Assembly mix "NEBuilder HiFi DNA Assembly Cloning Kit" was used for the construction of all biosensors, before transformation into *E. coli* DH5α cells. A combination of colony PCR and restriction digestion was used to ensure the correct assembly and transformation of each biosensor gene construct. Plasmid DNA was extracted from each *E. coli* DH5α strain using the "Monarch Plasmid Miniprep Kit" before transformation into *S. cerevisiae* strain BY4742. Growth on Sd media without supplemented uracil was used to select for *S. cerevisiae* cells which had successfully been transformed by a version of the PRS416 plasmid. A complete list of biosensor gene constructs, individual components used in each sensor, and the primers used for construction are shown in Table 2. Primer sequences used for amplification are shown in Supplementary Table 1.

| Table 2: Overview of biosensors constructed | | |
|---|---|---|
| Name | Components Used | Primers Used |
| SCP-*eGFP* | p*TEF1*, *SV40 NLS*, *TETR*, *SL7* linker, *VP16* activation domain, *ADH1* terminator (Inserted at the *Sma*I site of PRS416)<br><br>*tetO*x7, SCP, *eGFP*, *CYC1* terminator (Inserted at the *Eco*53KI site of PRS416) | 9, 100, 101, 102, 103, 95, 96, 16,<br><br>104, 105, 106, 43, 44, 20 |
| *BMOR-VP16* | p*TEF1*, *SV40 NLS*, *BMOR*, *SL7* linker, *VP16* activation domain, *ADH1* terminator (Inserted at the *Sma*I site of PRS416)<br><br>pBMO, SCP, *eGFP*, *CYC1* terminator (Inserted at the *Eco*53KI site of PRS416) | 9, 91, 92, 93, 94, 95, 96, 16<br><br>17, 43 44,20 |
| *VP16-BMOR* | p*TEF1*, *VP16* activation domain, *SL7* linker, *BMOR*, *SV40 NLS*, *ADH1* terminator (Inserted at the *Sma*I site of PRS416)<br><br>pBMO, SCP, *eGFP*, *CYC1* terminator (Inserted at the *Eco*53KI site of PRS416) | 9, 10, 11, 12, 13, 14, 15, 16,<br><br>17, 43, 44, 20 |
| p*MET17-TETR-tetO* | p*MET17*, *TETR*, *SV40 NLS*, *ADH1* terminator (Inserted at the *Sma*I site of PRS416)<br><br>p*TEF1*, *tetO*, *eGFP*, *CYC1* terminator (Inserted at the *Eco*53KI site of PRS416) | 21, 25, 26, 27, 15, 16,<br><br>28, 29, 30, 47, 48, 20 |
| p*MET17-TETR-tetO*x7 | p*MET17*, *TETR*, *SV40 NLS*, *ADH1* terminator (Inserted at the *Sma*I site of PRS416)<br><br>p*TEF1*, *tetO*x7. *eGFP*, *CYC1* terminator (Inserted at the *Eco*53KI site of PRS416) | 21, 25, 26, 27, 15, 16,<br><br>28, 33, 34, 49, 50, 20 |
| p*MET17-eGFP* | p*MET17*, *eGFP*, *CYC1* terminator (Inserted at the *Sma*I site of PRS416)<br><br>p*TEF1*, mCherry, *ADH1* terminator (Inserted at the *Eco*53KI site of PRS416) | 21, 45, 46, 24,<br><br>78, 79, 80, 81, 82, 83, |

## 4.7: Biosensor Assessment

The maximum and minimum signal output (dynamic range), and maximum and minimum activating metabolite concentration (linear range) of each biosensors was assessed using eGFP fluorescence measurements via flow cytometry. Unless otherwise stated *S. cerevisiae* strains were pre-cultured in 5mL of Sd media in 15mL falcon tubes at 30$^{o}$C with shaking at 200 rpm overnight before inoculation. Strains were inoculated in triplicate in 10 mL Sd media with or without biosensor inducing compounds at an $OD_{600}$ of 0.05. Cultures were grown in 50mL falcon tubes at 30$^{o}$C with shaking at 200 rpm. 250 µL samples were taken at various time points for use in flow cytometry. Samples were analysed using a Beckman Coulter CytoFLEX S flow cytometer using the recommended gain values based on calibration beads supplied by Beckman Coulter. Samples were diluted 50 µL in 250 µL milli-Q water when unmeasured events increased above 10%. A BY4742 *S. cerevisiae* strain containing the PRS416 plasmid was used to establish a representative gating for forward and side scatter parameters. This gating was used in recording 10000 events for each sample in flow cytometry experiments.

Fluorescence output was measured at the 525 nm and 610 nm wavelengths to quantify eGFP, and mCherry expression respectively, measured in arbitrary fluorescence units (AFU). Fluorescence from biosensor plasmid containing *S. cerevisiae* cells was normalised to *S. cerevisiae* cells containing the PRS416 plasmid to account for auto fluorescence. Normalisation was done via subtracting the average AFU from three biological replicates of the PRS416 plasmid containing strains from each biosensor plasmid containing AFU value. For each growth condition tested on a biosensor containing strain, an identical growth condition was used on the PRS416 containing strain for normalisation.

## 4.8: Statistical Analysis

Statistical comparison of fluorescence were done within the GraphPad Prism 6 software package using unpaired t tests with Welch's correction.

# 5: Results

## 5.1: Sc 2.0 Synthetic Chromosome XIV

The Sc 2.0 project aims to create a strain of yeast with a fully synthetic genome. To do this, each synthetic chromosome is being constructed in separate *S cerevisiae* strains before being combined. Construction of synthetic chromosomes for the Sc 2.0 project is an expensive and time consuming process, spanning many labs around the world. The SCRaMbLE system which has been designed as part of the synthetic genome represents an extremely important tool for ALE experiments.

To aid in the development of the SCRaMbLE system, one of the aims of this master's project was to assist in the construction of synthetic chromosome 14 as part of the Sc 2.0 project. This involved the integration of approximately 80 kilobases of synthetic DNA into chromosome 14 of *S. cerevisiae* strain 14 G-T. *S. cerevisiae* strain 14 G-T already had approximately 470 kilobases of native chromosome 14 replaced with 440 kilobases of synthetic chromosome 14. Replacement with synthetic DNA was done in a stepwise fashion using a series of "mega-chunks". *S. cerevisiae* strain 14 G-T already had mega-chunks G through T integrated leaving megachunks A-F, and U-X still requiring integration.

Part of this master's project involved the integration of megachunks U, V, and W for the construction of synthetic chromosome 14. The stepwise addition of these megachunks was achieved via yeast assembly of component chunks and integration using homologous recombination. After each attempted integration of a megachunk, initial screening was done to identify yeast colonies which had taken up the synthetic DNA and lost the corresponding native sequence. Two sets of PCR probes had previously been designed to differentiate between Sc 2.0 and native DNA after each attempted megachunk integration [56]. That is, one set of probes would amplify if the DNA at a specific site was synthetic, but not if it was native, the other set of probes would amplify if the DNA at the same site was native, but not if it was synthetic. Utilizing these probes, qPCR was used for initial screening of colonies that may have successfully integrated a megachunk, and lost the corresponding native sequence. An example of this is seen in Figure 6.

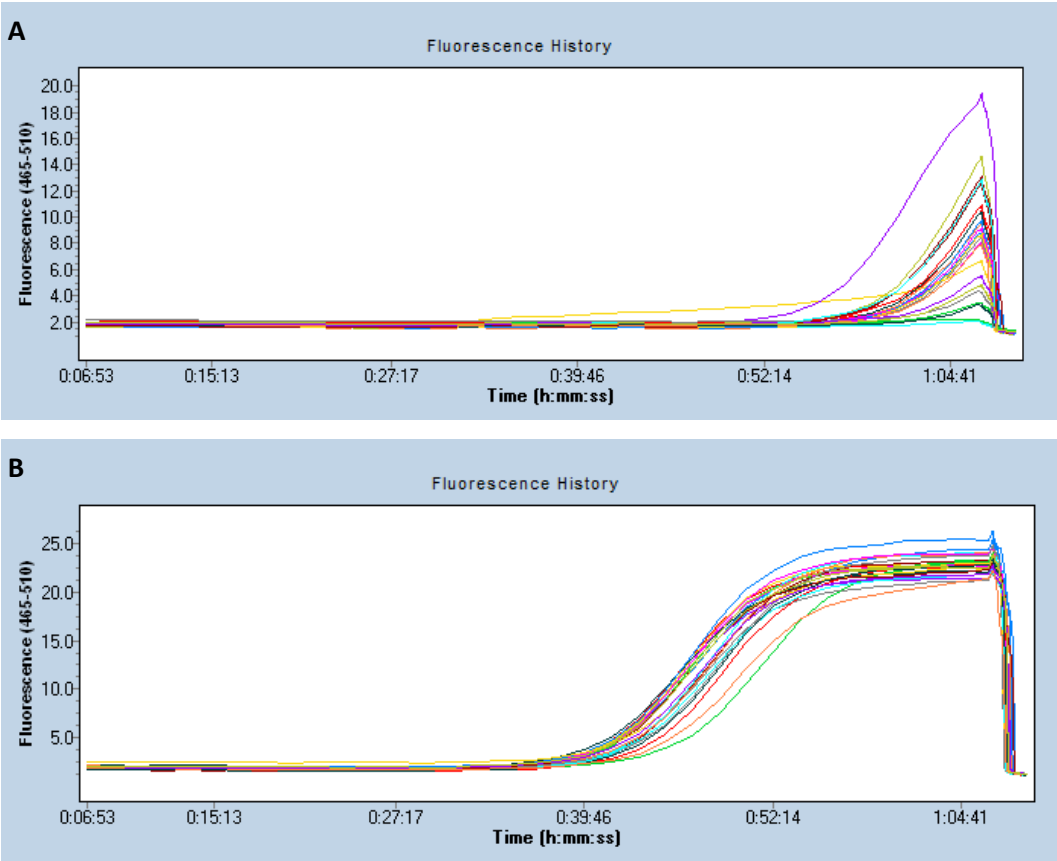


Figure 6: An example of the qPCR output from one colony which underwent successful megachunk transformation. **A** shows the amplification of gDNA using primers designed to amplify only on the native chromosome 14 sequence. The late peak in amplified products after approximately 55 minutes was indicative of non-specific amplification. Indicating that native chromosome 14 had likely been displaced. **B** shows the amplification of gDNA using primers designed to amplify on megachunk U. The earlier onset amplification of products after approximately 38 minutes, and the more uniform increase in fluorescence between primer pairs indicated that megachunk U had likely been successfully integrated.

To confirm that megachunks U, V, and W had integrated correctly, removing the corresponding native sequences, whole genome sequencing was performed. The generated reads were mapped to the S288c reference genome for chromosomes 1-13 and 15-16, with the native chromosome 14 sequence replaced by the Sc 2.0 chromosome 14 sequence. This resulted in an assembly with an average read coverage depth of 174. Analysis of read coverage and agreement to the Sc 2.0 chromosome 14 template were then completed manually to assess the accuracy of megachunk integrations.

There was a marked increase in read disagreement for approximately 148 bases between position 642 098- 641 951 in chunk U3. Many reads showed small amounts of disagreement to the template, however no one error was well repeated within the reads. This is likely due

to the highly repetitive region in this sequence. Issues with read disagreement from repetitive regions are a well-known artefact of illumina sequencing, and as the overall consensus still matched the synthetic template, this region was considered successfully integrated.

There was a single nucleotide polymorphism (SNP) at position 647 721 in chunk U3 with almost all reads showing the same disagreement with the template. This impacts two open reading frames (ORF) which are present on opposite DNA strands. These are YNR024W, and YNR025C. In YNR025C this SNP causes a synonymous sequence change in codon 48 from TTT to TTC, retaining phenylalanine as the encoded amino acid. In YNR024W codon 170, the sequence AAG is changed to AAA. This is also a synonymous change, retaining lysine as the encoding amino acid. It is unknown at which stage in the construction process this SNP occurred. However, as this SNP does not alter the final amino acid sequence of either ORF it was deemed acceptable, and did not require alteration.

Part of Sc 2.0 synthetic chromosome design involved placing custom sites for PCR tag analysis to differentiate between wild type and synthetic DNA. At position 660 212 to 660 237 in chunk U5 the PCR tag site indicated that wild type DNA was still present at this location. This indicated that synthetic DNA may not have replaced the wild type sequence for a region up to 1576 base pairs between the upstream and downstream PCR tags. This effects the 5' upstream regions of two ORFs, YNR032W and YNR031C. The major design goals of Sc 2.0 involved the removal of non-essential introns, transposons, the conversion of all stop codons to either TAA/TGA, and addition of PCR tag sites, and *loxP* sites 3' of ORFs. Fortunately none of the major design goals are impacted by this failed integration. As such the sequence is still in line with the original design and did not require alteration.

Finally, there was another PCR tag site indicating that native DNA was still present in a region spanning 692 246 to 692 150 in chunk V4. This occurred 3' of the ORF YNR051C, resulting in a missing *loxP* site and a stop codon failing to be re-coded. As these are two very important goals of the Sc 2.0 project this will require amelioration.

## 5.2: Synthetic Minimal Core Promoter

The butanol biosensors designed in this project made use of a *S. cerevisiae* synthetic core promoter (SCP). This SCP was designed to contain the minimal architecture necessary to allow initiation of transcription [62]. The structure of the SCP from 5' to 3' is as follows, 45 base pairs of a neutral AT-rich region, TATA box, 30 bp of experimentally validated random spacing sequence, and the *PGK*1 transcription start site. This structure allows the SCP to function as an unbiased minimal promoter, relatively free from native regulatory interference. Paired with an upstream activating sequence/transcription factor binding site, the SCP can then act as an inducible promoter. In this project, the design of the butanol sensors relied heavily on the SCP to work in conjunction with the Vp-16 activation domain. These two components have never before been used together, and as such their ability to induce transcription was unknown. Therefore, a construct was built to verify that a Vp16 activation domain, brought into close proximity of the SCP, was capable of inducing transcription of an *eGFP* reporter gene (Figure 7).

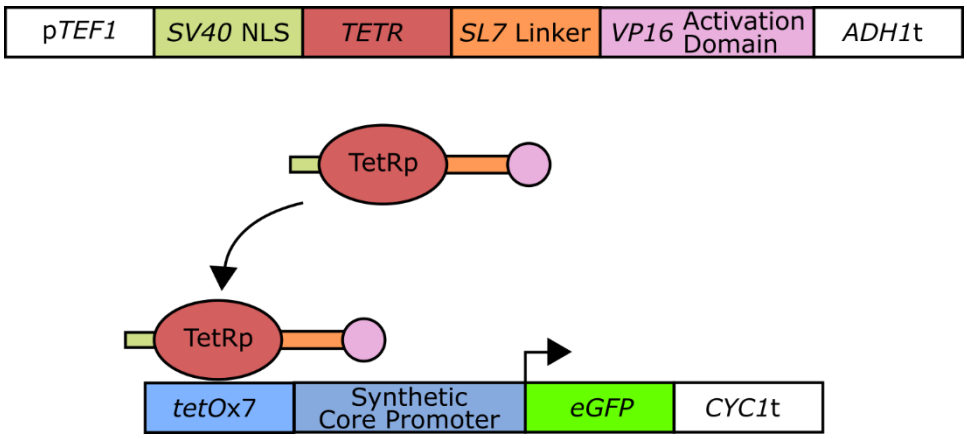


Figure 7: Simplified representation of the "SCP-*eGFP*" biosensor. Top and bottom band show DNA at different positions on a PRS416 plasmid (Table 2). TetRp, represented here as an oval, is expressed from the *TEF1* promoter as a fusion protein with the SV40 nuclear localisation sequence, SL7 linker, and Vp16 activation domain. Upon translation of the fusion protein, localisation to the *tetO*x7 sequence induces transcription of *eGFP* from the synthetic core promoter.

This construct (designated SCP-*eGFP*) made use of the protein TetR, and its known binding affinity to a DNA sequence *tetO* [66]. Seven repeats of the *tetO* sequence was placed upstream of the SCP, which was controlling the expression of an *eGFP* gene. The TetR protein was constitutively expressed from the *TEF1* promoter as a fusion protein with a SV40 nuclear localisation sequence, SL7 linking domain, and a Vp16 activation domain. In theory, this fusion protein would bind to the *tetO* binding sequence, bringing the Vp16 activation domain into close proximity of the SCP, inducing transcription of the *eGFP* reporter gene. In effect, this would appear the same as constitutively expressing *eGFP*. Furthermore, it is known that the binding of TetRp to the *tetO* sequence is inhibited by the compound doxycycline [66]. As such, if the expression of *eGFP* from the SCP was dependent on the binding of TetRp, then the addition of doxycycline would reduce expression. To test that the Vp16 activation domain was capable of inducing expression from the SCP, the SCP-*eGFP* construct was expressed from a PRS416 plasmid in *S.cerevisiae* BY4742 cells grown in Sd media, with and without 1µg/mL doxycycline (Figure 8).

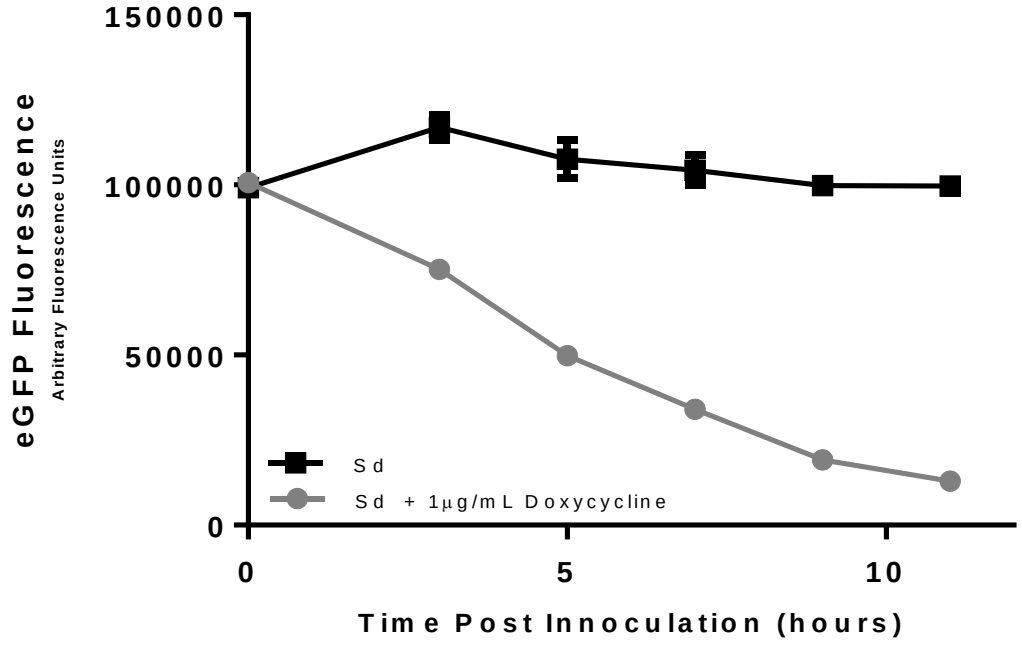


Figure 8: Fluorescence observed in *S. cerevisiae* BY4742 cells containing the SCP-*eGFP* biosensor plasmid. Cells were grown in Sd media with or without 1 µg/mL doxycycline for 11 hours. Error bars shown are standard deviation from three biological replicates.

Figure 8 shows the fluorescence of SCP-*eGFP* containing cells normalised to the PRS416 negative control plasmid. It can be seen that SCP-*eGFP* containing cells have a high level of fluorescence when grown in YNB. This suggests that the SCP is capable of inducing transcription when in close proximity to a Vp-16 activation domain. Furthermore, when grown in the presence of 1µg/mL doxycycline this fluorescent activity was reduced significantly (7.7 fold difference after 11 hours P = 0.0002). This demonstrates that the observed fluorescent activity is dependent on the TetRp-Vp16 activation domain. Together, this data showed that the Vp16 activation domain could be used in conjunction with the SCP for the proposed butanol biosensors.

## 5.3: Building a butanol biosensor for the ALE of butanol production

Using the BmoRp transcription factor from *T. butanovorans* as a basis, a butanol biosensor was designed for use in *S. cerevisiae* strain BY4742. In its native context, the BmoRp transcription factor responds to increased butanol concentrations by binding to an upstream regulatory sequence pBMO, inducing transcription of an alkane monooxygenase [52]. However, as *T. butanovorans* is a prokaryote, directly copying this transcription factor promoter pair was unlikely to be functional in *S. cerevisiae*. As such, the butanol biosensors designed in this project used the codon optimised BmoR protein as a fusion with an SL7 linking domain, and a Vp16 activation domain. The p*BMO* sequence could then be used as a binding site for the butanol activated BmoR-SL7-Vp16 fusion protein. Downstream of the p*BMO* site a SCP was placed, driving the expression of an *eGFP* reporter. Together, under conditions of increased butanol concentrations the BmoR-SL7-Vp16 fusion protein would bind to the p*BMO* site, inducing transcription from the nearby SCP, producing *eGFP*. This biosensor was designated "*BMOR-VP16*" (Figure 9).

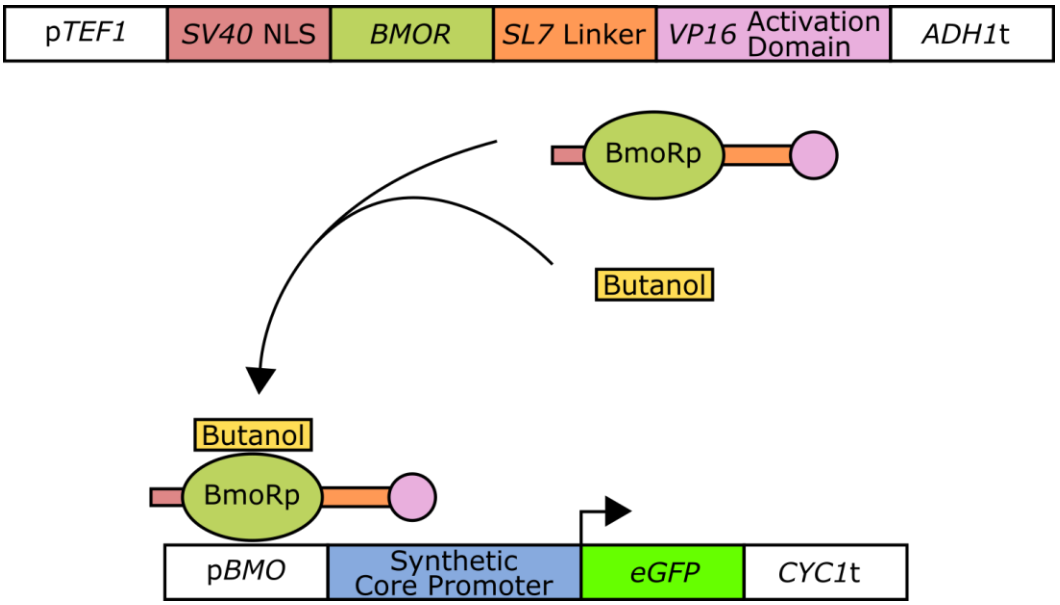


Figure 9: Simplified representation of the "*BMOR-VP16*" biosensor. Top and bottom band show DNA at different positions on a PRS416 plasmid (Table 2). BmoRp, represented here as an oval, is expressed from the *TEF1* promoter as a fusion protein with the SV40 nuclear localisation sequence, SL7 linker, and Vp16 activation domain. Upon translation of the fusion protein, localisation to the p*BMO* sequence induces transcription of *eGFP* from the synthetic core promoter.

The effectiveness of this sensor was assayed using a flow cytometric measurement of fluorescence emitted by *S. cerevisiae* BY4742 cells containing the *BMOR-VP16* biosensor on a PRS416 plasmid. Cells were cultured in Sd media with various concentrations of butanol for 11 hours (Figure 10). All fluorescence values were normalised to BY4742 cells containing an empty PRS416 plasmid grown under identical conditions.

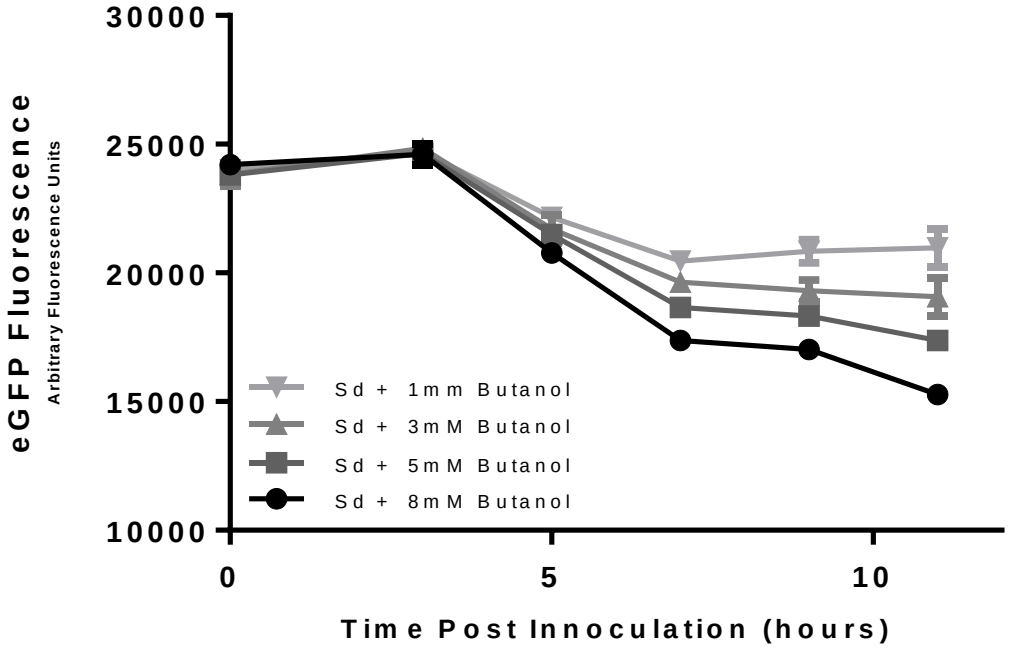


Figure 10: Fluorescence observed in *S. cerevisiae* BY4742 cells containing the "*BMOR-VP16*" biosensor plasmid. Cells were grown in Sd media with varying concentrations of butanol for 11 hours. Error bars shown are standard deviation from three biological replicates.

As can be seen in Figure 10 cells containing the *BMOR-VP16* sensor did not show an increase in fluorescence after induction with a range of butanol concentrations. Furthermore, it appeared that fluorescence decreased with increasing butanol concentration, the inverse of the desired response. Of particular interest is the very high level of basal fluorescence being produced by these cells. Even without induction, cells containing the *BMOR-VP16* biosensor showed fluorescence values approximately 13 times higher than cells not containing the biosensor. One explanation for this was that the BmoR-Vp16 fusion protein was constantly active. Rather than butanol causing a conformational change allowing the SV40-BmoRp-SL7-Vp16 fusion protein to bind to the p*BMO* sequence, it may be constantly active, regardless of butanol concentration. This would cause constant production of a high level of eGFP. This theory would also explain the growth defect present in the *BMOR-VP16* containing cells (Figure 11). Figure 11 demonstrates that BY4742 cells containing the *BMOR-VP16* sensor have a marked reduction in growth over the 11 hour time period, independent of butanol concentration. This growth reduction may be caused by cells constantly expressing large quantities of eGFP resulting in a significant metabolic burden.

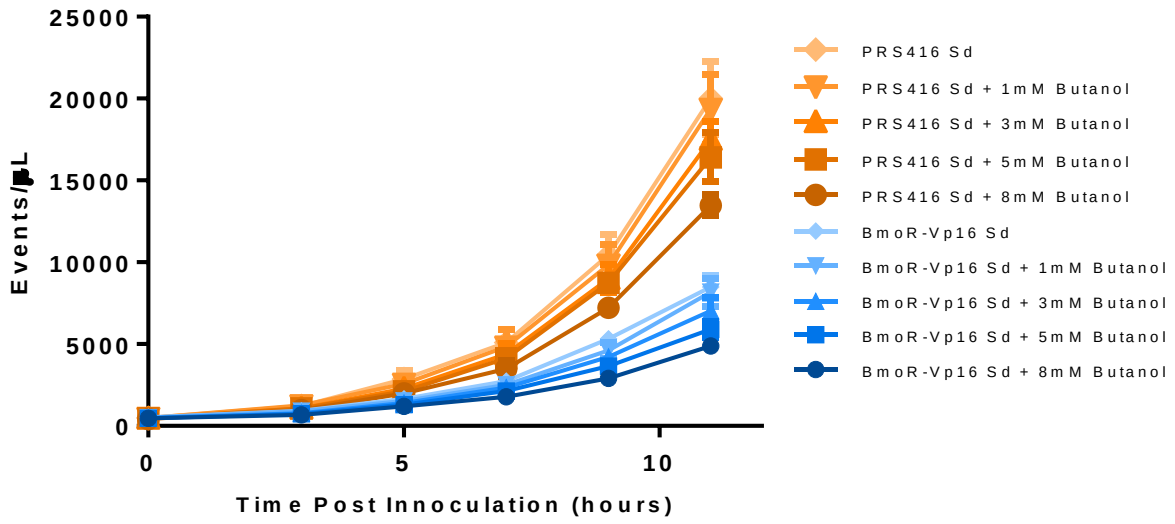


Figure 11: Growth of *S. cerevisiae* BY4742 cells containing the *BMOR-VP16* biosensor as determined by events/µL recorded in flow cytometric analysis. *BMOR-VP16* (blue), and *S. cerevisiae* cells with the PRS416 plasmid (orange). Cells were grown in Sd media with varying concentrations of butanol for 11 hours. Error bars shown are standard deviation from three biological replicates.

One possible reason that the *BMOR-VP16* version of the biosensor might be constantly active is because the combination of the SL7 linker, Vp16 activation domain and the SV40 nuclear localization sequence are disrupting the native structure and function of the BmoR protein. Additionally, it is unknown with what orientation BmoR binds to the p*BMO* sequence, and by extension, the position of the Vp16 activation domain relative to the SCP. To attempt to fix these potential disruptions, the order of these domains were inverted, resulting in a sensor with *VP16* activation domain, *SL7* linker, *BMOR*, and finally *SV40* nuclear localisation sequence (Figure 12). This version of the sensor will be referred to as "*VP16-BMOR*".

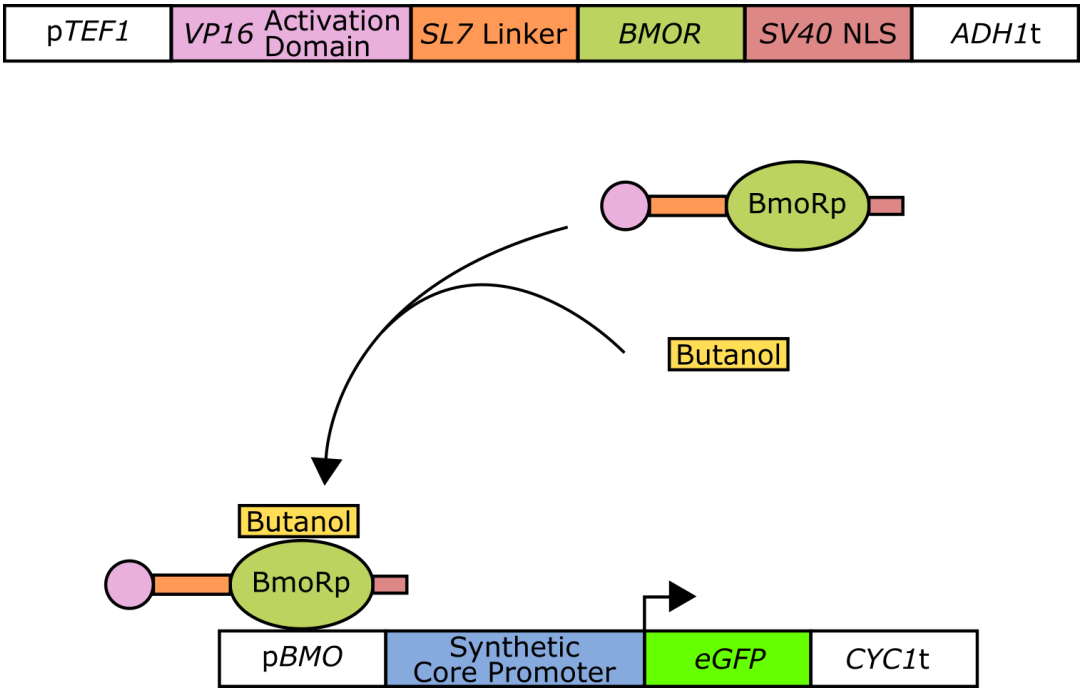


Figure 12: Simplified representation of the "*VP16-BMOR*" biosensor. Top and bottom band show DNA at different positions on a PRS416 plasmid (Table 2). BmoRp, represented here as an oval, is expressed from the *TEF1* promoter as a fusion protein with the SV40 nuclear localisation sequence, SL7 linker, and Vp16 activation domain. Upon activation of the fusion protein, localisation to the p*BMO* sequence induces transcription of *eGFP* from the synthetic core promoter.

*VP16-BMOR* was assayed in a similar fashion to *BMOR-VP16*. Flow cytometric analysis was completed on *S. cerevisiae* BY4742 cells containing the *VP16-BMOR* biosensor on a PRS416 plasmid grown in Sd media with various concentrations of butanol (Figure 13). Fluorescence values were normalised to BY4742 cells containing an empty PRS416 plasmid grown under identical conditions.

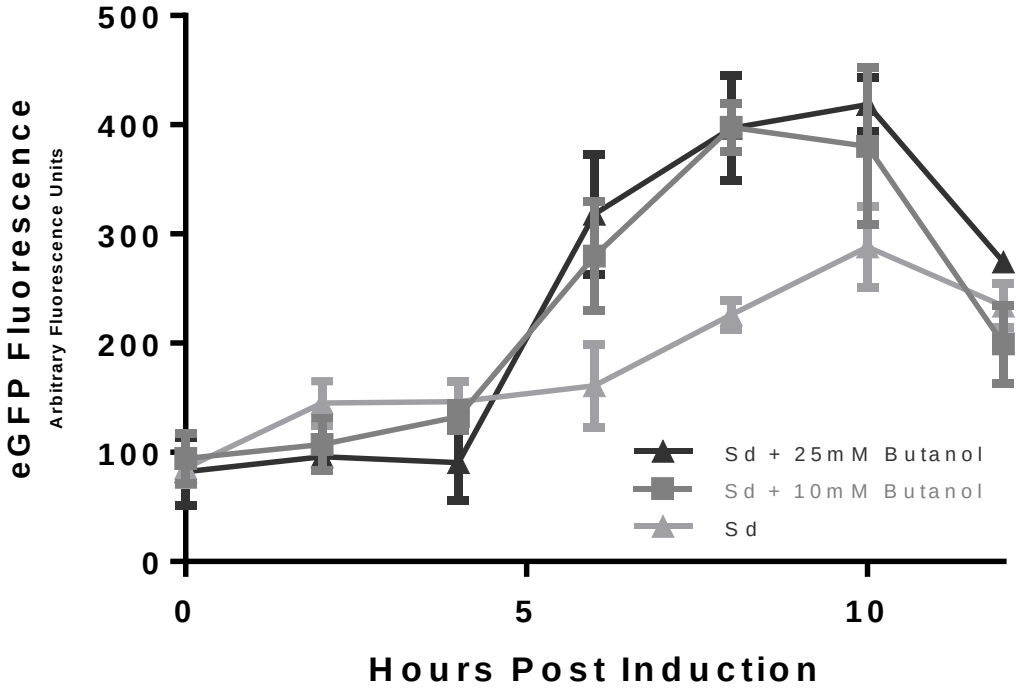


Figure 13: Fluorescence observed in *S. cerevisiae* BY4742 cells containing the "*VP16-BMOR*" biosensor plasmid. Cells were grown in Sd media with varying concentrations of butanol for 12 hours. Error bars shown are standard deviation from three biological replicates.

Figure 13 shows the response of *S. cerevisiae* BY4742 cells containing the *VP16-BMOR* butanol biosensor to 25mM and 10mM exogenous butanol. The *VP16-BMOR* version of the biosensor showed a significant improvement over the initial design. A 1.7 fold difference in fluorescence can be detected at the 8 hour time point (P = 0.0014). The drop in relative fluorescence after this time point might be the culture beginning to enter stationary phase, downregulating the *TEF1* promoter which is expressing the Vp16-SL7-BmoR fusion protein. Regardless of the loss of signal at the 10 hour time point, the significant signal at 8 hours is viable for use in a butanol biosensor.

Preliminary analysis was done to investigate the linear range of the *VP16-BMOR* biosensor. This was done by culturing *S. cerevisiae* cells in Sd medium with varying concentrations of butanol ranging from 0mM to 11.2mM and measuring fluorescence after 8 hours of growth (Figure 14).

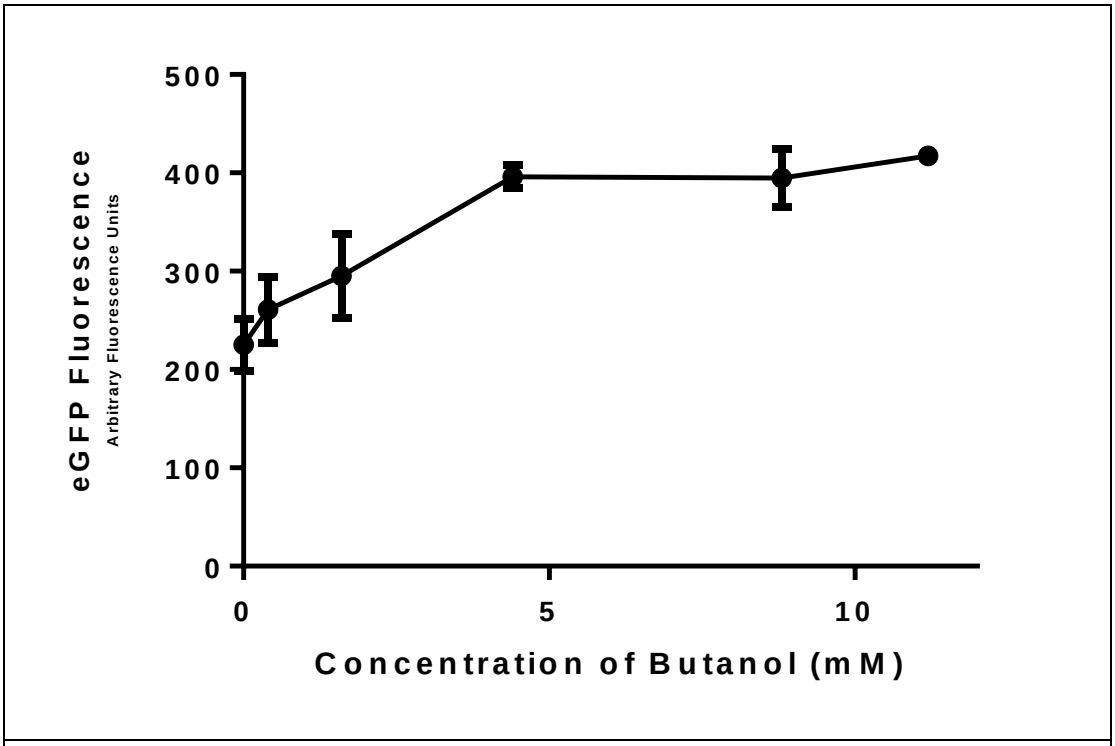


Figure 14: Fluorescence observed in *S. cerevisiae* BY4742 cells containing the "*BMOR-VP16*" biosensor plasmid. Cells were grown in Sd media with varying concentrations of butanol with sampling occurring at the 8$^{th}$ hour post inoculation. Error bars shown are standard deviation from three biological replicates.

Whilst this is only a preliminary assay of linear range and is not sufficient to make definitive statements regarding the minimum and maximum concentrations detectible by this sensor, some basic inferences can be made. Figure 14 shows that the sensor is likely saturated at a butanol concentration between 1.1mM and 4.4mM.

## 5.3: Building a methionine biosensor

Design of the methionine biosensor was based on the promoter for the native *S. cerevisiae* gene *MET17*. The *MET17* gene encodes a sulfhydrylase, catalysing the last step in the sulfate reduction pathway [67]. In response to elevated levels of methionine *MET17* is down regulated, along with the majority of genes involved in the sulfur network. This allows the cell to downregulate *de-novo* methionine synthesis if it is available in the immediate environment. The reduction in expression from the *MET17* promoter in response to methionine served as the basis of several biosensor designs.

One version of the biosensor attempted to invert the signal coming from the *MET17* promoter so that an increase in methionine concentration would result in an increase in signal output (Figure 15). In this sensor, p*MET17* drives the expression of a *TETR* transcriptional repressor. This repressor would then bind to a *tetO* binding sequence placed between p*TEF1* and an *eGFP* gene. Thus, when p*MET17* is transcriptionally active under conditions of low methionine, the TetR protein would be expressed, allowing it to bind to the *tetO* sequence, inhibiting p*TEF1* from transcribing *eGFP*. Furthermore, when p*MET17* is transcriptionally repressed under conditions of high methionine, the TetR protein is not transcribed, allowing p*TEF1* to induce the expression of *eGFP*. As placing a *tetO* binding sequence between the *TEF1* promoter and the *eGFP* gene may disrupt transcription, two variants of this sensor were made. One had a short *tetO* binding sequence of 42 base pairs (designated p*MET17-TETR-tetO*), and the other used the traditional 7 repeats of the binding sequence (designated p*MET17-TETR-tetO*x7).

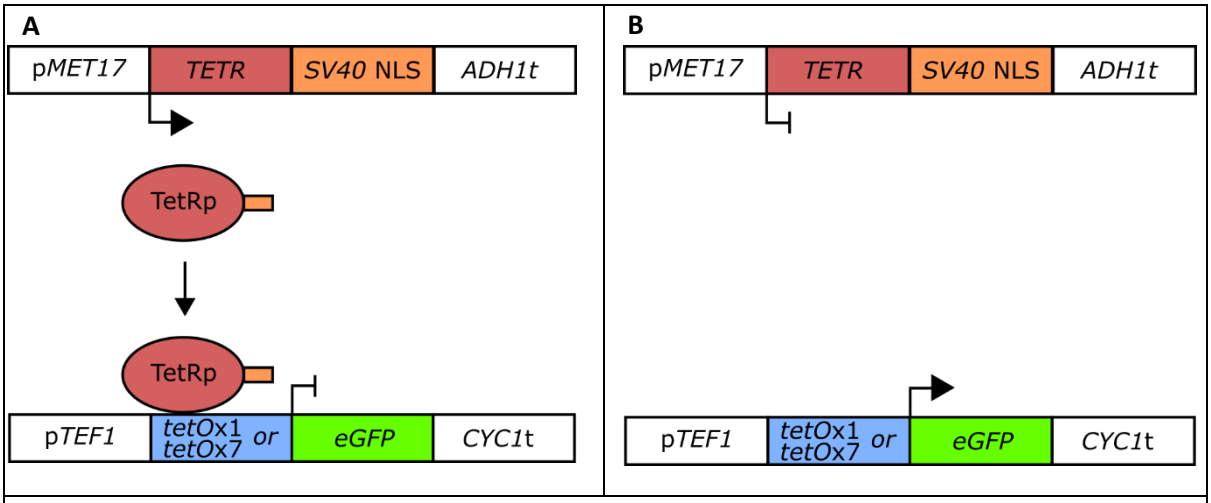


Figure 15: Simplified representation of the "p*MET17-TETR-tetO*" and "p*MET17-TETR-tetO*x7" biosensors. **A** shows the predicted function under conditions of low methionine concentrations. **B** shows the predicted function under conditions of high methionine concentration. Top and bottom bands show DNA at different positions on a PRS416 plasmid (Table 2). TetRp, represented here as an oval, is expressed from the *MET17* promoter as a fusion protein with the SV40 nuclear localisation sequence. Upon translation of the fusion protein localisation to the *tetO* or *tetOx7* sequences inhibits the transcription of *eGFP*

To assess the effectiveness and the response time of p*MET17-TETR-tetO* and p*MET17-TETR-tetO*x7 biosensors, a time course was undertaken measuring the response of *S. cerevisiae* BY4742 cells containing a version of the biosensors expressed from a PRS416 plasmid in the presence of 5mM or 1mM methionine (Figure 16). Fluorescence was normalised to BY4742 cells containing an empty PRS416 plasmid grown under identical conditions.

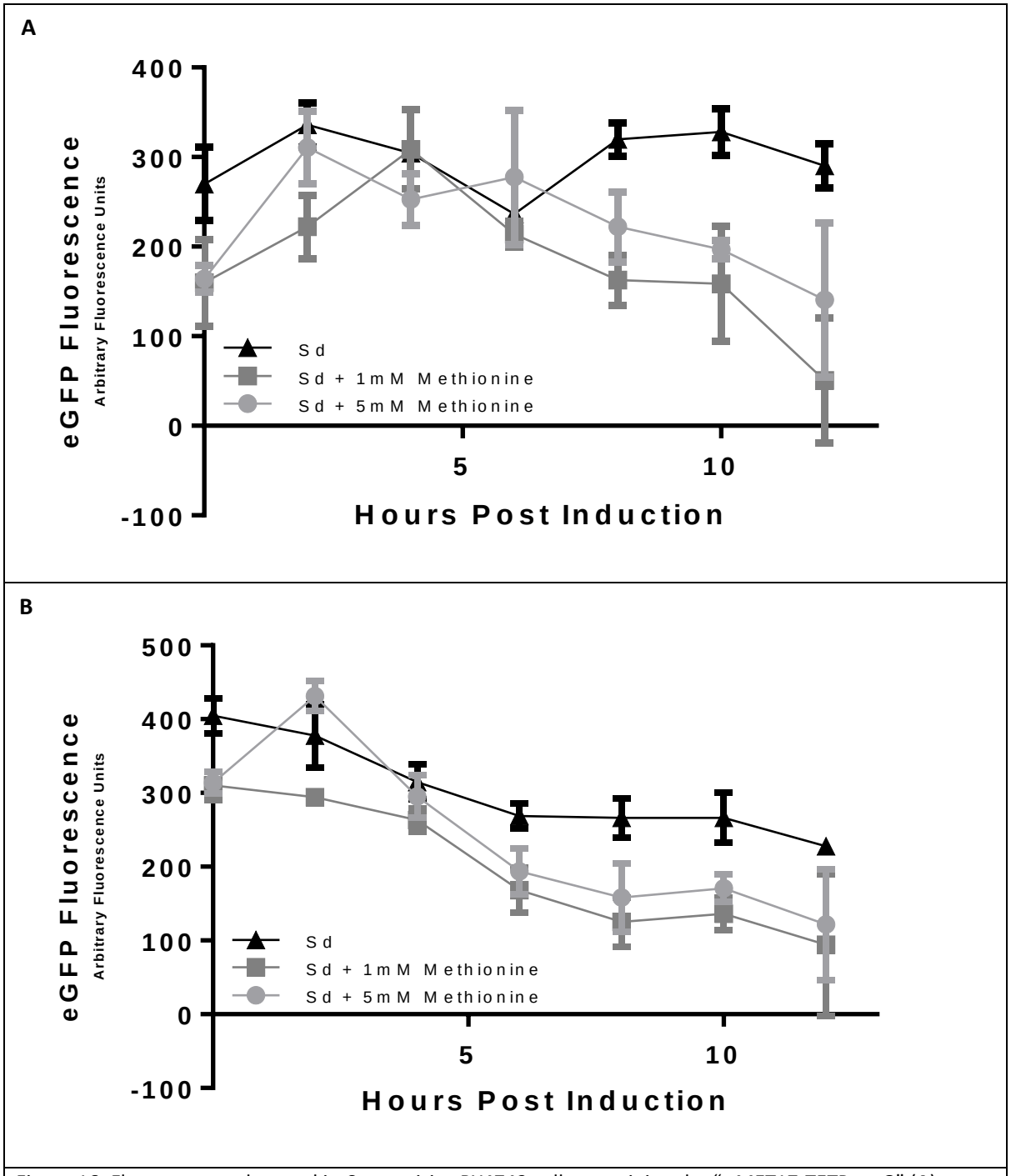


Figure 16: Fluorescence observed in *S. cerevisiae* BY4742 cells containing the "p*MET17-TETR-tetO*" (**A**) or "p*MET17-TETR-tetO*x7" (**B**) biosensor plasmid. Cells were grown in Sd media with varying concentrations of methionine for 12 hours. Error bars shown are standard deviation from three biological replicates.

Figure 16 shows that neither the p*MET17-TETR-tetO* or p*MET17-TETR-tetO*x7 versions of the biosensor were capable of detecting exogenous methionine. This could be because after p*MET17* stops transcribing *TETR* it may take longer than 12 hours for the TetRp to degrade, persisting in the cell and repressing *eGFP* transcription. It could also be that as p*MET17* is being downregulated the concentration of TetRp does decrease, but not enough to allow

p*TEF1* to read through the intervening *tetO* sequences and transcribe *eGFP*. Both of these explanations suggest that the concentration of TetR protein within the cell may be too high in the first 12 hours. Additionally, it is possible that the presence of the *tetO* sequences between the *TEF1* promoter and the *eGFP* start codon are inhibiting transcription.

To investigate these theories, an experiment was carried out with p*MET17-TETR-tetO* and p*MET17-TETR-tetO*x7 containing cells. This assessed their response to different concentrations of doxycycline. As was mentioned previously, doxycycline can be used to disrupt the binding of the TetR protein to the *tetO* sequence [66]. Therefore, providing an overabundance of doxycycline would keep the TetR protein unbound from the *tetO* sequence allowing p*TEF1* to transcribe *eGFP* regardless of methionine concentrations. If eGFP increased after treatment with doxycycline, this would suggest that there is an overabundance of TetRp under normal culture conditions. Furthermore, if saturating levels of doxycycline didn't increase *eGFP* expression then this would suggest that the problem rests with the ability of p*TEF1* to induce transcription with the intervening *tetO* sequences. A time course was undertaken measuring the response of *S. cerevisiae* BY4742 cells containing p*MET17-TETR-tetO* or p*MET17-TETR-tetO*x7 biosensors on PRS416 plasmids. Cells were cultured for 10 hours in Sd media with 0, 1, or 3 µg/mL doxycycline and fluorescence assayed using flow cytometry (Figure 17). Fluorescence was normalised to BY4742 cells containing an empty PRS416 plasmid grown under identical conditions.

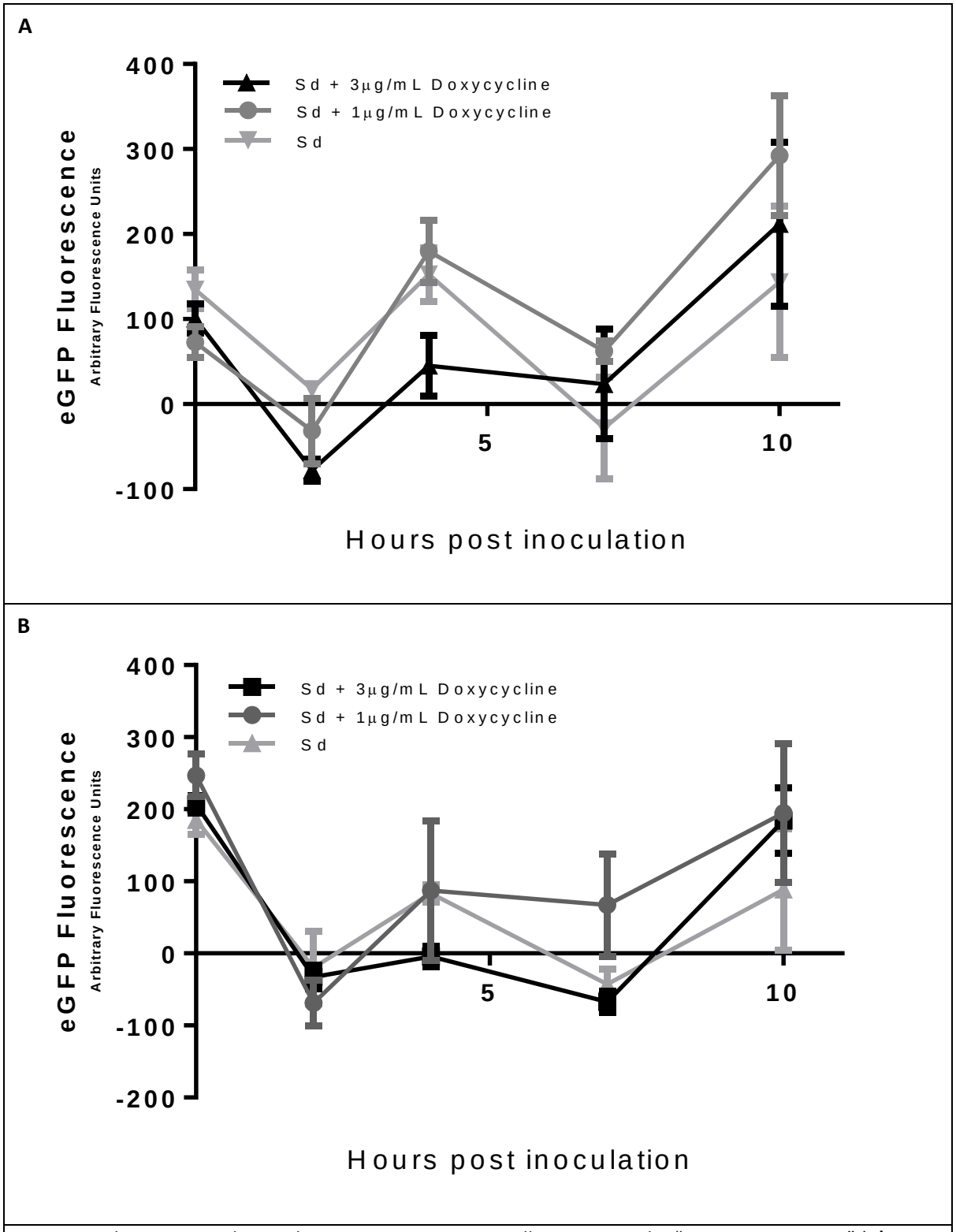


Figure 17: Fluorescence observed in *S. cerevisiae* BY4742 cells containing the "p*MET17-TETR-tetO*" (**A)** or "p*MET17-TETR-tetO*x7" (**B**) biosensor plasmid. Cells were grown in Sd media with varying concentrations of doxycycline for 10 hours. Error bars shown are standard deviation from three biological replicates.

Figure 17 showed that fluorescence was not significantly different for either version of the biosensor, irrespective of doxycycline concentration. 1 µg/mL is the standard concentration used in Tet-On and Tet-Off experimentation, and was shown in Figure 8 to be effective at

significantly reducing TetRp binding to the *tetO* sequence while testing the SCP-*eGFP* construct. Given that both 1 and 3 µg/mL doxycycline concentrations were tested with the methionine biosensors, it seems likely that overabundance of TetRp was not causing the lack of eGFP expression observed. Instead this result suggests that the placement of the *tetO* sequences between p*TEF1* and the *eGFP* gene disrupted either transcription or translation of *eGFP*. Meaning that even if *TETR* is successfully down-regulated via p*MET17* in response to methionine, p*TEF1* would not be able to induce transcription of *eGFP*.

As the p*MET17-TETR-tetO* and p*MET17-TETR-tetO*x7 biosensors appeared to have some issues regarding *eGFP* transcription, an alternative biosensor was also developed. Rather than invert the signal coming from the *MET17* promoter, this alternative methionine sensor simply used the loss of expression from p*MET17* as the signal. The biosensor, referred to as "p*MET17-eGFP*", used p*MET17* to drive the expression of *eGFP* whilst *mCherry* was constitutively expressed by p*TEF1* (Figure 18). This allowed flow cytometric analysis to measure the loss of eGFP fluorescence as p*MET17* was down-regulated in response to methionine, whilst simultaneously providing a mCherry signal for normalisation. Inclusion of a mCherry signal for normalization was done to reduce the amount of false positives in the population which weren't expressing *eGFP* for more general reasons. For example, without mCherry normalization, slower growth, inhibited protein production, or the formation/accumulation of toxic by-products would likely result in reduced *eGFP* expression and therefore would incorrectly appear as methionine detecting cells.

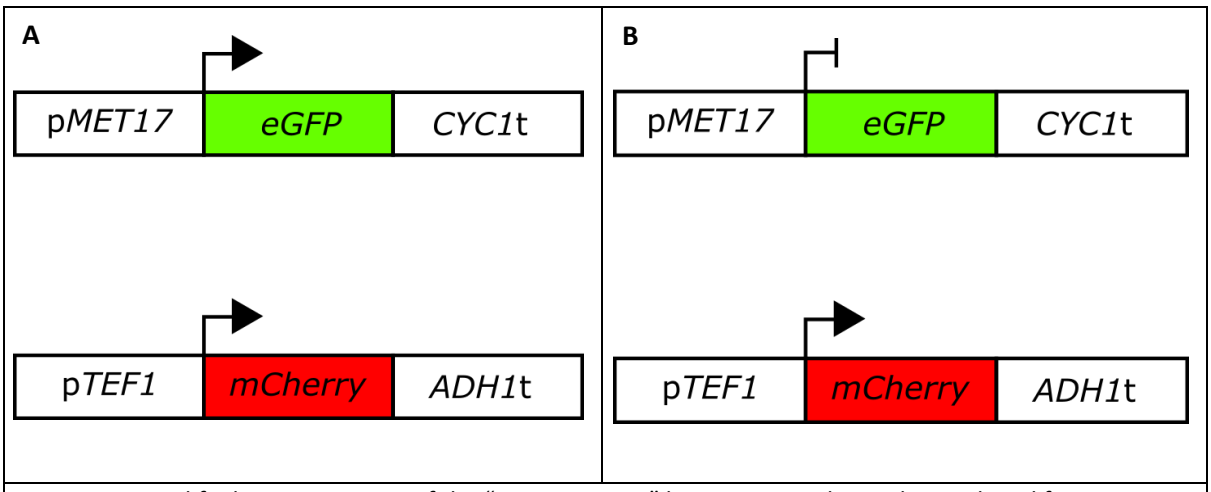


Figure 18: Simplified representation of the "p*MET17-eGFP*" biosensors. **A** shows the predicted function under conditions of low methionine concentrations. **B** shows the predicted function under conditions of high methionine concentration. Top and bottom bands show DNA at different positions on a PRS416 plasmid (Table 2).

Cells containing the p*MET17-eGFP* biosensor on the PRS416 plasmid were grown in Sd medium with either 0mM or 5mM methionine for 11 hours. Activity of the biosensor was assayed using flow cytometric measurement of fluorescence (Figure 19). Fluorescence was normalised to BY4742 cells containing an empty PRS416 plasmid grown under identical conditions.

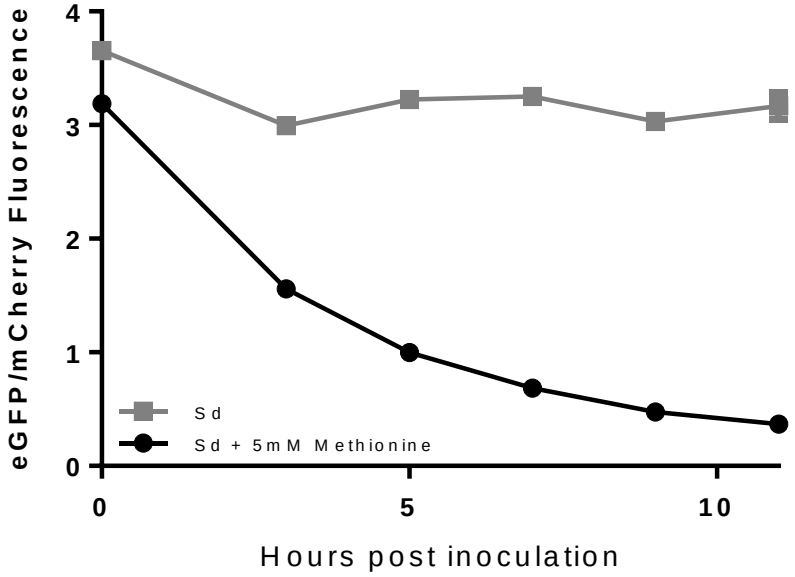


Figure 19: Fluorescence observed in *S. cerevisiae* BY4742 cells containing the "p*MET17-eGFP"* biosensor plasmid. Cells were grown in Sd media with or without 5mM methionine for 11 hours. Error bars shown are standard deviation from three biological replicates.

Figure 19 shows the drop in eGFP fluorescence relative to mCherry of *S. cerevisiae* BY4742 cells containing the p*MET17-eGFP* biosensor in response to 5mM of methionine. There was an 8.6 fold difference in fluorescence between induced and un-induced cells after 11 hours (P value = 0.0056). This is a significant difference in *eGFP* expression in response to methionine, and demonstrates that the p*MET17-eGFP* biosensor has strong capability to be used as a methionine biosensor.

A preliminary assessment of the p*MET17-eGFP* biosensor's linear range was also conducted. *S. cerevisiae* BY4742 cells containing the biosensor on the PRS416 plasmid were grown in Sd medium with concentrations of methionine ranging from 0mM to 5mM. Their fluorescent response was measured using flow cytometry (Figure 20). Fluorescence was normalised to BY4742 cells containing an empty PRS416 plasmid grown under identical conditions.

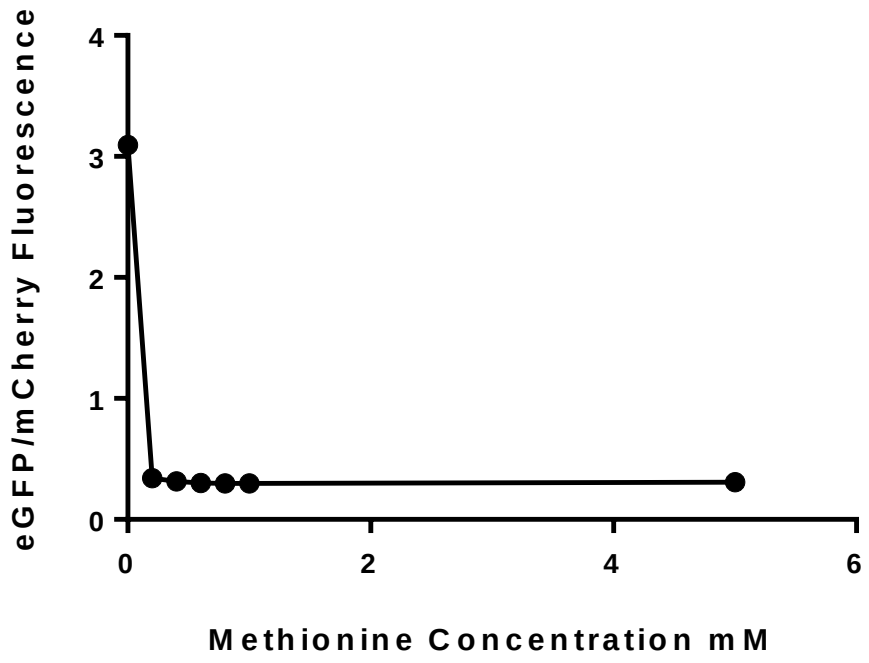


Figure 19: Fluorescence observed in *S. cerevisiae* BY4742 cells containing the "p*MET17-eGFP"* biosensor plasmid. Cells were grown in Sd media with varying concentrations of methionine with sampling being performed after 11 hours. Error bars shown are standard deviation from three biological replicates.

Figure 20 shows the loss of fluorescence after 11 hours of *S. cerevisiae* BY4742 cells containing the p*MET17-eGFP* biosensor. It can be seen that the concentrations used in this preliminary assessment were far too high, as the sensor is saturated with the addition of 0.2mM of methionine and shows no further reduction in fluorescence above that concentration.

# 6: Discussion

## 6.1: Sc 2.0

The Sc 2.0 project, and by extension the SCRaMbLE technique, will create a *S. cerevisiae* genome that is extremely amenable to ALE. SCRaMbLE can increase or decrease the copy number of non-essential genes in an un-biased fashion via *loxP* mediated recombination. Flux through metabolic pathways is extremely complex, and the most obvious genetic changes to improve it are often insufficient. The ability to randomly alter gene copy number allows unbiased assessment of all possible enzymatic inputs into a metabolic pathway. This allows the screening of non-obvious genetic solutions to metabolic flux problems.

Significant progress in the construction of Sc 2.0 chromosome 14 has been achieved in this master's project. 80 kilobases of synthetic DNA encoding 53 genes were able to be integrated into Sc 2.0 strain 14 G-T, representing an 18% increase in completion and generating the strain Sc 2.0 strain 14 G-W. There were two minor alterations from the designed sequence which have no major impact on the functionality of the final strain. The first was a SNP in two ORFs in chunk U3 which encode synonymous amino acid translation. The second was a failed integration in chunk U5 where the synthetic design does not differ from the wild type sequence. However, a more significant alteration is also present in strain Sc 2.0 14 G-W. Part of chunk V4 was not correctly integrated, and as a result a *loxP* site is missing, and the stop codon from ORF YNR051C has not been recoded. As such, this strain will require minor correction.

This will be done in two steps, the first is the re-transformation of chunk V4. This will be integrated over the affected sequence using homologous recombination. As chunk V4 has the Leu2 leucine auxotrophic rescue gene with it, successful recombinants can be easily selected for (Figure 21). Next, The CRISPR- Cas9 system will be used to induce a double stranded break in Leu2 whilst the cell is simultaneously transformed with the chunk W1. This will result in the homologous recombination of chunk W1 displacing the Leu2 gene from chunk V4 (Figure 21).

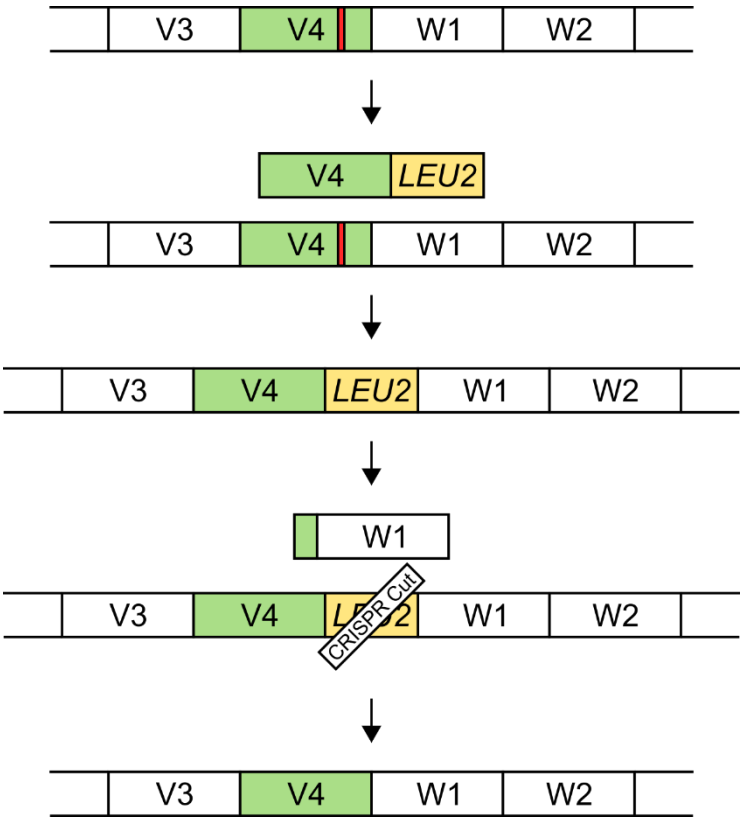


Figure 21: Simplified representation of the strategy that will be used to correct the SNP present in chunk V4 of the strain Sc 2.0 Syn XIV G-W strain.

Once these alterations are conducted, Sc 2.0 chromosome 14 will be 72% complete. The remaining mega-chunks to be integrated will then be A-F, and X. There are currently 6 completed synthetic yeast chromosomes out of the total 16, which are Syn VI, II, X, V, XII, and III [32, 68-72]. The construction of Sc 2.0 is a monumental advancement in *S. cerevisiae* research and for synthetic biology. Sc 2.0 will allow the construction of a minimal *S. cerevisiae* genome, which could be used to create a platform industrial strain, and advance fundamental *S. cerevisiae* research. Additionally, the information gained in synthetic genome construction has paved the way for other synthetic eukaryotic genomes to be assembled. Furthermore, the creation of the SCRaMbLE system has significant ramifications in the field of metabolic engineering using ALE. One of the largest hurdles facing ALE for metabolic engineering is ability to generate genetic diversity in a way that is biased towards gain of function mutations. That is, completely random mutation generates mostly deleterious genotypes. SCRaMbLE is poised to circumvent this problem by allowing a large genetic diversity to be sampled, without using completely random mutations. The integration of Lox P sites 3' of ORFs means that with the addition of a Cre recombinase protein, whole ORFs can be duplicated, deleted, or rearranged. In the context of ALE for

metabolic engineering, this allows the relative concentration of catalytic enzymes in production pathways to be altered due to differences in gene copy number. Without SCRaMbLE, it would likely be impossible to replicate the range of possible catalytic enzyme concentrations. For example, SNPs in a promoter sequence are capable of reducing transcriptional activity, but they are far less likely to induce increased transcription. Furthermore, random SNPs have a very high chance of occurring within an ORF, most likely resulting in a loss of function mutation. For these reasons, SCRaMbLE likely represents a superior technique for the generation of enzyme concentration diversity. With the aid of biosensors, SCRaMbLE can optimize the flux through complex metabolic pathways via routes that are non-obvious to a human engineer.

## 6.2: Butanol biosensor

Butanol offers a possible alternative to petrol for internal combustion engines. It can be used in conventional engines without requiring mechanical modification. If butanol could be produced in industrially relevant volumes many countries could transition away from fossil fuel powered transportation. *S. cerevisiae* offers a possible cell factory host for butanol production, but requires a functional butanol biosensor if ALE is to be used.

There are two proposed butanol biosensors for use in *S. cerevisiae* developed by Shi S, Choi YW, Zhao H, Tan MH and Ang EL [73] whose mechanisms of action are not understood. These were developed by transcriptomic analysis of *S. cerevisiae* strains in response to butanol exposure. This identified promoters $P_{YDL167C\text{-}T3}$ and $P_{YIL104C\text{-}T4}$, which are activated in response to butanol, but not to other short chain alcohols. These promoters can then be used to drive expression of a selective marker such as *eGFP*. However, these two promoters have a major limitation for use in biosensors. They control the expression of a putative RNA binding protein, and a small nucleolar ribonucleoprotein complex assembly factor, which are expressed in response to cellular stresses and hypoxia respectively. That is, neither promoter is specifically activated by butanol directly, simply by the downstream effects of butanol on fairly general stress responses. Whilst Shi S, Choi YW, Zhao H, Tan MH and Ang EL [73] did ensure that neither promoter was activated by other short chain alcohols, they did not consider any other off target activation. That is, it's possible that a wide variety of compounds could activate sensors using these promoters, not just butanol. In an ALE

experiment, this could result in selecting for any compound which triggers the correct stress response, rather than butanol production. Unwanted activation of these promoters poses major hurdles to their use as biosensors, especially in ALE experiments. Conversely, the *VP16-BMOR* butanol biosensor described in this study has the potential to circumvent the limitations of these existing butanol biosensors.

This project investigated using the butanol activated transcription factor BmoRp from *T. butanovorans* as part of a butanol biosensor in *S. cerevisiae*. This involved the fusion of BmoRp with the well characterised Vp16 activation domain to induce transcription from a *S. cerevisiae* synthetic core promoter. Of the two designs tested, the *VP16-BMOR* biosensor was successful in responding to exogenous butanol concentrations (Figure 13 and 14). The *BMOR-VP16* design showed no significant response to exogenous butanol, but had very high un-induced fluorescence values. This might be indicative that this configuration is resulting in a constitutively active fusion protein. This may also be causing metabolic burden to the cell, hence its reduced growth and sensitivity to butanol (Figure 11). However, this plasmid still needs to be sequenced to ensure that unknown SNPs were not responsible for this activity.

Promisingly, the *VP16-BMOR* version of the sensor is functional, showing a significant 1.7 fold difference in fluorescence between induced and un-induced populations (Figure 13). The sensor is most highly activated after 8 hours (Figure 13). Preliminary investigation into the linear range of this sensor has shown that its maximum detectible butanol concentration is somewhere between 1.6 mM and 4.4 mM (Figure 14). This is somewhat lower than the highest recorded production of butanol in *S. cerevisiae* at 11.265 mM [51]. Given how close maximum production and the *VP16-BMOR* detection limit are, this biosensor should be amenable to directed protein evolution to increase the maximum detectable concentration.

In its current state, the *VP16-BMOR* biosensor could be used in strains producing less than 1.6 mM butanol to identify mutations which increase butanol production. Each separate mutation could then be engineered into a single *S. cerevisiae* strain in an attempt to increase butanol production beyond 11.265 mM. However, not all mutations will act

additively, and some may even inhibit each other. Therefore, it would be more efficient, to have a biosensor which could detect higher concentrations of butanol.

Work will need to be done to increase linear range of the *VP16-BMOR* biosensor. This goal could be accomplished through the use of directed protein evolution. Random mutagenesis of the *VP16-BMOR* transcription factor via error prone PCR could generate a library of biosensor mutants which have a range of butanol affinities. Sensor variants with increased affinity would require a lower concentration of butanol to induce activation, lowering the minimum concentration possible for detection. Whilst sensor variants with decreased affinity for butanol would require increased butanol concentrations to induce transcription, increasing the maximum detectable concentration. This mutated sensor would then be ideal for increasing butanol production in *S. cerevisiae* via ALE.

More testing still needs to be done before a conclusive statement can be made about *VP16-BMOR*'s use in ALE. Whilst it has been shown that it is capable of detecting exogenously supplied butanol, whether it is capable of detecting endogenously produced butanol is still unknown. Therefore, a plasmid containing the sensor should be transformed into a known butanol overproducing strain such as *S. cerevisiae* "COM", which can produce up to 11.265 mM [51]. Furthermore, a BmoRp based biosensor developed in *E. coli* is known to respond with varying strengths to other short chain alcohols. The activation of the *VP16-BMOR* biosensor in response to these same short chain alcohols should also be investigated to ensure that butanol remains the primary activator in this design.

Despite the work that remains, *VP16-BMOR* represents the first butanol biosensor for use in *S. cerevisiae* for which the mechanism of action is understood. *VP16-BMOR* has the potential to be an excellent butanol biosensor for use in *S. cerevisiae*. With a known mechanism of action, it is unlikely to be activated by a wide range of cellular metabolites, representing a significant improvement on the biosensors suggested by Shi S, Choi YW, Zhao H, Tan MH and Ang EL [73]. Additionally, as the BmoR protein and its corresponding p*BMO* binding sequence have prokaryotic origins, they are more likely to remain transcriptionally isolated, free from possible interference from native transcriptional regulators.

In a more general context, the development of this sensor has also offered a design principle for using prokaryotic transcription factors as eukaryotic biosensors. Previous work in this area has involved quite complicated transcription factor engineering, requiring detailed understanding of the transcription factor to be used [74]. However, the fusion of a Vp16 activation domain via flexible SL7 linker to activate a generic SCP can easily be engineered for any known transcription factor. Whilst this technique might not work in all cases, it is a simple first step that should be considered when trying to use prokaryotic transcription factors to drive expression in a eukaryotic host.

## 6.3: Methionine Biosensor

Methionine is a limiting amino acid in the diets of poultry. Its absence leads to immune dysfunction and reduced growth. Attempts at producing this amino acid through synthetic chemistry produces a racemic mixture of D- and L- methionine, which when fed to poultry has been linked to toxic side effects [37]. Engineering *S. cerevisiae* to overproduce methionine offers one possible solution to this problem. Methionine produced from biological sources would exist primarily as the L enantiomer. In order to do so using ALE in *S. cerevisiae*, a methionine biosensor was required.

The approach used in this project made use of the *S. cerevisiae MET17* promoter, which is expressed under conditions of low methionine, and repressed under conditions of high methionine (Figure 15). ALE experiments use different methods to apply selective pressure. The predominant method discussed here has been the expression of *eGFP* in response to a compound of interest, and subsequent sorting using FACs. However, other common responses to a compound of interest used in ALE are the expression of an auxotrophic complementation gene, or expression of a gene increasing growth rate [75]. In those situations, a biosensor which results in gene down-regulation cannot be used. As would be the case for a biosensor based on the *MET17* promoter. As such, two biosensors which invert the signal produced by p*MET17* were designed and named p*MET17-TETR-tetO*, and p*MET17-TETR-tetO*x7. The design of these sensors used p*MET17* to drive the expression of the transcriptional repressor *TETR*. TetRp should then bind to a variable length *tetO* sequence which has been placed between the *TEF1* promoter and an *eGFP* gene. As such, binding of TetRp to the *tetO* sequence should inhibit the ability of p*TEF1* to transcribe *eGFP*.

Under conditions of increased methionine concentrations, p*MET17* should stop transcription of *TETR*, allowing p*TEF1* to read through the *tetO* sequence and transcribe *eGFP*. The only difference between these sensors was how many repeats of the *tetO* sequence separated p*TEF1* from *eGFP*.

Assessment of both biosensors revealed that they cannot detect exogenously supplied methionine at concentrations that should have triggered the down-regulation of p*MET17* (Figure16) [42]. There were several possibilities that could explain why these sensors were un-responsive. One possibility was that TetRp was being produced during the pre-culture in Sd media and was still present in concentrations too high to allow p*TEF1* to read through the *tetO* sequence and transcribe *eGFP*. Another possibility was that even with the reduction in transcription of *TETR* via p*MET17*, basal transcription of the protein was still too high to allow de-repression at the *tetO* sites. However, both of these theories were shown to be unlikely after culturing biosensor containing cells in Sd media with doxycycline (Figure 17). As doxycycline inhibits the ability of TetRp to bind to the *tetO* sequence this demonstrated that the lack of *eGFP* expression in these sensors is likely TetR protein independent. Thus the failure of these sensors to produce eGFP in response to exogenous methionine is probably the result of the *tetO* sequences interfering with the ability of upstream p*TEF1* to induce transcription. The TetRp/*tetO* system has been for gene repression previously by introducing *tetO* binding sequences within the promoter itself [76]. That is, engineering a promoter to contain *tetO* sequences rather than placing them between the promoter and the reporter gene. This may be a necessary line of inquiry to produce a functional inverter based methionine biosensor in the future.

Due to the uncertainty about the feasibility of the inverted p*MET17* based biosensors, a simplified methionine sensor was also constructed. The p*MET17-eGFP* biosensor, used p*MET17* to drive the expression of *eGFP*, whilst simultaneously, a constitutively expressed *TEF1* promoter drove expression of mCherry. In response to increased methionine, this sensor would show a decrease in *eGFP* expression whilst maintaining mCherry. The constitutive expression of mCherry would allow normalisation of eGFP to generic protein expression levels. This would reduce the amount of false positives with low *eGFP* expression being identified as methionine over producers. As was mentioned above, the loss of signal

using this sensor is not ideal for use in all ALE based engineering strategies, but would be sufficient for a FACs based approach.

The p*MET17-eGFP* based sensor worked well, showing an 8.6 fold difference between induced and un-induced cultures after 11 hours (Figure 19). Furthermore, the preliminary assay of the linear range of the sensor has shown a probable maximum concentration detection somewhere below 0.2mM (Figure 20). However, this requires more investigation assaying a wider range of methionine concentrations.

There is a limitation in the applicability of p*MET17-eGFP* to ALE for methionine production. The *MET17* promoter is down regulated in response to methionine, but the manner in which this occurs is via conversion into the metabolite cysteine, which in turn triggers a regulatory response via $SCF^{Met30}$ [42]. The metabolites in the final steps of the sulfur network exist in an interconnected pool, where cysteine, methionine, S-adenosyl methionine, S-adenosyl homocysteine, homocysteine, and cystathionine are constantly being re-cycled and interconverted [38]. As such, increased methionine concentrations induce the transcriptional response seen in p*MET17* via their conversion to cysteine, and then through signalling by $SCF^{Met30}$. This means that it's possible a subset of mutations seen in ALE experiments may reduce the flux into the methionine branch of the pathway, re-routing it to cysteine. These would result in an increase in biosensor output but would be a false positive. However, as these two metabolites are only separated by 4 enzymatic steps the bulk of mutations which effect broader flux into/through the sulfur network would still result in increased methionine production [38]. This limitation means that during extended use of this biosensor for ALE some proportion of identified methionine overproducers will in fact be cysteine overproducers. These false positive would eventually be identified and removed during fermentation and quantification of methionine production levels using high performance liquid chromatography. This does not pose a great hindrance to this sensors use in ALE, but is an important consideration moving forward.

The p*MET17-eGFP* biosensor represents the first methionine biosensor that has been developed for use in *S. cerevisiae*. Although there are minor issues regarding cysteine activation, this sensor should be able to be utilised for ALE experiments and high throughput screening for the production of methionine in *S. cerevisiae*.

# 7: References


1. Becker J, Wittmann C: **Advanced Biotechnology: Metabolically Engineered Cells for the Bio-Based Production of Chemicals and Fuels, Materials, and Health-Care Products.** *Angewandte Chemie International Edition* 2015, **54:**3328-3350.
2. Chubukov V, Mukhopadhyay A, Petzold CJ, Keasling JD, Martín HG: **Synthetic and systems biology for microbial production of commodity chemicals.** *npj Systems Biology and Applications* 2016, **2**.
3. Clomburg JM, Crumbley AM, Gonzalez R: **Industrial biomanufacturing: The future of chemical production.** *Science* 2017, **355**.
4. Menezes AA, Montague MG, Cumbers J, Hogan JA, Arkin AP: **Grand challenges in space synthetic biology.** *Journal of The Royal Society Interface* 2015, **12**.
5. Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, Geertman JMA, van Dijken JP, Pronk JT, van Maris AJA: **Malic acid production by Saccharomyces cerevisiae: Engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export.** *Applied and Environmental Microbiology* 2008, **74:**2766-2777.
6. Saitoh S, Ishida N, Onishi T, Tokuhiro K, Nagamori E, Kitamoto K, Takahashi H: **Genetically engineered wine yeast produces a high concentration of L-lactic acid of extremely high optical purity.** *Applied and Environmental Microbiology* 2005, **71:**2789-2792.
7. Koopman F, Beekwilder J, Crimi B, van Houwelingen A, Hall RD, Bosch D, van Maris AJA, Pronk JT, Daran JM: **De novo production of the flavonoid naringenin in engineered Saccharomyces cerevisiae.** *Microbial Cell Factories* 2012, **11:**15.
8. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, et al: **High-level semi-synthetic production of the potent antimalarial artemisinin.** *Nature* 2013, **496:**528-532.
9. Nielsen J, Keasling Jay D: **Engineering Cellular Metabolism.** *Cell* 2016, **164:**1185-1197.
10. Keasling JD: **Manufacturing Molecules Through Metabolic Engineering.** *Science* 2010, **330:**1355-1358.
11. Krömer JO, Sorgenfrei O, Klopprogge K, Heinzle E, Wittmann C: **In-Depth Profiling of Lysine-Producing Corynebacterium glutamicum by Combined Analysis of the Transcriptome, Metabolome, and Fluxome.** *Journal of Bacteriology* 2004, **186:**1769-1784.
12. Georgi T, Rittmann D, Wendisch VF: **Lysine and glutamate production by Corynebacterium glutamicum on glucose, fructose and sucrose: roles of malic enzyme and fructose-1,6-bisphosphatase.** *Metab Eng* 2005, **7:**291-301.
13. Tong IT, Liao HH, Cameron DC: **1,3-Propanediol production by Escherichia coli expressing genes from the Klebsiella pneumoniae dha regulon.** *Applied and Environmental Microbiology* 1991, **57:**3541-3546.
14. Thykaer J, Christensen B, Nielsen J: **Metabolic network analysis of an adipoyl-7-ADCA-producing strain of Penicillium chrysogenum: elucidation of adipate degradation.** *Metab Eng* 2002, **4:**151-158.
15. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, Khandurina J, Trawick JD, Osterhout RE, Stephen R, et al: **Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol.** *Nat Chem Biol* 2011, **7:**445-452.
16. Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature* 2006, **440:**940-943.
17. Matsuda F, Ishii J, Kondo T, Ida K, Tezuka H, Kondo A: **Increased isobutanol production in Saccharomyces cerevisiae by eliminating competing pathways and resolving cofactor imbalance.** *Microbial Cell Factories* 2013, **12:**119-119.
18. Branduardi P, Longo V, Berterame NM, Rossi G, Porro D: **A novel pathway to produce butanol and isobutanol in Saccharomyces cerevisiae.** *Biotechnology for Biofuels* 2013, **6:**68.

19. Chen X, Nielsen KF, Borodina I, Kielland-Brandt MC, Karhumaa K: **Increased isobutanol production in Saccharomyces cerevisiae by overexpression of genes in valine metabolism.** *Biotechnology for Biofuels* 2011, **4:**21.
20. Portnoy VA, Bezdan D, Zengler K: **Adaptive laboratory evolution — harnessing the power of biology for metabolic engineering.** *Current Opinion in Biotechnology* 2011, **22:**590-594.
21. Chatterjee R, Yuan L: **Directed evolution of metabolic pathways.** *Trends in Biotechnology* 2006, **24:**28-38.
22. Portnoy VA, Herrgård MJ, Palsson BØ: **Aerobic Fermentation of d-Glucose by an Evolved Cytochrome Oxidase-Deficient Escherichia coli Strain.** *Applied and Environmental Microbiology* 2008, **74:**7561-7569.
23. Otero JM, Cimini D, Patil KR, Poulsen SG, Olsson L, Nielsen J: **Industrial systems biology of Saccharomyces cerevisiae enables novel succinic acid cell factory.** *PLoS One* 2013, **8:**e54144.
24. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO: **In silico design and adaptive evolution of Escherichia coli for production of lactic acid.** *Biotechnology and Bioengineering* 2005, **91:**643-648.
25. Williams TC, Pretorius IS, Paulsen IT: **Synthetic Evolution of Metabolic Productivity Using Biosensors.** *Trends Biotechnol* 2016, **34:**371-381.
26. Binder S, Schendzielorz G, Stäbler N, Krumbach K, Hoffmann K, Bott M, Eggeling L: **A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level.** *Genome Biology* 2012, **13:**R40.
27. Raman S, Rogers JK, Taylor ND, Church GM: **Evolution-guided optimization of biosynthetic pathways.** *Proc Natl Acad Sci U S A* 2014, **111:**17803-17808.
28. Stemmer WP: **DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution.** *Proc Natl Acad Sci U S A* 1994, **91:**10747-10751.
29. Coco WM, Levinson WE, Crist MJ, Hektor HJ, Darzins A, Pienkos PT, Squires CH, Monticello DJ: **DNA shuffling method for generating highly recombined genes and evolved enzymes.** *Nat Biotechnol* 2001, **19:**354-359.
30. Sieber V, Martinez CA, Arnold FH: **Libraries of hybrid proteins from distantly related sequences.** *Nat Biotechnol* 2001, **19:**456-460.
31. Shen Y, Stracquadanio G, Wang Y, Yang K, Mitchell LA, Xue Y, Cai Y, Chen T, Dymond JS, Kang K, et al: **SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes.** *Genome Res* 2016, **26:**36-49.
32. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, et al: **Total synthesis of a functional designer eukaryotic chromosome.** *Science* 2014, **344:**55-58.
33. Jankowski J, Kubińska M, Zduńczyk Z: **Nutritional and immunomodulatory function of methionine in poultry diets–a review.** *Annals of Animal Science* 2014, **14:**17-32.
34. Castellini C, Mugnai C, Dal Bosco A: **Effect of organic production system on broiler carcass and meat quality.** *Meat Science* 2002, **60:**219-225.
35. Mottet A, Tempio G: **Global poultry production: current state and future outlook and challenges.** *World's Poultry Science Journal* 2017, **73:**245-256.
36. Goulart CC, Costa FGP, Silva JHV, Souza JG, Rodrigues VP, Oliveira CFS: **Requirements of digestible methionine + cystine for broiler chickens at 1 to 42 days of age.** *Revista Brasileira de Zootecnia* 2011, **40:**797-803.
37. Kalbande V, Ravikanth K, Maini S, Rekhe D: **Methionine supplementation options in poultry.** *International Journal of Poultry Science* 2009, **8:**588-591.
38. Thomas D, Surdin-Kerjan Y: **Metabolism of sulfur amino acids in Saccharomyces cerevisiae.** *Microbiol Mol Biol Rev* 1997, **61:**503-532.

39. Cherest H, Davidian JC, Thomas D, Benes V, Ansorge W, Surdin-Kerjan Y: **Molecular Characterization of Two High Affinity Sulfate Transporters in Saccharomyces Cerevisiae.** *Genetics* 1997, **145:**627-635.
40. Ljungdahl PO, Daignan-Fornier B: **Regulation of amino acid, nucleotide, and phosphate metabolism in Saccharomyces cerevisiae.** *Genetics* 2012, **190:**885-929.
41. Lee TA, Jorgensen P, Bognar AL, Peyraud C, Thomas D, Tyers M: **Dissection of Combinatorial Control by the Met4 Transcriptional Complex.** *Molecular Biology of the Cell* 2010, **21:**456-469.
42. Menant A, Barbey R, Thomas D: **Substrate-mediated remodeling of methionine transport by multiple ubiquitin-dependent mechanisms in yeast cells.** *The EMBO Journal* 2006, **25:**4436-4447.
43. Crowley TJ: **Causes of Climate Change Over the Past 1000 Years.** *Science* 2000, **289:**270-277.
44. Shafiee S, Topal E: **When will fossil fuel reserves be diminished?** *Energy Policy* 2009, **37:**181-189.
45. Hong C, Tontonoz P: **Liver X receptors in lipid metabolism: opportunities for drug discovery.** *Nat Rev Drug Discov* 2014, **13:**433-444.
46. Rakopoulos DC, Rakopoulos CD, Giakoumis EG, Papagiannakis RG, Kyritsis DC: **Influence of properties of various common bio-fuels on the combustion and emission characteristics of high-speed DI (direct injection) diesel engine: Vegetable oil, bio-diesel, ethanol, n-butanol, diethyl ether.** *Energy* 2014, **73:**354-366.
47. Durre P: **Biobutanol: an attractive biofuel.** *Biotechnol J* 2007, **2:**1525-1534.
48. Krivoruchko A, Serrano-Amatriain C, Chen Y, Siewers V, Nielsen J: **Improving biobutanol production in engineered Saccharomyces cerevisiae by manipulation of acetyl-CoA metabolism.** *J Ind Microbiol Biotechnol* 2013, **40:**1051-1056.
49. Schadeweg V, Boles E: **n-Butanol production in Saccharomyces cerevisiae is limited by the availability of coenzyme A and cytosolic acetyl-CoA.** *Biotechnol Biofuels* 2016, **9:**44.
50. Si T, Luo Y, Xiao H, Zhao H: **Utilizing an endogenous pathway for 1-butanol production in Saccharomyces cerevisiae.** *Metab Eng* 2014, **22:**60-68.
51. Shi S, Si T, Liu Z, Zhang H, Ang EL, Zhao H: **Metabolic engineering of a synergistic pathway for n-butanol production in Saccharomyces cerevisiae.** 2016, **6:**25675.
52. Kurth EG, Doughty DM, Bottomley PJ, Arp DJ, Sayavedra-Soto LA: **Involvement of BmoR and BmoG in n-alkane metabolism in 'Pseudomonas butanovora'.** *Microbiology* 2008, **154:**139-147.
53. Dietrich JA, Shis DL, Alikhani A, Keasling JD: **Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis.** *ACS Synth Biol* 2013, **2:**47-58.
54. Winston F, Dollard C, Ricupero-Hovasse SL: **Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C.** *Yeast* 1995, **11:**53-55.
55. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics* 2012, **28:**1647-1649.
56. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CLV, Chandrasegaran S, Cai Y, et al: **Design of a synthetic yeast genome.** *Science* 2017, **355:**1040-1044.
57. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA, 3rd: **One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome.** *Proc Natl Acad Sci U S A* 2008, **105:**20404-20409.

58. Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, et al: **The Reference Genome Sequence of Saccharomyces cerevisiae: Then and Now.** *G3: Genes|Genomes|Genetics* 2014, **4:**389-398.
59. Peng B, Williams TC, Henry M, Nielsen LK, Vickers CE: **Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities.** *Microb Cell Fact* 2015, **14:**91.
60. Curran KA, Karim AS, Gupta A, Alper HS: **Use of High Capacity Terminators in Saccharomyces cerevisiae to Increase mRNA half-life and Improve Gene Expression Control for Metabolic Engineering Applications.** *Metabolic engineering* 2013, **19:**88-97.
61. Lee S, Lim WA, Thorn KS: **Improved Blue, Green, and Red Fluorescent Protein Tagging Vectors for S. cerevisiae.** *PLOS ONE* 2013, **8:**e67902.
62. Redden H, Alper HS: **The development and characterization of synthetic minimal yeast promoters.** *Nature Communications* 2015, **6:**7810.
63. Hodel AE, Harreman MT, Pulliam KF, Harben ME, Holmes JS, Hodel MR, Berland KM, Corbett AH: **Nuclear Localization Signal Receptor Affinity Correlates with in Vivo Localization in Saccharomyces cerevisiae.** *Journal of Biological Chemistry* 2006, **281:**23545-23556.
64. Robinson CR, Sauer RT: **Optimizing the stability of single-chain proteins by linker length and composition mutagenesis.** *Proceedings of the National Academy of Sciences* 1998, **95:**5929-5934.
65. Hall DB: **The VP16 Activation Domain Interacts with Multiple Transcriptional Components as Determined by Protein-Protein Cross-linking in Vivo.** *Journal of Biological Chemistry* 2002, **277:**46043-46050.
66. Garí E, Piedrafita L, Aldea M, Herrero E: **A Set of Vectors with a Tetracycline-Regulatable Promoter System for Modulated Gene Expression in Saccharomyces cerevisiae.** *Yeast* 1997, **13:**837-848.
67. Spiropoulos A, Bisson LF: **MET17 and Hydrogen Sulfide Formation inSaccharomyces cerevisiae.** *Applied and Environmental Microbiology* 2000, **66:**4421-4426.
68. Zhang W, Zhao G, Luo Z, Lin Y, Wang L, Guo Y, Wang A, Jiang S, Jiang Q, Gong J, et al: **Engineering the ribosomal DNA in a megabase synthetic chromosome.** *Science* 2017, **355**.
69. Xie Z-X, Li B-Z, Mitchell LA, Wu Y, Qi X, Jin Z, Jia B, Wang X, Zeng B-X, Liu H-M, et al: **"Perfect" designer chromosome V and behavior of a ring derivative.** *Science* 2017, **355**.
70. Wu Y, Li B-Z, Zhao M, Mitchell LA, Xie Z-X, Lin Q-H, Wang X, Xiao W-H, Wang Y, Zhou X, et al: **Bug mapping and fitness testing of chemically synthesized chromosome X.** *Science* 2017, **355**.
71. Shen Y, Wang Y, Chen T, Gao F, Gong J, Abramczyk D, Walker R, Zhao H, Chen S, Liu W, et al: **Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome.** *Science* 2017, **355**.
72. Mitchell LA, Wang A, Stracquadanio G, Kuang Z, Wang X, Yang K, Richardson S, Martin JA, Zhao Y, Walker R, et al: **Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond.** *Science* 2017, **355**.
73. Shi S, Choi YW, Zhao H, Tan MH, Ang EL: **Discovery and engineering of a 1-butanol biosensor in Saccharomyces cerevisiae.** *Bioresource Technology* 2017.
74. Skjoedt ML, Snoek T, Kildegaard KR, Arsovska D, Eichenberger M, Goedecke TJ, Rajkumar AS, Zhang J, Kristensen M, Lehka BJ, et al: **Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast.** *Nat Chem Biol* 2016, **12:**951-958.
75. Wang Q, Tang S-Y, Yang S: **Genetic biosensors for small-molecule products: Design and applications in high-throughput screening.** *Frontiers of Chemical Science and Engineering* 2017, **11:**15-26.
76. Blount BA, Weenink T, Vasylechko S, Ellis T: **Rational Diversification of a Promoter Providing Fine-Tuned Expression and Orthogonal Regulation for Synthetic Biology.** *PLOS ONE* 2012, **7:**e33279.

# 8: Supplementary Data

| Supplementary Table 1: Primers used in this project | |
|---|---|
| Name | Sequence (5'-3') |
| AC0001 | GGGGCCATCACCGTTTTGT |
| AC0002 | CCTCTTCAGCTGCATCAAGGA |
| AC0003 | TGCCCCTTACGAAATGAGGTC |
| AC0004 | GACGATATCTGAAGGGTGG |
| AC0005 | GTGTAGAACAGAGCCCGG |
| AC0006 | TGAGCATTTTGCTTACTGC |
| AC0007 | GGAAAGAGAAAAAACAGCTTAGC |
| AC0008 | TTCAAACCTCATTTTCTGGTACAT |
| AC0009 | ggtatcgataagcttgatatcgaattcctgcagcccGCACACACCATAGCTTCAAA |
| AC0010 | accatccaaatgcaacatTTGTAATTAAAACTTAGATTAGATTGCT |
| AC0011 | tctaagttttaattacaaATGTTGCATTTGGATGGTGAAGA |
| AC0012 | gaattcttgcatcttagaACCACCACCACCAGAACCAG |
| AC0013 | ggttctggtggtggtggtTCTAAGATGCAAGAATTCGCT |
| AC0014 | taaatcataagaaattcgTTAAACTTTTCTTTTTTTTTTTGGAGTACCA |
| AC0015 | aaaaaaagaaaagtttaaCGAATTTCTTATGATTTATGA |
| AC0016 | cgcggtggcggccgctctagaactagtggatccccGAGCGACCTCATGCTATACCT |
| AC0017 | tccactagttctagagcggccgccaccgcggtggagCCTCGGCGGACAGCG |
| AC0018 | agatgtaaatgaaaccatCCGGCGCCTCCACTCA |
| AC0019 | cgtgagtggaggcgccggATGGTTTCATTTACATCTTTATTAGCCG |
| AC0020 | gcaattaaccctcactaaagggaacaaaagctggagGGTAAATGGTTCCAAGGCCG |
| AC0021 | ggtatcgataagcttgatatcgaattcctgcagcccTTATTTTTTGCTTTTTCTCTTGAGGTCA |
| AC0022 | agatgtaaatgaaaccatTGTATGGATGGGGGTAATAGA |
| AC0023 | attacccccatccatacaATGGTTTCATTTACATCTTTATTAGCCG |
| AC0024 | cgcggtggcggccgctctagaactagtggatccccGGTAAATGGTTCCAAGGCCG |
| AC0025 | cttgtcagaaatgaacatTGTATGGATGGGGGTAATAGA |
| AC0026 | attacccccatccatacaATGTTCATTTCTGACAAGGTTTCT |
| AC0027 | taaatcataagaaattcgTTAAACTTTTCTTTTTTTTTTGGGTCGT |
| AC0028 | gcggccgccaccgcggtggagGCACACACCATAGCTTCAAA |
| AC0029 | actgatagggaTTGTAATTAAAACTTAGATTAGATTGCT |
| AC0030 | tttaattacaaTCCCTATCAGTGATAGAGAAA |
| AC0031 | aatgaaaccatGTGGTAAACTCGACTTTCACT |
| AC0032 | gagtttaccacATGGTTTCATTTACATCTTTATTAGCCG |
| AC0033 | agtttacaagttaattaaTTGTAATTAAAACTTAGATTAGATTGCT |
| AC0034 | tctaagttttaattacaaTTAATTAACTTGTAAACTCCCTATCAGTGA |
| AC0035 | agatgtaaatgaaaccatAGCTTGATTAGAATACTCGACTTTCAC |
| AC0036 | gagtattctaatcaagctATGGTTTCATTTACATCTTTATTAGCCG |
| AC0037 | agatgtaaatgaaaccatTAGATAATTACTTCCTTGATGATCTCCGG |
| AC0038 | caaggaagtaattatctaATGGTTTCATTTACATCTTTATTAGCCG |
| TETo sequence | TCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCAC |
| AC0039 | ATAGAGCACTCGATCTTCCC |
| AC0040 | TGGACGTTAATCACTTGCG |
| AC0041 | CATCACAAAAATCGACGC |
| AC0042 | GCCTGGTATCTTTATAGTCC |
| AC0043 | ttcttcacctttagacatTAGATAATTACTTCCTTGATGATCTCCGG |
| AC0044 | caaggaagtaattatctaATGTCTAAAGGTGAAGAATTATTCACTGG |
| AC0045 | ttcttcacctttagacatTGTATGGATGGGGGTAATAGA |

| | |
|---|---|
| AC0046 | attacccccatccatacaATGTCTAAAGGTGAAGAATTATTCACTGG |
| AC0047 | cctttagacatGTGGTAAACTCGACTTTCACT |
| AC0048 | gagtttaccacATGTCTAAAGGTGAAGAATTATTCACTGG |
| AC0049 | ttcttcacctttagacatAGCTTGATTAGAATACTCGACTTTCAC |
| AC0050 | gagtattctaatcaagctATGTCTAAAGGTGAAGAATTATTCACTGG |
| AC0051 | GCGATGGATAGTTCATTAGACG |
| AC0052 | TGATCGTTATCCAGTCGG |
| AC0053 | ACCTCGACAGCATGCAAGCTTGGTTTTAGAGCTAGAAATAGCAAGTTA |
| AC0054 | CCAAGCTTGCATGCTGTCGAGGTGATCATTTATCTTTCACTGCG |
| AC0055 | cttgtcagaaatgaacatACCACCACCACCAGAACCAG |
| AC0056 | ggttctggtggtggtggtATGTTCATTTCTGACAAGGTTTCT |
| AC0057 | aaagtcttatcaatctccTTAAACTTTTCTTTTTTTTTTTGGGTCGT |
| AC0058 | aaaaaaagaaaagtttaaGGAGATTGATAAGACTTTTCTAGTTGCA |
| AC0059 | cgcggtggcggccgctctagaactagtggatcccccCATTACAGATAGCGCCGATCA |
| AC0060 | TAAAGTAATACTTCTTCGTACG |
| AC0061 | TTAACACCAGAAATTCCAG |
| AC0062 | TTAACACCAGAAGGCAACGACCC |
| AC0063 | TAATCGTCGTCGAAGTCTCTATACG |
| AC0064 | TTAACACCAGAAATCAAGAACTC |
| AC0065 | TAAACGCTTGGTAATAGACGC |
| AC0066 | GAACAGACCAAGGACAGC |
| AC0067 | GTCTTTTTTTGTCGTTGTTCG |
| AC0068 | CGTCTATCATTACACGTATGC |
| AC0069 | ATAGACAAAGATAGCTTCGC |
| AC0070 | AGCGTGTAGGAAGGCTGGG |
| AC0071 | GTTTGGCTCGGGTTGTGCAG |
| AC0072 | GTCCAACAGGTGGTTACATTTTTTAGAGCTAGAAATAGCAAGTTA |
| AC0073 | AATGTAACCACCTGTTGGACGATCATTTATCTTTCACTGCG |
| AC0074 | AACACCTTTAAATCTGTCAATTTTAGAGCTAGAAATAGCAAGTTA |
| AC0075 | TTGACAGATTTAAAGGTGTTGATCATTTATCTTTCACTGCG |
| AC0076 | TTTGCTACCTATGTAACCATTTTTAGAGCTAGAAATAGCAAGTTA |
| AC0077 | ATGGTTACATAGGTAGCAAAGATCATTTATCTTTCACTGCG |
| AC0078 | tccactagttctagagcggccgccaccgcggtggagGCACACACCATAGCTTCAAA |
| AC0079 | ttcacccttagaaaccatTTGTAATTAAAACTTAGATTAGATTGCT |
| AC0080 | tctaagttttaattacaaATGGTTTCTAAGGGTGAAGAAGACA |
| AC0081 | taaatcataagaaattcgTTACTTGTACAATTCGTCCATACCAC |
| AC0082 | gacgaattgtacaagtaaCGAATTTCTTATGATTTATGA |
| AC0083 | gcaattaaccctcactaaagggaacaaaagctggagGAGCGACCTCATGCTATACCT |
| AC0084 | ttcttcacctttagacatTTGTAATTAAAACTTAGATTAGATTGCT |
| AC0085 | tctaagttttaattacaaATGTCTAAAGGTGAAGAATTAGT |
| AC0086 | tttaccgtaagtagcatcCTGAATAAATTATGGAAAGGAAATGT |
| AC0087 | tttccataatttattcagGATGCTACTTACGGTAAATTGACCT |
| AC0088 | cgcggtggcggccgctctagaactagtggatcccccTAAATGGTTCCAAGGCCGGC |
| AC0089 | tttaccgtaagtagcatcCTACAAGAATAATATACGAAACGT |
| AC0090 | cgtatattattcttgtagGATGCTACTTACGGTAAATTGACCT |
| AC0091 | cttagaaacttttcttttTTTTTTTGGCATTTGTAATTAAAACT |
| AC0092 | ttacaaatgccaaaaaaaAAAAGAAAAGTTTCTAAGATGCAAGA |
| AC0093 | agaagaaccagaagtaccAGTACCAATTCTAGATTGAGACCAGT |
| AC0094 | caatctagaattggtactGGTACTTCTGGTTCTTCTGGTTCT |
| AC0095 | taaatcataagaaattcgTTAACCACCATATTCATCAATACCCA |
| AC0096 | gatgaatatggtggttaaCGAATTTCTTATGATTTATGA |

| | |
|---|---|
| AC0097 | ggtatcgataagcttgatatcgaattcctgcagcccCCTCGGCGGACAGCG |
| AC0098 | cttagaaacttttctttttTAGATAATTACTTCCTTGATGATCTCCGG |
| AC0099 | caaggaagtaattatctaAAAAGAAAAGTTTCTAAGATGCAAGA |
| AC0100 | gaacataactttttctttttTTTTTTTGGCATTTGTAATTAAAACT |
| AC0101 | ttacaaatgccaaaaaaaAAAAGAAAAGTTATGTTCATTTCTGACA |
| AC0102 | agaagaaccagaagtaccGTCGTCACCCTTTCTTGGAC |
| AC0103 | ccaagaaagggtgacgacGGTACTTCTGGTTCTTCTGGTTCT |
| AC0104 | ACTAGTTCTAGAGCGGCCGCCACCGCGGTGGAGTTAATTAACTTGTAAACTCCCTATCAG |
| AC0105 | atattacaagttaattaaAGCTTGATTAGAATACTCGACTTTCAC |
| AC0106 | gagtattctaatcaagctTTAATTAACTTGTAATATTCTAATCAAGCT |
| AC0107 | acaaataaattttaaggtCAATTTACCGTAAGTAGCATCCTGA |
| AC0108 | gctacttacggtaaattgACCTTAAAATTTATTTGTACTACTGGT |
| AC0109 | AATCTAAGTTTTAATTACAAATGTCTAAAGGTGAAGAATTAGTATGTTATTTATTCTTCA |
| AC0110 | GTACAAATAAATTTTAAGGTCAATTTACCGTAAGTAGCATCCTGAATAAATTATGGAAAG |
| AC0111 | tctaagttttaattacaaATGTCTAAAGGTGAAGAATTATTCACTGG |

| Supplementary Table 2: Sequence of components used in biosensor construction | |
|---|---|
| *ADH1t* | **C**GAATTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAA ATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTC |
| *eGFP + CYC1T* | ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAATTAGATGGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAA GGTGAAGGTGATGCTACTTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATGGCCAACCTTAGTCACTACTTTC GGTTATGGTGTTCAATGTTTTGCGAGATACCCAGATCATATGAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAAGAAAG AACTATTTTTTTCAAAGATGACGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTTAGTTAATAGAATCGAATTAAAAGGTA TTGATTTTAAAGAAGATGGTAACATTTTAGGTCACAAATTGGAATACAACTATAACTCTCACAATGTTTACATCATGGCTGACAAACAAAAGAATG GTATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATTGGTGATGGTC CAGTCTTGTTACCAGACAACCATTACTTATCCACTCAATCTGCCTTATCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTTGTTAGAATTTG TTACTGCTGCTGGTATTACCCATGGTATGGATGAATTGTACAAATAAGGCGCGCCTTTTCCTTTGTCGATATCATGTAATTAGTTATGTCACGCTTA CATTCACGCCCTCCTCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTTAATAGTTATGTT AGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAAACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAG GTTTTGGACGCGCTTTGGCCGGCCTTGGAACCATTTACC |
| *pMET17* | TTATTTTTTGCTTTTTCTCTTGAGGTCACATGATCGCAAAATGGCAAATGGCACGTGAAGCTGTCGATATTGGGGAACTGTGGTGGTTGGCAAAT GACTAATTAAGTTAGTCAAGGCGCCATCCTCATGAAAACTGTGTAACATAATAACCGAAGTGTCGAAAAGGTGGCACCTTGTCCAATTGAACACG CTCGATGAAAAAAATAAGATATATATAAGGTTAAGTAAAGCGTCTGTTAGAAAGGAAGTTTTTCCTTTTTCTTGCTCTCTTGTCTTTTCATCTACTA TTTCCTTCGTGTAATACAGGGTCGTCAGATACATAGATACAATTCTATTACCCCCATCCATACA |
| *TETR* | ATGTTCATTTCTGACAAGGTTTCTTCTATGACTAAGTTGCAACCAAACACTGTTATTAGAGCTGCTTTGGACTTGTTGAACGAAGTTGGTGTTGAC GGTTTGACTACTAGAAAGTTGGCTGAAAGATTGGGTGTTCAACAACCAGCTTTGTACTGGCACTTCAGAAACAAGAGAGCTTTGTTGGACGCTTT GGCTGAAGCTATGTTGGCTGAAAACCACACTCACTCTGTTCCAAGAGCTGACGACGACTGGAGATCTTTCTTGATTGGTAACGCTAGATCTTTCA GACAAGCTTTGTTGGCTTACAGAGACGGTGCTAGAATTCACGCTGGTACTAGACCAGGTGCTCCACAAATGGAAACTGCTGACGCTCAATTGAG ATTCTTGTGTGAAGCTGGTTTCTCTGCTGGTGACGCTGTTAACGCTTTGATGACTATTTCTTACTTCACTGTTGGTGCTGTTTTGGAAGAACAAGCT GGTGACTCTGACGCTGGTGAAAGAGGTGGTACTGTTGAACAAGCTCCATTGTCTCCATTGTTGAGAGCTGCTATTGACGCTTTCGACGAAGCTG GTCCAGACGCTGCTTTCGAACAAGGTTTGGCTGTTATTGTTGACGGTTTGGCTAAGAGAAGATTGGTTGTTAGAAACGTTGAAGGTCCAAGAAA GGGTGACGAC |
| *tetO* | TCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCAC |

| | |
|---|---|
| *tetOx7* | TTAATTAACTTGTAAACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTATTCTAATCAAGCT |
| p*TEF1* | GCACACACCATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATCGCCGTACCACTTCAAAACACCCAAGCACAGCATACTAAATTTCCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAGAAAAAAGAGACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTGAAAATTTTTTTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAGTTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCATTTTTCTTGTTCTATTACAACTTTTTTTACTTCTTGCTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTAATTACAA |
| *VP16* AD | ATGTTGCATTTGGATGGTGAAGATGTTGCTATGGCTCATGCTGATGCTTTGGATGATTTTGATTTGGATATGTTGGGTGATGGTGATTCTCCAGGTCCAGGTTTTACTCCACATGATTCTGCTCCATATGGTGCTTTGGATATGGCTGATTTTGAATTTGAACAAATGTTTACTGATGCTTTGGGTATTGATGAATATGGTGGT |
| *SL7* Linker | GGTACTTCTGGTTCTTCTGGTTCTGGTTCTGGTGGTTCTGGTTCTGGTGGTGGTGGT |
| *BMOR* | TCTAAGATGCAAGAATTCGCTAGATTGGAAACTGTTGCTTCTATGAGAAGAGCTGTTTGGGACGGTAACGAATGTCAACCAGGTAAGGTTGCTGACGTTGTTTTGAGATCTTGGACTAGATGTAGAGCTGAAGGTGTTGTTCCAAACGCTAGACAAGAATTCGACCCAATTCCAAGAACTGCTTTGGACGAAACTGTTGAAGCTAAGAGAGCTTTGATTTTGGCTGCTGAACCAGTTGTTGACGCTTTGATGGAACAAATGAACGACGCTCCAAGAATGATTATTTTGAACGACGAAAGAGGTGTTGTTTTGTTGAACCAAGGTAACGACACTTTGTTGGAAGACGCTAGAAGAAGAGCTGTTAGAGTTGGTGTTTGTTGGGACGAACACGCTAGAGGTACTAACGCTATGGGTACTGCTTTGGCTGAAAGAAGACCAGTTGCTATTCACGGTGCTGAACACTACTTGGAATCTAACACTATTTTCACTTGTACTGCTGCTCCAATTTACGACCCATTCGGTGAATTCACTGGTATTTTGGACATTTCTGGTTACGCTGGTGACATGGGTCCAGTTCCAATTCCATTCGTTCAAATGGCTGTTCAATTCATTGAAAACCAATTGTTCAGACAAACTTTCGCTGACTGTATTTTGTTGCACTTCCACGTTAGACCAGACTTCGTTGGTACTATGAGAGAAGGTATTGCTGTTTTGTCTAGAGAAGGTACTATTGTTTCTATGAACAGAGCTGGTTTGAAGATTGCTGGTTTGAACTTGGAAGCTGTTGCTGACCACAGATTCGACTCTGTTTTCGACTTGAACTTCGGTGCTTTCTTGGACCACGTTAGACAATCTGCTTTCGGTTTGGTTAGAGTTTCTTTGTACGGTGGTGTTCAAGTTTACGCTAGAGTTGAACCAGGTTTGAGAGTTCCACCAAGACCAGCTGCTCACGCTAGACCACCAAGACCAGCTCCAAGACCATTGGACTCTTTGGACACTGGTGACGCTGCTGTTAGATTGGCTATTGACAGAGCTAGAAGAGCTATTGGTAGAAACTTGTCTATTTTGATTCAAGGTGAAACTGGTGCTGGTAAGGAAGTTTTCGCTAAGCACTTGCACGCTGAATCTCCAAGATCTAAGGGTCCATTCGTTGCTGTTAACTGTGCTGCTATTCCAGAAGGTTTGATTGAATCTGAATTGTTCGGTTACGAAGAAGGTGCTTTCACTGGTGGTAGAAGAAAGGGTAACATTGGTAAGGTTGCTCAAGCTCACGGTGGTACTTTGTTCTTGGACGAAATTGGTGACATGGCTCCAGGTTTGCAAACTAGATTGTTGAGAGTTTTGCAAGACAGAGCTGTTATGCCATTGGGTGGTAGAGAACCAATGCCAGTTGACATTGCTTTGGTTTGTGCTACTCACAGAAACTTGAGATCTTTGATTGCTCAAGGTCAATTCAGAGAAGACTTGTACTACAGATTGAACGGTTTGGCTATTTCTTTGCCACCATTGAGACAAAGATCTGACTTGGCTGCTTTGGTTAACCACATTTTGTTCCAATGTTGTGGTGGTGAACCACACTACTCTGTTTCTCCAGAAGTTATGACTTTGTTCAAGAGACACGCTTGGCCAGGTAACTTGAGACAATTGCACAACGTTTTGGACGCTGCTTTGGCTATGTTGGACGACGGTCACGTTATTGAACCACACCACTTGCCAGAAGACTTCGTTATGGAAGTTGACTCTGGTTTGAGACCAATTGAAGAAGACGGTTCTACTGCTGCTCACAGAGCTAGACAACCAGCTTCTGGTTCTGGTCCAGCTAAGAAGTTGCAAGACTTGGCTTTGGACGCTATTGAACAAGCTATTGAACAAAACGAAGGTAACATTTCTGTTGCTGCTAGACAATTGGGTGTTTCTAGAACTACTATTTACAGAAAGTTGAGACAATTGTCTCCAACTGGTTGTCACAGACCAGCTCACTGGTCTCAATCTAGAATTGGTACT |
| p*BMO* | CCTCGGCGGACAGCGCGGAAGATTGGAAACAGCCCGAGCGTGCGTGCCTCGGGCTGCATCCTTGCCACACCCAACCGGATTCGTCGGACCGCTCGACATTCGCGTTCGCTCCCGCGGCGCCGCGGGTGTACCGTTGCGTTACAGATGTACCCTTCTTTAACGTGTAACACACGCCTGGAGCGGCCAAGAGCCCCGCACCTTGCGGCGCGTCTTCCCCAGGGGCCCACCGGTTGCGGCCTTTTGCTGCGACCGTCCATGCTGGCACGACACTTGCTGAAAGCGTTAGAGCGGAATCGGTCCG |
| SCP | TTAATTAACTTGTAATATTCTAATCAAGCTTATAAAAGAGCACTGTTGGGCGTGAGTGGAGGCGCCGGAGATCATCAAGGAAGTAATTATCTA |

| | |
|---|---|
| mCherry | ATGGTTTCTAAGGGTGAAGAAGACAACATGGCTATTATTAAGGAATTCATGAGATTCAAGGTTCACATGGAAGGTTCTGTTAACGGTCACGAATT<br>CGAAATTGAAGGTGAAGGTGAAGGTAGACCATACGAAGGTACTCAAACTGCTAAGTTGAAGGTTACTAAGGGTGGTCCATTGCCATTCGCTTGG<br>GACATTTTGTCTCCACAATTCATGTACGGTTCTAAGGCTTACGTTAAGCACCCAGCTGACATTCCAGACTACTTGAAGTTGTCTTTCCCAGAAGGT<br>TTCAAGTGGGAAAGAGTTATGAACTTCGAAGACGGTGGTGTTGTTACTGTTACTCAAGACTCTTCTTTGCAAGACGGTGAATTCATTTACAAGGT<br>TAAGTTGAGAGGTACTAACTTCCCATCTGACGGTCCAGTTATGCAAAAGAAGACTATGGGTTGGGAAGCTTCTTCTGAAAGAATGTACCCAGAA<br>GACGGTGCTTTGAAGGGTGAAATTAAGCAAAGATTGAAGTTGAAGGACGGTGGTCACTACGACGCTGAAGTTAAGACTACTTACAAGGCTAAG<br>AAGCCAGTTCAATTGCCAGGTGCTTACAACGTTAACATTAAGTTGGACATTACTTCTCACAACGAAGACTACACTATTGTTGAACAATACGAAAG<br>AGCTGAAGGTAGACACTCTACTGGTGGTATGGACGAATTGTACAAGTAA |