The mycobacteriophage Bxb1 integrase as a catalyst for DNA inversion in *Saccharomyces cerevisiae*

Marshall Timmermans

A Thesis in the Department of Biology

Presented in partial fulfillment of the requirements for the Degree of Master of Science (Biology)

At Concordia University

Montreal, Quebec, Canada

Summer 2016

© Marshall Timmermans, 2016

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Marshall Timmermans

Entitled: The Mycobacteriophage Bxb1 Integrase as a Catalyst for DNA Inversion in *Saccharomyces cerevisiae*

And submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

Complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

______________________________________ Chair
Dr. Grant Brown

______________________________________ Examiner
Dr. Malcolm Whiteway

______________________________________ Examiner
Dr. David Walsh

______________________________________ Examiner
Dr. David Kwan

______________________________________ Supervisor
Dr. Vincent Martin

Approved by ____

Chair of Department of Graduate Program Director

____________________2016 ____________________________________

Dean of Faculty

ABSTRACT

“The mycobacteriophage Bxb1 integrase as a catalyst for DNA inversion in *Saccharomyces cerevisiae*”

Authored by Marshall Timmermans

Synthetic biology aims to develop methods of genetic regulation not seen in nature to manipulate biological systems, allowing them to achieve incredibly complex feats of biotechnology. One genetic element which has been receiving some attention by synthetic biologists is a class of enzymes known as serine recombinases. These enzymes have been shown to catalyze integration, excision, and inversion of DNA in a site-specific and directional manner. Due to the variety of activities of these enzymes and their irreversible mechanism of action, these enzymes could be used to construct gene regulation devices analogous in function to the digital memory of modern computers. This “genetic memory” could potentially offer improvements to the field of metabolic engineering and industrial biotechnology. This thesis aims to explore the use of DNA inversion catalyzed by the serine integrase from the BXB1 mycobacteriophage as a method of gene control in the brewer’s yeast *Saccharomyces cerevisiae.* Three assays were developed wherein a yeast-enhanced Green Fluorescent Protein (yeGFP) coding sequence could be inverted with respect to a promoter to determine the parameters of the BXB1 integrase inversion in yeast. The results from these assays show that this reporter system could be made to “activate” GFP fluorescence on a plasmid, but could not be found to inactivate GFP fluorescence on a plasmid or activate the chromosomally integrated GFP reporter. These results, call into serious question the

utility of the BXB1 serine integrase, and possibly serine integrases in general, as tools for synthetic biology in *Saccharomyces cerevisiae.*

Acknowledgements:

There are quite a lot of people who have loaned me their time, expertise and support over the course of the years leading up to this thesis, and I would like to make the first (of many, to be sure) repayment on those loans by acknowledging them here. To thank everyone who deserved it properly would expand this thesis's page count beyond readability, but here goes: Firstly, I would like to thank Dr. Vincent Martin for giving me a space in his research group in which to explore and experiment with molecular biology. His mentorship, instruction, and have expanded both my understanding of biology and challenged me to become a better biologist, and he has passed on to me a passion for natural products metabolism and metabolic engineering which I will carry with me in the years to come. I can't thank him enough for his guidance over the course of this thesis work. My colleagues in the Martin lab have also been invaluable for sharing their expertise, lab techniques, and camaraderie over the years. Special thanks are owed to Euan Burton, Lauren Narcross, Nick Gold, Elena Fossati, Kane Larue, Damien Biot-Pelletier, Zach Wiltshire, Shoham Mookerjee, Alex Campbell, Leanne Bourgeois, Audrey Morin, David Colatriano, Mindy Melgar, Andrew Wieczorek and James Scrivens for their friendship and support. A general "thank you" should also be extended to the entire Department of Biology at Concordia University. It has been a rare pleasure to dwell within an entire department of intensely knowledgeable scientists and all-around great people. Lastly, but most importantly, I would like to thank my parents, Eugene Timmermans and Judy Gatin. Without their genes I wouldn't be here and without their love and support I could not have done anything.

CONTENTS

Chapter 1: Introduction ... 1

1.1: Tools for modulating gene epxression in synthetic biology and metabolic engineering ... 1

1.1.1: Metabolic Flux and Inducible vs Constitutive Promoters ... 2

1.1.2: Balancing Metabolic Pathways by Combinatorial Assembly ... 3

1.1.3: Dynamic Control of Metabolic Flux Using Biosensors ... 5

1.1.4: Control of Metabolic Flux by Engineering Protein Stability ... 8

1.2: Engineering Genetic Memory ... 10

1.2.1: Integration and LSTRs ... 16

1.2.2: Excision ... 20

1.2.3: Inversion ... 23

1.3: Thesis Objective- The Mycobacteriophage BXB1 serine integrase as a regulator of gene expression in S. cerevisiae ... 26

Chapter 2: Materials and Methods ... 28

2.1: List of Plasmids, Oligos, and Strains ... 28

2.2: Polymerase Chain Reaction of Reporter Construct cassettes ... 31

2.3: Assembly of BXBInt Plasmids ... 37

2.4: Quantification of GFP fluorescence from TDH3YEGFP, YEGFPATTBP, and YEGFPATTRL ... 39

2.5: Testing the function of BXBint in S. cerevisiae via Flow Cytometry ... 40

2.6: Time Course ... 41

2.6: Integration of the Reporter Construct into yeast Chromosome14 ... 42

2.7: Fluorometry of Putative RCGFPattBP Integrants ... 48

Chapter 3: Results ... 48

3.1: Effect of Att sites on yeGFP expression levels ... 48

3.2: Serine Integrase Mediated Inversion of Plasmid DNA ... 50

3.3: FyeGFPattBP: Turning Off Gene Expression Using BXBInt ... 54

3.4: Integration of Reporter Construct Into Chromosome 14 ... 59

Chapter 4: DISCUSSION ... 63

4.1: Impact of Att Sites on yeGFP Expression ... 63

4.2: Activation of Expression By DNA Inversion ... 68

4.3: Deactivation of gene expression by BXBInt mediated Inversion of YEGFPATTBP ... 70

4.4: Integration of RCGFPattBP ... 75

Chapter 5: Conclusions ... 77

Activity of BXBInt in saccharomyces cerevisiae. 77
Restraints on BXBINT inversion.......... 78
Inversion of Chromosomal DNA 80
Overall Conclusions.......... 81
REFERENCES.......... 82

Table 1: List of Tables and Figures:

Figure 1-1: Hypothetical effects of genetic memory on metabolite production in microbial culture	15
Figure 1-2: Attachment (Att) sites of BXB1 Integrase (BXBInt)	22
Figure 1-3: Proposed mechanism of LSTR mediated integration of circularized DNA into a chromosome	23
Figure 1-4: Illustration of the two mechanisms LSTR mediated Excision	26
Figure 1-5: Excision mediated by BXBInt.	29
Figure 1-6: Inversion of DNA mediated by BXBInt	30
Table 2.1: List of plasmids	34
Table 2.2. Strains used in this study	35
Table 2.3: Oligonucleotides used in this Thesis	36
Figure 2-1: Schematic representation of DNA Assembler based cloning strategy	41
Figure 2-2: Schematic representations of the four constructs assembled to validate the activity of BXBInt in *Saccharomyces cerevisiae*	43
Figure 2-3: Schematic representation of the strategy implemented to integrate the RCGFPATTBP construct into Flagfeldt Site 16	51
Figure 2-4: PCR reactions used to map the integration cassette from purified genomic DNA of G418 resistant colonies	53
Figure 3-1: Flow cytometry of TDH3YEGFP, YEGFPATTBP, and YEGFPATTRL	55
Figure 3-2: Flow cytometry of BXBInt mediated inversion.	56
Figure 3-3: Time course of BXBInt inversion	59
Figure 3-4: Effectiveness of C-terminal tagged BXBInt variants	60
Figure 3-5: Long term expression of BXBInt mediated inversion	61
Figure 3-5: Fluorescence of YEGFPATTBP inversion.	62
Figure 3-6: Fluorescence of YEGFPATTBP after 24, 72, and 96 hrs of induction	63
Figure 3-6: Ethidium bromide stained agarose gel images of Site 16 integration PCR characterization	67
Figure 3-7: Fluorescence of F16RCGFPATTBP integrants + BXBHA	68
Figure 3-8: Fluorometry of integrated RCGFPATTBP cassette compared to plasmid borne RCGFPATTBP cassette.	69
Figure 4-1: RNA secondary structure of Att Sites	71

CHAPTER 1: INTRODUCTION

1.1: TOOLS FOR MODULATING GENE EXPRESSION IN SYNTHETIC BIOLOGY AND METABOLIC ENGINEERING

Synthetic biology is a relatively novel sub-discipline of biotechnology, which aims to rationally engineer gene expression in living systems to create novel emergent behaviours generating new patterns of gene expression based on chemical inputs and biological outputs.[1] Due to recent advances in technology related to DNA sequencing, synthesis, and cloning, the ability to rationally engineer complex biological systems that can be modelled with a high degree of predictability has advanced the field of biotechnology to levels unimagined even a few years ago. One area of synthetic biology focuses on refactoring existing biological signalling pathways to produce gene expression behaviours not found in nature. Stable oscillation of gene expression in response to chemical feedback loops,[2] "logic gate" or computational expression of multiple genes in response to various environmental stimuli,[3] and the optimization of transgenic multi-enzyme metabolic pathways through the differential transcript-level regulation of each step in that pathway[4] are all examples of such engineered systems.

Metabolic engineering is the practice of genetically engineering organisms to produce high quantities of enzymatically synthesized small molecules and macromolecules.[5] This practice, like more theoretical aspects of synthetic biology, is concerned with specifically and differentially expressing enzymes in order to direct the flow of intracellular metabolites towards a desired product. One common property between synthetic biology and metabolic engineering is the desire to control the expression of multiple genes simultaneously within a living organism. However, the

most complex of these synthetic gene regulatory systems have been developed in bacteria such as *Escherichia coli* and *Bacillus subtilis.*[6] The differences in structure and complexity between prokaryotic and eukaryotic gene regulation mean that advances in synthetic biology research focusing on prokaryotic organisms cannot necessarily be directly applied to eukaryotes, and the engineering of complex phenotypes in eukaryotes has lagged behind prokaryotic engineering in terms of sophistication.[6] Having a diverse toolbox of these genetic mechanisms applied to organisms more suitable for industrial fermentation, such as *Saccharomyces cerevisiae,* would provide additional tools for molecular biologists and genetic engineers to perform complex manipulations of living systems.

1.1.1: Metabolic Flux and Inducible vs Constitutive Promoters

Microbial production of important molecules has long been faced with a serious problem when it comes to expressing genes that are detrimental to the health of the microbe producing it. The strategy of overexpressing a gene product by regulating the gene under the power of a strong constitutive promoter may be suitable for the microbial production of many gene products, but if the gene product in question inhibits the growth of the microbe then constitutive expression of the gene product will make large-scale production of that molecule impractical. For example, heterologous expression of the *Xanthophyllomyces dendrorhous* carotenoid biosynthetic pathway in *S. cerevisiae* was found to induce cell membrane stress and triggered the pleiotropic drug resistance response, which slowed cell growth substantially compared to un-engineered *S. cerevisiae.*[7] Similarly, heterologous expression of the antimalarial drug precursor artemisinic acid pathway in yeast initially resulted in decreased stability of the plasmid

expressing this pathway, and induced oxidative and osmotic stress responses as well as the pleiotropic drug response in addition to reducing the cellular growth rate.[8] When growth of the microbe in fermentation is inhibited, the total yield of the desired product will be limited. The solution in the case of heterologous production of proteins in *E. coli* or yeast has been generally accepted to be to regulate the gene responsible for the toxic product under the power of an inducible promoter, such as the promoter from the *E. coli Lac* operon, or the yeast *GAL1* galactose-inducible promoter.[9] In this way cell division of the production microbe can occur unimpeded by the presence of the toxic gene product until the culture has reached a high density, after which expression of the toxic gene product can be induced and the quantity of gene product resulting from the fermentation can be increased compared to a constitutive expression strategy.

The inducible promoters used in these types of fermentation have their disadvantages. Chemicals such as Isopropyl β-D-1-thiogalactopyranoside (IPTG), a synthetic analogue of lactose used to induce the *Lac* promoter in *E. coli,* can be expensive when applied in large-scale bioreactors, raising the cost of large-scale fermentation of strains bearing *Lac*-regulated genes of interest. Promoters dependent on non-glucose carbohydrates, changes in pH, or changes in temperature may alter the growth of the production strain in undesirable ways[10] and may be difficult to implement in large scale bioreactors. Engineering gene expression strategies, which do not perturb the growth of the production strain, should therefore be a goal of genetic engineers seeking to maximize microbial production of potentially toxic gene products.

1.1.2: Balancing Metabolic Pathways by Combinatorial Assembly

One of the major approaches towards balancing metabolic flux through a pathway focuses on steady-state engineering of gene expression by combinatorial optimization of enzyme expression levels in order to characterize and optimize metabolite production. This is done by discovering or engineering libraries of constitutive promoters of varying strengths, by discovering or engineering variants of other genetic elements such as ribosome binding sites with different strengths, or by differentially controlling the copy number of DNA coding for metabolically active enzymes. Using these expression level variants libraries of strains expressing different levels of each enzyme in a metabolic pathway are generated that can be used to determine the optimal flux through that pathway.[11–13] This approach has been applied towards generating strains of *S. cerevisiae* expressing a heterologous xylose utilization pathway, directing flux of carbon from xylose towards ethanol production without compromising cell growth.[14] This combinatorial approach has also been successful in optimizing a 6-gene pathway for the production of farnesyl pyrophosphate (FPP) in *E. coli* through combinatorial engineering of ribosome binding sites of various strengths.[15] An approach towards combinatorial optimization of a heterologous mevalonic acid biosynthetic pathway in *S. cerevisiae* was capable of increasing production of amorpha-4,11, diene (a precursor for the antimalarial drug artemisinin) through integration of multiple copies of each biosynthetic gene in this pathway into the *S. cerevisiae* chromosome through δ-integration.[16] While there is a distinct advantage to this approach of combinatorial optimization, in that generating a random library can be achieved by transforming a strain with a mix of DNA elements to create a random library, there are also some disadvantages. Depending on the number of biosynthetic

genes that an experimenter may want to modulate, and depending on the number of promoter elements used to vary expression of each step, the libraries generated in these experiments can grow to be quite large, and the only way to determine whether the experiment has worked is to screen as many of the variants as possible. There is another disadvantage in that if the regulatory elements used to generate the combinatorial library are constitutive, there may be some combinations of expression levels that are lethal to the cell. These lethal variants may provide useful insight into the chemistry of the pathway, but obtaining direct information about these variants may not be simple due to the inherent bias in the library generation process. Engineering systems of genetic expression that are precisely controllable, rather than constitutive, would go a long way towards addressing this bias.

1.1.3: Dynamic Control of Metabolic Flux Using Biosensors

Another approach to metabolic optimization is to engineer metabolic control of a heterologous pathway through the use of biosensors and feedback loops.[17] Biosensor in this context refers to a protein or RNA molecule that is capable of specific allosteric binding to some small molecule metabolite, synthetic small molecule, or metal ion, and this binding instigates some change in the expression or activity of another gene or protein.[18] The principle behind this approach is that it is possible to discover or engineer a biosensor capable of recognizing some indicator of the concentration of specific metabolite within a cell. This biosensor would then either provide some quantitative measure of the metabolic flux towards the desired product such as through the generation of pigments or fluorogenic compounds allowing for high-throughput non-chemical screening of genetic variants of the desired biosynthetic pathway, or by

altering the expression of some point(s) of that pathway in order to increase flux toward the desired product.[19] This approach to metabolic engineering adds an element of temporal control over the expression of biosynthetic genes, allowing for a dynamic response to changes in the intracellular metabolite profile of the cell or changes in the extracellular environment. One illustrative example of this strategy is overproduction of the heterologous lycopene pathway in *E. coli* through incorporation of the *E. coli Ntr* regulon into the heterologous lycopene pathway.[20] This regulon in *E. coli* is capable of sensing the intracellular concentration of acetyl phosphate (ACP), which was seen as an indicator of glucose availability for pathways other than lycopene. By regulating the expression of two rate-limiting steps in the lycopene pathway under the power of the *glnAp2* promoter, a response element of the *Ntr* regulon, such that the expression of these lycopene synthetic enzymes would increase whenever concentrations of ACP increased above some threshold. This achieved a feedback loop between lycopene production and biomass accumulation capable of finding some balance between the fluxes of these two processes.

The application of biosensors to form the sort of “closed loop” feedback mechanisms, where a feedback mechanism is able to both sense a metabolite related to the metabolic process being modified and directly affect the expression of that pathway, has been limited in the years following the publishing of the *Ntr* lycopene biosensor.[18] Part of the difficulty in developing true closed loop biosensors is in generating a sensor that is able to directly affect expression of the pathway that it senses. So called “open loop” biosensors which allosterically bind to some inducer not related to the pathway being engineered and then affect some change in activity or

expression of that pathway have been identified as possible tools for specific modulation of the activities of multiple enzymes in a pathway, for example by controlling flux through the Embden-Meyerhof, Entner-Doudoroff, and pentose phosphate pathways in *E. coli* through control of the extracellular concentrations of iron, magnesium, and quorum sensing molecules.[21] The open loop approach to biosensors does not provide the sort of autonomous dynamic control provided by the closed loop example, but still enable experimenters to manually fine tune the expression of multiple steps in a biosynthetic process. Biosensors also show promise in enabling high-throughput screening of directed evolution and protein engineering experiments, which aim to increase flux through a pathway. Biosensors have been developed that sense some metabolite in a biosynthetic pathway and produce a fluorometric or colourimetric indication of that metabolite's concentration, thus allowing experimenters to use high throughput fluorometry or spectrophotometry in place of relatively low-throughput chromatographic or mass-spectrometry based screening in order to determine metabolite concentration.[22] This can involve generating whole cell sensors in the form of auxotrophic strains expressing a fluorescent protein, which will "sense" the presence of the molecule that they are auxotrophic for by continuing to grow and express the fluorescent protein,[23] or by identifying regulons that allosterically bind to a small molecule and directly induce expression of some reporter molecule.[24] One such example was the creation and utilization of an intracellular methionine and valine transcription factor based biosensor capable of inducing expression of green fluorescent protein (GFP), which enabled the imposition of artificial selection by fluorescence associated cell sorting (FACS) during a random mutagenesis directed evolution

experiment attempting to overproduce the essential amino acid L-valine.[25] A propionate biosensor coupled to the expression of GFP was also used to enable the heterologous expression of the CO_2-fixing 3-hydroxyproprionate pathway from *Chloroflexus aurantiacus* in *E. coli,* which required expression and mutagenesis of four metabolic pathways and several deletions in the host strain in order to succeed.[26] Biosensors needn't always involve allosteric binding of metabolites to some transcription effector protein in order to provide an indication of the metabolic state of a cell. By expressing a DOPA dioxygenase from the flowering plant *Mirabilis jalapa* in *S. cerevisiae,* which produces the purple pigment betaxanthin from DOPA, experimenters were able to create a biosensor indicating the flux of carbon towards the dopamine pathway.[27] This enabled the rapid characterization of around 200,000 tyrosine hydroxylase mutants by monitoring the appearance of the coloured pigment as a proxy for tyrosine hydroxylase activity and allowing for the downstream production of high titers of (*S*)-reticuline from a heterologous pathway in yeast.

In all cases, biosensors appear to offer significant advantages for metabolic engineering, although generating a specific biosensor is non-trivial and entirely depends on the chemical and biological context of the metabolite being detected, and so far the potential of closed-loop true feedback systems has yet to be realized. The ability to detect metabolites without having to subject cultures to low-throughput chemical analysis in order to determine the metabolic state of a cell is currently one of the most useful, and most used application of biosensors.

1.1.4: Control of Metabolic Flux by Engineering Protein Stability

In addition to transcriptional manipulation for metabolic optimization, it is also possible to control metabolic flux at the protein level by altering the stability of enzymes in a biosynthetic pathway. By increasing the half-life of weakly expressed or weakly functional enzymes in a pathway, or by reducing the half-life of problematic enzymes in a pathway, the overall abundance of that enzyme in the host cell can be manipulated to optimize metabolic flux. This can be achieved by creating fusion proteins of some metabolic enzyme tagged with peptide sequences, which will target that enzyme for degradation or prevent it from being degraded. Increasing the stability of a weakly functional enzyme, MIOX, in a heterologous three-enzyme α-D-glucaric acid pathway in *E. coli* by appending an N-terminal SUMO tag, which prevents ubiquitination, resulted in a twofold increase in the conversion of this product from glucose.[28] Reducing the stability of phosphofructokinase has been achieved in *E. coli* by generating fusions of phosphofructokinase with an SsrA tag, which can target that enzyme for degradation upon the induction of expression of the SsrB adaptor protein; the result being an inducible reduction in the rate of central metabolism and the diversion of carbon flux towards the *myo*-inositol synthesis pathway in a strain with a stable glucose-6-phosphate 1-dehydrogenase (*zwf*) knockout.[29] Engineering protein stability for metabolic engineering is a strategy that has been shown to be effective in these and numerous other contexts, but does suffer from the drawback of all protein fusion experiments in the effect of the fusion tag on the specific activity of the enzyme must be determined experimentally for each enzyme fusion.

1.2: ENGINEERING GENETIC MEMORY

One recent development in the field of synthetic biology may result in better induction strategies for large scale fermentation of microbes bearing engineered metabolic pathways, namely the concept of “Genetic Memory”.[3,30] The general concept of genetic memory is that expression of a gene is induced in a permanent, rather than transient, manner. The use of the term “memory” is an analogy to digital information storage, not cognitive or neurological memory, in that genetic expression can be bistable in the sense of being either on or off and remaining in that state until some new stimulus causes it to change states again. This concept of genetic memory could potentially have applications in the field of metabolic engineering complimentary to inducible systems of metabolic control, such as open loop biosensors. For large-scale fermentation of inducible gene products, traditional induction requires that the induction conditions must be maintained throughout the entire production phase of the fermentation. Inducers, which are also carbon sources for the microbe, such as galactose, must be continually added to the fermentation media as the microbe consumes it. Even for non-metabolized inducers such as IPTG or acyl homoserine lactones (AHL), they must be added continuously to the fermentation media to replace inducer, which may be chemically inactivated. These features of inducer kinetics may lead to inefficiencies in product formation, particularly in the case of small molecules rising from multi-enzyme bioconversion pathways where factors such as metabolic burden and redox imbalance already limit product formation.

In contrast to traditional inducible promoters, gene expression systems based on the concept of genetic memory would provide an irreversible change in expression

levels upon addition of an inducer which is maintained after the inducer is taken away. Figure 1-1 shows an example of a hypothetical yeast strain expressing some valuable product from a cassette regulated a traditional promoter compared to a strain with genetic memory.

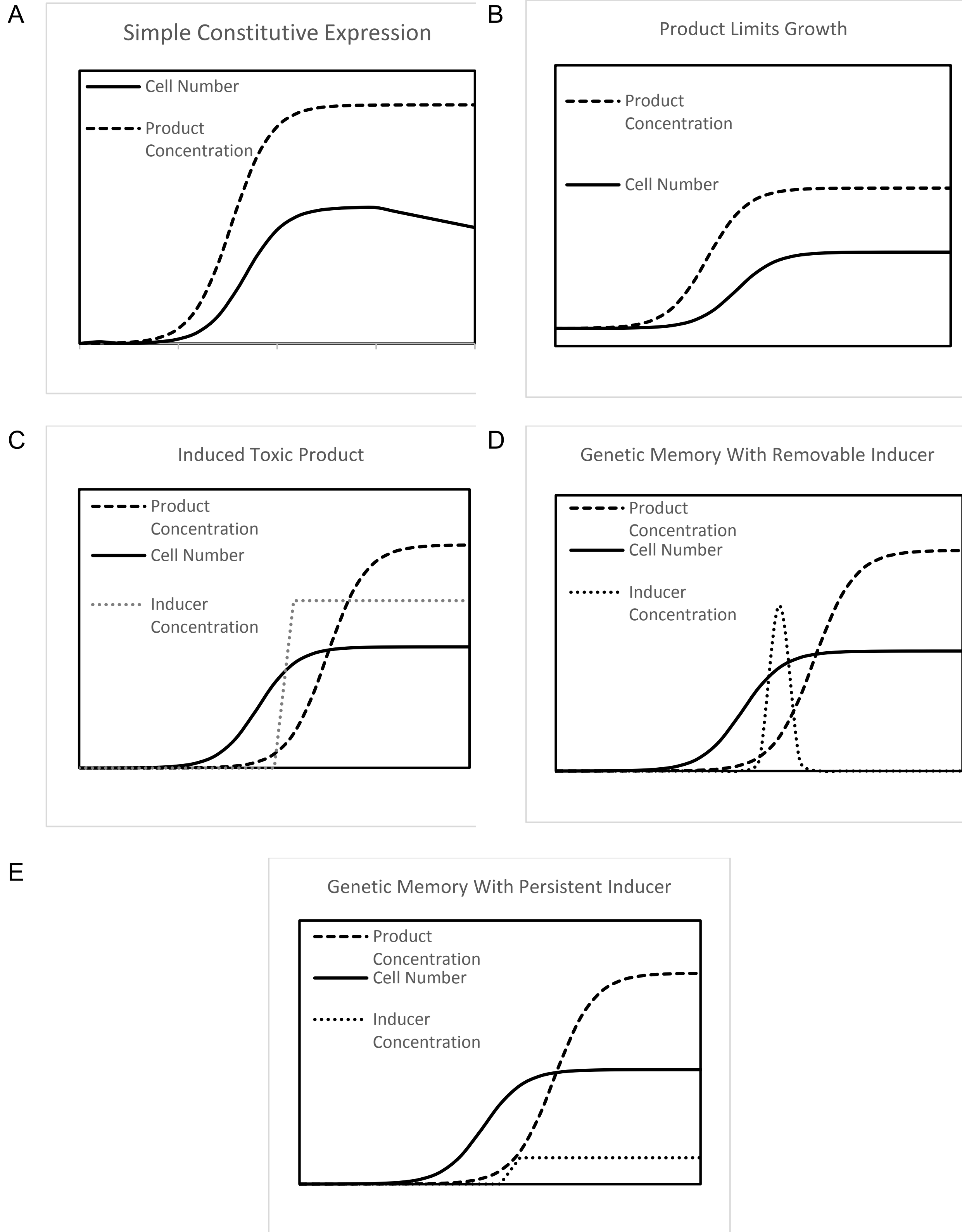


Figure 1-1: Hypothetical effects of Genetic Memory on metabolite production in microbial culture. A: Situation where some metabolite product is expressed constituitively and whose concentration increases in tandem with the cell number of the production strain. B: Case where the metabolite product is inhibitory or toxic to the cell, the growth rate of the cell is limited by the production of the product. C: Where that toxic product is regulated by an inducible promoter, cell growth can occur unimpeded, but decreases upon induction. Inducer concentration must be maintained throughout the production phase of the fermentation. D: If the product is regulated by an irreversible genetic memory system, a metabolizable inducer such as galactose can be pulsed in the fermentation media, without the need to constantly feed that inducer to the fermentaion. E: Non-metabolized inducers may only be

required in small quantities to create an induction effect equivalent to higher concentrations of inducer for a transient inducible promoter.

For the transient promoter, an inducer would have to be added continuously to the fermentation media throughout the course of production. In the case of the galactose-inducible *GAL1* promoter of *S. cerevisiae,* for example, this would mean a permanent shift towards using galactose as a carbon source exclusively, which may be undesirable. For an expression cassette with genetic memory, the switch to galactose need only be temporary before switching back to a different carbon source, and the effects of the galactose induction would persist in the absence of galactose. In the case of non-metabolized inducers, such as IPTG or AHLs, adding sufficient inducer to a large scale fermentation of a strain with a transient induction system may be costly. In a system with genetic memory, it is possible that far lower concentrations of inducer may result in the same effect, as the concentration of inducer would only need to cross a minimum threshold to instigate the genetic memory. If a genetic memory system could be broadly applied to any inducible promoter system, then inducible promoters could be used in situations where they may otherwise have been impractical.

Attempts have been made to engineer genetic memory in bacteria based on irreversible modification of DNA by Site Specific Recombination (SSR).[30–32] SSR is a widespread mechanism of DNA modification that serves many functions, such as a

component of the life cycle of many viruses,[33] reorganization of genomic information through transposition,[34] and altering expression of genes through structural DNA modification known as phase variation.[35] Generally speaking, SSR refers to directed, irreversible or selectively reversible recombination of double stranded DNA between two specific recombination sequences, catalyzed by a recombinase enzyme.

Recombinases can be divided into two main categories based on their catalytic residue and mechanism of molecular action, the tyrosine and the serine recombinases. Both classes of enzymes have cognate DNA binding sequences, although there is little sequence similarity between the binding sites of even closely-related recombinases.[36] There are three types of SSR activities found in nature: Integration, Excision, and Inversion of DNA. The Large Serine-Type Recombinase (LSTR) integrases are a subclass of serine integrases, of which the mycobacteriophage BXBInt protein is a member is the focus of this thesis. Many examples of LSTRs have been identified showing all three SSR activities to some degree.[31] The ability of the LSTP integrases to show multiple activities is based on the non-repetitive nature of the cognate DNA binding sequences of these proteins. Depending on the orientation of these attachment sequences and the presence of accessory proteins, DNA can be either integrated, excised, or inverted.[37] There have been around 80 LSTRs characterized so far and expression of these enzymes in *E. coli* have identified 13 including BXBInt, which show a high degree of orthogonality with regard to each other.[38] The variety and orthogonality of these enzymes have led to them being incorporated into complex computational genetic devices which modify *E. coli* plasmids.[31] While the utility of these computational

devices may currently be speculative, the use of LSTRs to alter gene expression both inducibly and irreversibly may have useful applications for metabolic engineering.

There are a number of features of serine integrases that would make them useful tools for molecular biology and biotechnology. Firstly, the activity of serine integrases is meant to be irreversible. Once recombination has been catalyzed there is little or no detectable reversal of that recombination, due to the fact that the DNA sequences to which the enzyme binds in order to catalyze recombination are not identical to each other and thus these sequences are destroyed during recombination. This could be considered a form of genetic memory, resulting in permanent modification of DNA from one state of inversion to another. Genetic memory could provide two main advantages over transient induction. Inducers such as alternative sugars, increased temperature, or changes in acidity can affect global gene expression patterns throughout the organism, which may alter how the microbe grows in undesirable ways. By allowing a gene expression cassette to switch states irreversibly in the presence of an inducer, the inducer may be removed while still allowing for the expression of the gene of interest and allowing for a more wild-type gene expression profile for the organism.

Secondly, the number of serine integrase family members, combined with the orthogonality of their activity, would allow experimenters to differentially regulate a number of genes in a complex metabolic pathway. Serine integrases are purported to catalyze inversion between DNA sequences known as Attachment (Att) sites named AttB and AttP sites as substrates regardless of what the sequence of the surrounding and intervening DNA is, so it is conceivable that a coding sequence could be “flipped on” while another coding sequence is simultaneously “flipped off”. This would require

only that the two coding sequences were cloned in a particular orientation with respect to a promoter, and that different serine integrases and their cognate Att sites regulated them. This type of system could potentially be useful in solving a number of problems unique to metabolic engineering, such as diverting metabolic flux towards a desired pathway by simultaneously upregulating the desired pathway and downregulating competing pathways, or by ablating expression of an enzyme resulting a toxic intermediate while simultaneously upregulating an enzyme that detoxifies that intermediate, or by the creation of “oscillating” patterns of gene expression where a heterologous pathway can be self-limiting through the expression of serine integrases and their cognate RDFs.

Thirdly, the ability of LSTR recombinases to catalyze recombination absent requirements for cofactors or host factors has made them popular for heterologous site-specific recombination. Frequently, LSTRs are monopartite enzymes that do not require host factors or multimerization in order to catalyze integration,[37] potentially making them ideal as tools for genetic manipulation. BXB1Int can catalyze chromosomal integration in a diversity of organisms such as human fibroblast cells,[39] *Drosophila,*[40] *Nicotiana,*[41] and wheat.[42] Interestingly, BXBInt can also catalyze unidirectional excision of tobacco chloroplast DNA,[43] *Arabidopsis* chromosomal DNA,[44] wheat chromosomal DNA,[42] and murine embryonic stem cell chromosomal DNA.[45] The successful heterologous expression of this enzyme in particular is encouraging for the use of this enzyme in *S. cerevisiae.*

1.2.1: Integration and LSTRs

Integration by SSR occurs between two molecules of dsDNA, where recombination between two attachment sites, cause the two molecules to be spliced into one. The sequence of the phage BXB1 attachment sites is shown in Figure 1-2 and a schematic representation of the integration activity of LSTRs is shown in Figure 1-3.

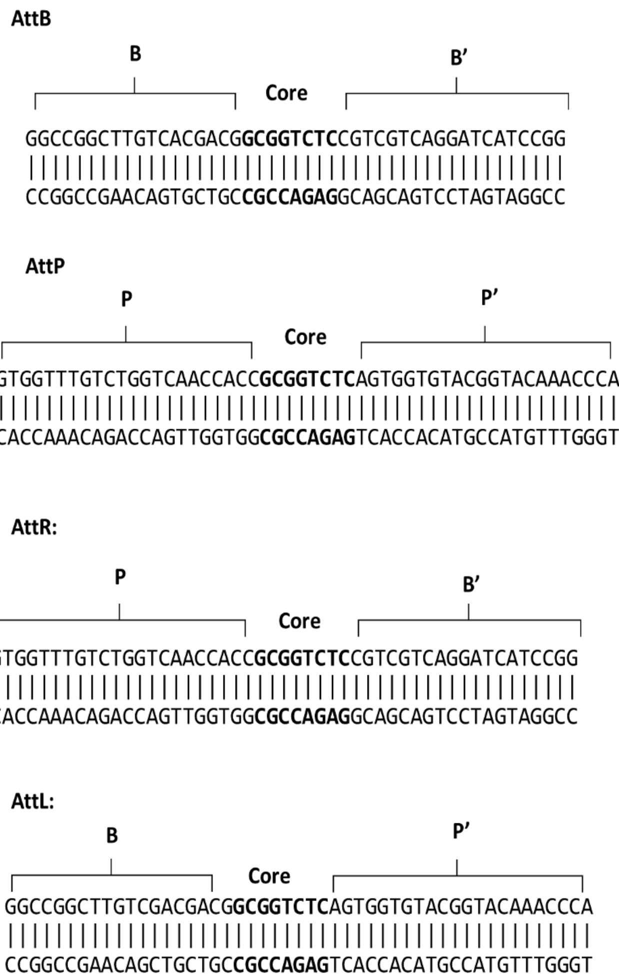


Figure 1-2: Attachment (Att) sites of BXB1 Integrase (BXBInt). The pre-recombination sites AttB and AttP share a core octonucleotide, each flanked by distinct palindromic sequences. Post recombination, the sites AttR and AttL are composed of elements of both AttB and AttP, but are distinct from the pre-recombination sites and are no longer palindromic. The structure of these sites contribute to the directionality of BXBInt recombination

Integration

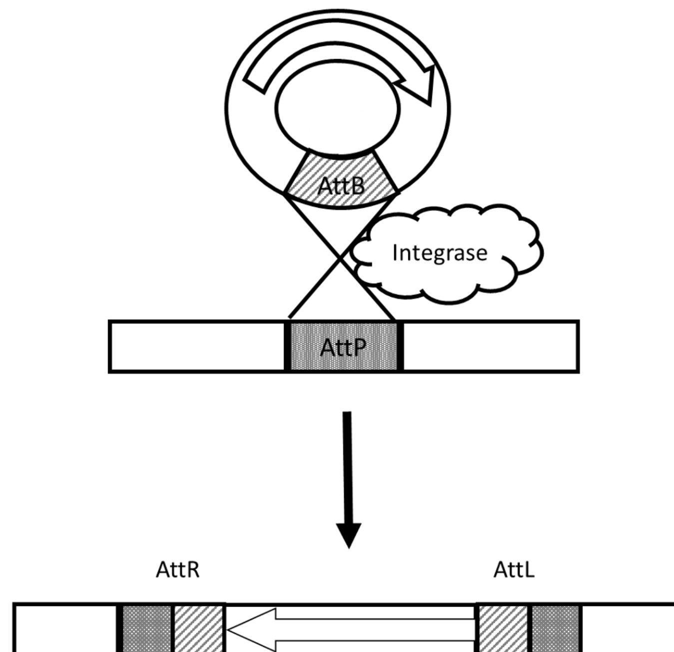


Figure 1-3: Proposed mechanism of LSTR mediated integration of circularized DNA into a chromosome. The integrase enzyme will specifically bind to its cognate binding sites, called Att sites on the plasmid (hashed lines) and chromosome (dark grey). Once bound to both sites the enzyme catalyzes recombination between strands, forming two new sequences composed of elements from each of the attachment sites and a single molecule of integrated DNA.

Many viruses use SSR as a method of integrating dsDNA copies of their genome into the chromosome of their host. For the LSTRs the principal components required for integration are the recombinase enzyme, and for two Att sites to exist on two separate molecules of DNA such as the phage genome and the genome of the host.[36] Of the

many examples of LSTR currently characterized, the INT protein from the mycobacteriophage BXB1 (BXBInt) shows particular promise as a genome engineering and synthetic biology tool.[39] BXBInt is a ~57 kDa monopartite serine recombinase of 500 amino acids of which the N-terminal 150 amino acids comprise the catalytic domain bearing the active serine residue.[46,47] BXBInt irreversibly and unidirectionally catalyzes recombination between two short (45-52) base pair DNA sequences known as AttP and AttB without the need for accessory proteins, host proteins, high-energy cofactors, divalent metals, or supercoiled DNA[48,49]. As diagrammed in Figure 1-3, this enzyme will bind to AttP and AttB sites on two molecules of DNA, and then catalyzes strand exchange between these two sites to splice them together. BXBInt is highly specific to its cognate Att sites, as even single point mutations at almost any nucleotide position of either Att site will reduce binding of the enzyme to the DNA substrate and inhibit synapsis.[50] The cognate Att sites for BXBInt are short compared to the recombination sites of some other serine and tyrosine recombinases. AttB is 45 base pairs and AttP is 52 base pairs. These sites share a non-palindromic core sequence of 8 nucleotides which is the location of strand exchange during synapsis, and which are flanked by 6 base pair inverted repeats.[48] Outside of these inverted repeats both AttB and AttP are non-symmetrical and non-palindromic and thus have a distinct orientation, although integration occurs regardless of the orientation of the sites.[48] The asymmetric nature of the AttB and AttP sites also contribute to the directionality of recombination activity. Post-recombination, the AttB and AttP sites will resolve into two distinct sequences, called AttR and AttL, which share the common core sequence of 8 base pairs of the pre-recombination sites, but whose inverted repeats and external sequences will be a

hybrid of both pre-recombination sites. The specificity of sequence recognition and proofreading of BXBInt is such that recombination will not be catalyzed between the AttR and AttL sites to any measurable degree.[49] The irreversible nature of BXBInt mediated recombination is sometimes referred to as highly directional recombination, a property which is shared by some, although notably not all, LSTRs.[36]

1.2.2: Excision

Excision is in some regards the reverse process of integration. Two recombination sites on one strand of DNA recombine to remove the intervening DNA sequence from the whole, resulting in two molecules of DNA. Excision, like integration, is an essential part of the life cycle of many viruses, particularly lysogenic bacteriophages, which will excise their dsDNA genome in preparation for encapsidation. Excision also function in many other contexts, such as transposons in bacterial and eukaryotic genomes which show excision and integration SSR independent of viral replication.[34] The mechanism by which LSTRs catalyze excision is similar in many respects to the mechanism of inversion, and understanding the mechanism of excision may give insights to the process of inversion, which is the focus of this thesis.

Excision mediated by LSTRs such as BXBInt can occur in two ways, both of which are shown in Figure 1-4.

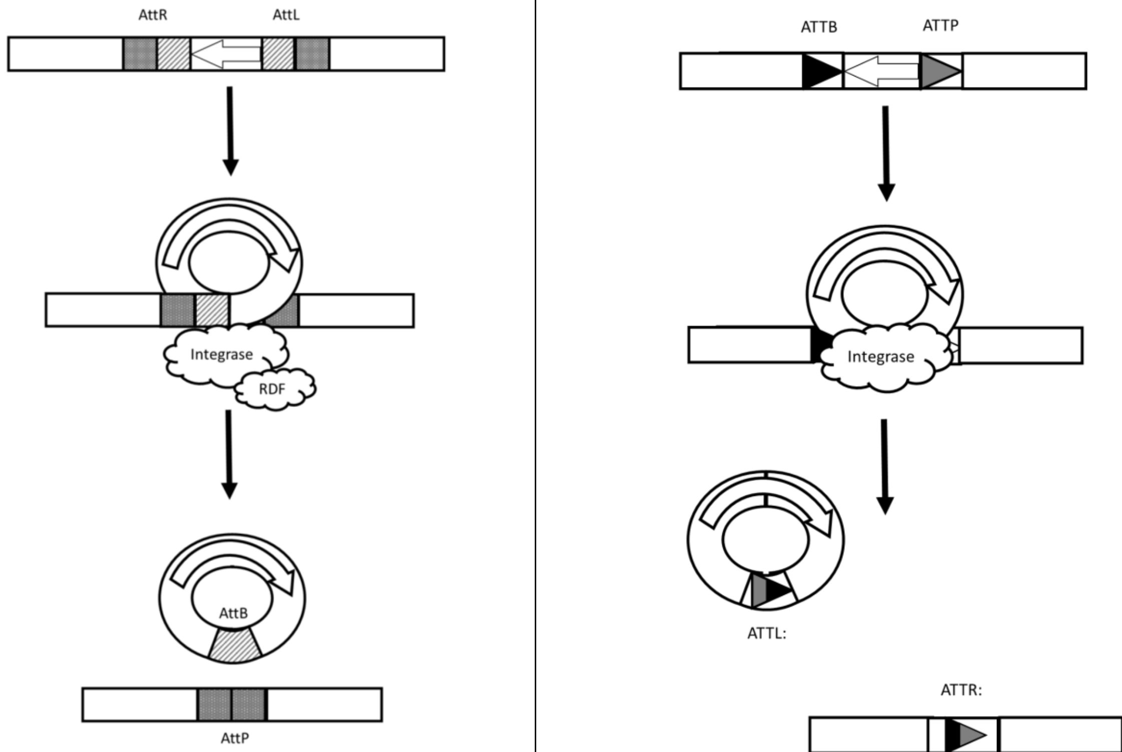


Figure 1-4: Illustration of the two mechanisms LSTR mediated Excision. RDF mediated excision is shown on the left. Two post-integration sites, AttR and AttL are shown as grey boxes. The integrase will bind to AttR and AttL, forming a loop of the intervening DNA but not catalyzing recombination. The RDF protein will then bind to the nucleoprotein complex formed by the Integrase, AttR and AttL. Once bound, the RDF will alter the function of the integrase to enable it to catalyze recombination between AttR and AttL, resulting in excision of the intervening DNA.

RDF independent excision is shown on the right. The two pre-recombination sites AttB and AttP are shown as grey triangles to indicate their orientation with respect to each other. When AttB and AttP are in this parallel orientation, Excision can occur without an RDF. This mechanism of excision also depends on the formation of a loop when the integrase binds to AttB and AttP, after which recombination occurs via a similar mechanism to integration, resolving in two molecules of DNA.

The mechanism of excision used during the BXB1 phage life cycle occurs after integration between AttB and AttP has occurred, and uses the post-recombination sequences AttR and AttL as substrates. When both AttR and AttL are on the same molecule of DNA, BXBInt will bind to these sites in conjunction with an accessory protein known generally as a Recombination Directionality Factor (RDF), which in the case of BXB1 is named gp47. This RDF binds to the BXBInt / Att site DNA complex,

stimulating strand exchange in BXBInt / AttR + AttL complexes and inhibiting integration.[51] The recombination between AttR and AttL will reconstruct the AttP and AttB sites.[50] It should be noted that in order for recombination to occur between AttR and AttL sites on one molecule of DNA, the intervening DNA must form a loop in order to bring AttR and AttL into proximity with BXBInt, although it is not clear as to whether this loop forms stochastically, whether BXBInt mediates the formation of this loop, or whether the RDF mediates the formation of this loop.

The other method by which BXBInt can mediate excision is by orientating AttB and AttP into the same molecule of DNA with a certain orientation with respect to each other. AttB and AttP are non-palindromic sequences, with a distinct orientation based on their core 8-nucleotide sequence. Like many members of the LSTR recombinases, when BXBInt AttB and AttP sites are in a "parallel" orientation with respect to each other on one molecule of DNA, BXBInt mediated recombination will occur between them, resolving in excision of the intervening DNA.[52] This excision, like RDF mediated excision, requires the DNA between Att sites to form a loop in order for the Att sites to come into contact with BXBInt and for recombination to occur. This form of excision is unidirectional and will result in two molecules of DNA, the original molecule bearing either AttR or AttL, and a circular excision fragment bearing the other post-recombination site. This excision activity of BXBInt was first tested in the fission yeast *Schizosaccharomyces pombe* using both plasmid based and chromosomally integrated AttP/AttB substrates, and this activity was found to be somewhat less efficient using chromosomal substrates, although other LSTR family members were much more efficient at chromosomal excision.[52]

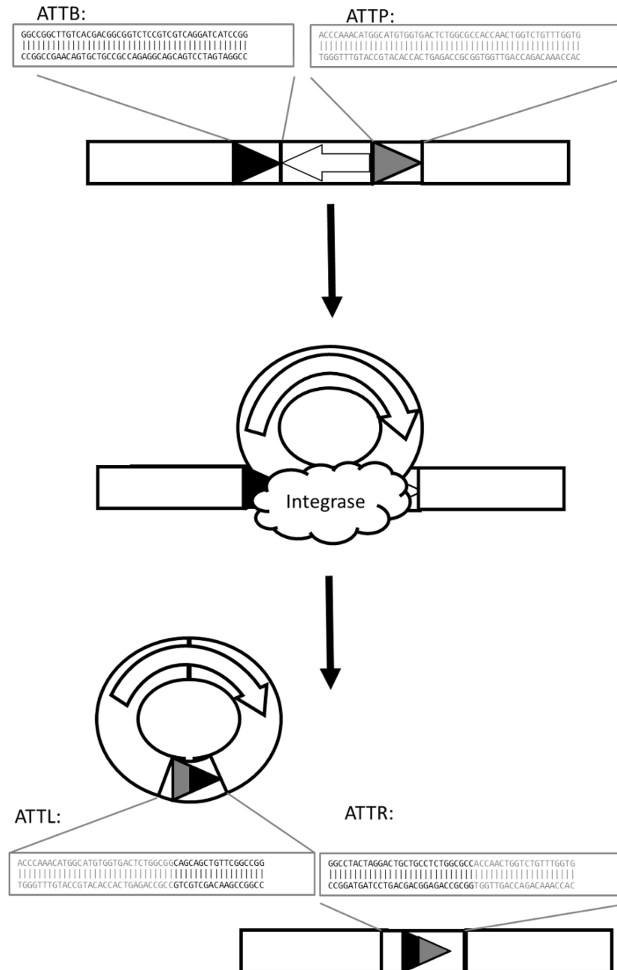


Figure 1-5: Excision mediated by BXBInt. The integrase enzyme binds to the attachment sites AttB and AttP, stabilizing the formation of a loop in the DNA. This protein/DNA complex is known as an invertosome. Once the invertosome is formed, strand exchange is mediated by the recombination. When the Att sites are in the parallel orientation (as shown by black and grey arrows), the result will be the excision of DNA between the two attachment sites.

1.2.3: Inversion

The third mechanism of SSR is Inversion, or “flipping” of DNA. DNA inversion is observed in nature as exemplified by the Hin recombinase protein of *Salmonella*

enterica. Hin recombinase functions to invert a ~900 bp region of the *Salmonella* chromosome which acts as a promoter to either one or two ORFs responsible for flagellar protein synthesis.[53] The mechanism of the Hin recombinase is thought to be representative of the mechanism of many serine recombinases: Hin-mediated DNA inversion involves a tetramer of the Hin protein, two dimers of an enhancer protein Fis, and a DNA-bending protein HU. These proteins work in concert to form a "loop" in the DNA strand known as an invertosome prior to catalyzing inversion[54]. The formation of the invertosome complex is thought to be similar to the mechanism of inversion catalyzed by other serine recombinases.

The mechanism for BXBInt mediated inversion is similar to that of non-RDF mediated excision. Excision and inversion activities for BXBInt are distinct from each other in the orientation of the AttB and AttP sites. Where excision occurs when AttB and AttP are in a "parallel" orientation with respect to each other on the same molecule of DNA, inversion occurs where the two sites are in an "antiparallel" orientation with respect to each other on the same molecule of DNA as shown in Figures 1.5 and 1.6:

Otherwise, the mechanism of BXB1Int inversion is similar to that of LSTR excision, where DNA forms an invertosome complex, bringing AttP and AttB into proximity of BXBInt which facilitates strand exchange between the two sites. This inversion results in a DNA sequence that has been inverted, and the two attachment sites resolve into distinct post-recombination sequences known as AttR and AttL.[48] Much like inversion and excision, this activity of BXBInt is highly directional, and may be reversed by co-expressing the RDF gp47 with BXBInt.

Inversion

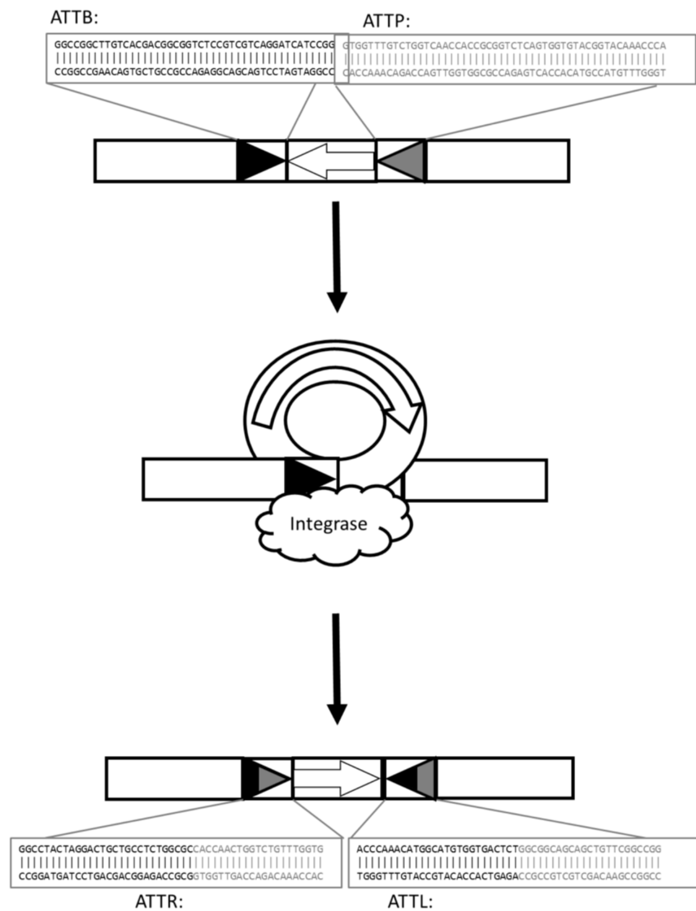


Figure 1-6: Inversion of DNA mediated by BXBInt. Much like the process of excision, the integrase binds to the two Att sites to form an invertosome. When the Att sites are in the antiparallel orientation, the result of strand exchange in the invertosome will be inversion of the DNA between the Att sites.

The inversion activity of LSTRs such as BXBInt has been a focus of interest for synthetic biologists.[30,31,55] This interest is due in part to the fact that inversion does not commonly occur with the more commonly used tyrosine recombinases, although site-directed mutagenesis of the *Cre* recombinase *LoxP* sites has been undertaken to

generate a small number of *LoxP* variants that favour inversion in mice.[56] In contrast to the tyrosine recombinases, over 80 LSTRs have been identified, many of which are known or suspected to catalyze inversion in addition to integration and excision. The demonstrated ability of these enzymes to catalyze DNA inversion *in vivo,* the orthogonality of attachment site binding of LSTRs, as well as the large number of characterized enzymes, leads to the possibility of performing multiple separate DNA inversion events in living cells.

1.3: THESIS OBJECTIVE

The main impetus of this thesis is to develop the use of serine integrases in yeast to catalyze recombination. These enzymes have a number of properties that may be well suited for novel methods of synthetic gene expression. Currently, there are no characterized systems for differential and orthogonal regulation of multiple transgenes in *S. cerevisiae.* Induction systems that exhibit these properties would be valuable tools for the discipline of metabolic engineering that frequently makes use of simultaneous up and downregulation of native and heterologous genes in order to divert carbon flux. The serine integrase/Att site system of re-orienting a coding sequence with respect to a promoter element would have both of these properties, as the ability to activate expression of a gene or de-activate expression of a gene depends only on the starting state of the coding sequence, and ideally should be unbiased with respect to the DNA between the Att sites and with respect to the ability to activate or de-activate gene expression.

The work presented in this thesis is an attempt to create a system of gene expression control using the inversion property of the serine integrase from BXB1, referred to as BXBInt or BXB integrase. The experiments presented will attempt to determine the parameters under which this enzyme is able to catalyze inversion in *S. cerevisiae.* Firstly, the effects of interpolating the BXBInt Att sites between the ORF of a yeast enhanced green fluorescent protein (yeGFP) ORF and the promoter and terminator of an expression cassette were quantified. To determine what conditions inversion occurs, three types of inversion targets were constructed in order to test the capabilities of BXBInt to catalyze recombination in situations, which may be applicable to other gene manipulation experiments. The first is the case of a plasmid-borne yeGFP ORF, which has been cloned in a reverse-compliment orientation with respect to a promoter and flanked by the BXBInt attachment sites AttB and AttP. This yeGFP ORF will rely on BXBInt to invert the ORF and bring it into an orientation that will allow for the correct transcription of the ORF. Successful inversion of the plasmid based reporter construct will be assayed by the expression of yeGFP only when BXBInt is expressed. Another construct to assay for BXBInt activity is composed of a plasmid borne yeGFP ORF cloned in the transcriptionally permissive "forward" orientation with respect to a promoter and flanked by BXB AttB and AttP. In this case, expression of BXBInt can be used to turn off yeGFP expression. The third inversion target will be integrated into a yeast chromosome rather than being localized on a plasmid. The testing for inversion activity of these variants of the typical BXBInt substrate will provide evidence for or against the suitability of BXBInt as a versatile tool for genome manipulation and genetic engineering.

CHAPTER 2: MATERIALS AND METHODS

2.1: *LIST OF PLASMIDS, OLIGONUCLEOTIDES, AND STRAINS*

Table 2.1: List of plasmids

Plasmid	Genotype	Description	Source
pGC966	CEN6/ARS4^{ori}, pMB1^{ori}, LEU2, Amp^{R}, loxP-Kan^{R}, PGAL1-HISstuffer-T$_{CYC1}$, ΔKpnI$^{(3555-2560)}$A(3558)G, ΔKpnI$^{(4509-4514)}$ A(4512)G	CEN plasmid with LEU2 marker, protein expression under control of *GAL1* promoter, with deleted *KpnI* site and 506 homology cloning linkers.	[57]
pGC967	CEN6/ARS4^{ori}, pMB1^{ori}, URA3, Amp^{R}, loxP-Kan^{R}, PGAL1-HISstuffer-T$_{CYC1}$, ΔKpnI$^{(3593-3598)}$A(3596)G	CEN plasmid with URA3 marker, protein expression under control of *GAL1* promoter, with deleted *KpnI* site and 506 homology cloning linkers.	[57]
pTDH3YEGFP	pGC966 with forward-sense yeast-enhanced Green Fluorescent Protein	Yeast Enhanced Green Fluorescent Protein (yeGFP) regulated by the strong constitutive promoter *TDH3*	This thesis
pYEGFPATTBP	pGC966 with forward-sense yeast-enhanced Green Fluorescent Protein and BXB1 Int Att B and AttP sites	yeGFP with the BXBInt AttB site interposed between the *TDH3* promoter and the yeGFP ORF, and the AttP site interposed between yeGFP ORF and *ENO2* terminator.	This Thesis
pYEGFPATTRL	pGC966 with forward-sense yeast-enhanced Green Fluorescent Protein and BXB1 Int Att R and Att L sites	Identical to pYEGFPATTBP, but with the BXBInt post-recombination sites AttR and AttL interposed between the ORF and promoter/terminator	This Thesis
pRCGFPATTBP	pGC966 with reverse compliment-sense yeast-enhanced Green Fluorescent Protein and BXB1 Int Att B and AttP sites	Identical to pYEGFPATTBP, but with the yeGFP ORF cloned in the reverse compliment orientation with respect to the promoter	This Thesis
pBXBHA	pGC967; *GAL1*p; BXBInt:HA	pGC967 harbouring the integrase from BXB1 under the power of the galactose-inducible *GAL1* promoter, with an N-terminal HA tag	This thesis
pBXBTy1	pGC967; *GAL1*p; BXBInt:Ty1NLS	pGC967 harbouring the integrase from BXB1 under the power of the galactose-inducible *GAL1* promoter, with an N-terminal Ty1 NLS tag	This Thesis
pBXBSV40	pGC967; *GAL1*p; BXBInt:SV40NLS	pGC967 harbouring the integrase from BXB1 under the power of the galactose-inducible *GAL1* promoter, with an N-terminal SV40 NLS tag	This Thesis

Table 2.2. Strains used in this thesis

Strains	Genotype	Source

CEN.PK113-16B	*MATα leu2-3 MAL2-8C SUC2*	EUROSCARF (Frankfurt, Germany)
CEN.PK113-13D	*MATα ura3-52 MAL2-8C SUC2*	EUROSCARF (Frankfurt, Germany)
CEN.PK113-17A	*MATα ura3-52 leu2-3,112 MAL2-8C SUC2*	EUROSCARF (Frankfurt, Germany)
TDH3YEGFP	CEN.PK113-16B harbouring pTDH3YEGFP	This thesis
YEGFPATTBP	CEN.PK113-16B harbouring pYEGFPATTBP	This thesis
YEGFPATTRL	CEN.PK113-16B harbouring pYEGFPATTRL	This thesis
RCGFPATTBP	CEN.PK113-16B harbouring pRCGFPATTBP	This thesis
BXBHA	CEN.PK113-13D harbouring pBXBHA	This thesis
BXBTY1	CEN.PK113-13D harbouring pBXBTY1	This thesis
BXBSV40	CEN.PK113-13D harbouring pBXBSV40	This thesis
BXBHA_ RCGFPATTBP	CEN.PK113-17A harbouring pBXBHA and pRCGFPATTBP	This thesis
BXBTY1_ RCGFPATTBP	CEN.PK113-17A harbouring pBXBTY1 and pRCGFPATTBP	This thesis
BXBSV40_ RCGFPATTBP	CEN.PK113-17A harbouring pBXBSV40 and pRCGFPATTBP	This thesis
BXBHA_ YEGFPATTBP	CEN.PK113-17A harbouring pBXBHA and pYEGFPATTBP	This thesis
F16TDH3YEGFP	CEN.PK113-13D with the yeGFP expression cassette from pTDH3YEGFP integrated into Flagfedt site 16	This thesis
F16RCGFPATTBP	CEN.PK113-13D with the Att site regulated reverse compliment yeGFP cassette from pTDH3YEGFP integrated into Flagfeldt site 16.	This thesis
F16RCGFPATTBP_BXBHA	CEN.PK113-13D with the Att site regulated reverse compliment yeGFP cassette from pTDH3YEGFP integrated into Flagfeldt site 16 and harbouring pBXBHA.	This thesis

Table 2.3: Oligonucleotides used in this thesis

Number	Oligo Name	Primer Sequence 5' to 3'
pTDH3ATTBP construction		
1.	MLT 506 TDH3 F	taaccctcactaaagggaacaaaagctggagctcgtttaaacggcgcgcctcgagtttatcattatcaatactgcc
2.	MLT TDH3 R	ctctagcgaagagcactagttctagagg
3.	MLT ENO2T F blunt	agtgcttttaactaagaattattagtcttttctg
4.	MLT Eno2t R 506	ataacttcgtataatgtatgctatacgaagttattaggtaccaggtatcatctccatctcccatatg
5.	MLT tdh3_YEGFP_noatt F	tagtctttttttagttttaaaacaccaagaacttagtttcgaaaaaaaatgtctatgtctaaaggtgaagaattattcact
6.	MLT eno2t_YEGFP_noatt R	aaataagcagaaaagactaataattcttagttaaaagcacttaattatttgtacaattcatccataccatg
pYEGFPATTBPconstruction		
7.	MLT tdh3_ATTB F	tctttttttagttttaaaacaccaagaacttagtttcgaggccggcttgtcacgac
8.	MLT attp_YEGFP_R	gggtttgtaccgtacaccactgagaccgcggtggttgaccagacaaaccactaattatttgtacaattcatccataccatg
9.	MLT attb_yeGFP F	ggccggcttgtcacgacggcggtctccgtcgtcaggatcatccggaaaaaaatgtctatgtctaaaggtgaagaattatt cactg
10.	MLT eno2t_ATTP R	gatgaaaaaataagcagaaaagactaataattcttagttaaaagcacttgggtttgtaccgtacacc
11.	MLT TDH3 attB R	ccggatgatcctgacgacggagaccgccgtcgtgacaagccggcctcgaaactaagttcttggtg
pYEGFPATTRL construction		
12.	MLT yeGFP Forward attRL F	ggccggcttgtcacgacggcggccgcggtggttgaccagacaaacaaaaaaatgtctatgtctatgtctaaaggtgaa gaattattc
13.	MLT yeGFP Forward attRL R	tgggtttgtaccgtacaccactgagatctccgtcgtcaggatcatccggatgttatttgtacaattcatccataccatg
14.	MLT eno2t_ATTL_R	aaaataagcagaaaagactaataattcttagttaaaagcacttgggtttgtaccgtacacc
15.	MLT tdh3_ATTR_F	tttagttttaaaacaccaagaacttagtttcgaggccggcttgtcacgac
pRCGFPATTBP construction		
16.	MLT yeGFP backwards F	ggcggtctccgtcgtcaggatcatccggaaaaaaatgtcttaattatttgtacaattcatccataccatg
17.	MLT yeGFP Backwards R	gtacaccactgagaccgcggtggttgaccagacaaaccacatgtctaaaggtgaagaattattcactg
pBXBHA and BXBInt expression cassette construction		
18.	MLT BxB1 Integ F	tcggtaatgggatccctctagaactagtgctcttcgctagagatgagagccctggtagtcatc
19.	MLT BxB1 Integ R	gctcactttaatttcggtttcgttatcttccaggctgcgtttttctacgacatcccggtgtgtag
20.	MLT TDH2t F	atttaactccttaagttactttaatgatttag
21.	MLT TDH2t R	gcgaaaagccaattagtgtg

22.	MLT TDH2t 506	ataacttcgtataatgtatgctatacgaagttattaggtaccgcgaaaagccaattagtgtg
23.	MLT Gal1p F No Linker	gtttaaacggcgcgcctagtac
24.	MLT Gal1p R No Linker	atccggggttttttctccttg
25.	MLT 506 Gal1p F	taaccctcactaaagggaacaaaagctggagctcgtttaaacggcgcgccgtttaaacggcgcgcctag
26.	MLT Gal1p bxb R	catcggtgacgcgggacaggcggatgactaccagggctctcatatccggggttttttctcc
27.	MLT BXB1 gal1pKOZ F	taacgtcaaggagaaaaaaccccggataaaaaaaatgagagccctggtagtcatcc
28.	MLT tdh2 BXB HA tag R	attaaagtaacttaaggagttaaatctaagcgtaatctggaacatcgtatgggtacgacatcccggtgtgtagccgttcg
29.	MLT BXB1 int :: TY1L R	ttaatttcggtttcattgtcctctagggatcttttttttgaatttattgtagtcgacatcccggtgtgtagc
30.	MLT TDH2t :: Ty1 L F	atccctagaggacaatgaaaccgaaattaaagtttctagagatacctggtagatttaactccttaagttactttaatgattta g
31.	MLT BXB int :: SV40NLS R	ctacaccttcctcttcttcttggggtccgacatcccggtgtgtagc
32.	MLT TDH2t :: SV40NLS F	gaccccaagaagaagaggaaggtgtagatttaactccttaagttactttaatgatttag

Integration construct assembly

33.	LN_16U5p_F	ttcgtgaaacacgtgggatacc
34.	LN_16U16u3p_R_univ_int_t ail	tcgtattaatttcgataagccaggttaacctgctccgttaattcgggtttcaatcactt
35.	LN_univ_int_f_16u3p_tail	tagacaaacgaagtgattgaaacccgaattaacggagcaggttaacctggcttatcgaa
36.	LN_univ_int_R_c1_tail	ctcaaagtaatgctgcagtctcggcgcgccggctacaattaatacataaccttatgtatc
37.	LN_16d5p_F_C6_tail	ggcaatcacatcaccatgagttgtgcggccgctgcctacgcaacactttagctg
38.	LN_16d3p	ttgttgggattccattgtgattaagg
39.	C1:506	gagactgcagcattactttgagaagtaaccctcactaaagggaacaaaagctggagctcgtttaaacggcgc gcc
40.	C6:506	ataacttcgtataatgtatgctatacgaagttattaggtaccgcggccgcacaactcatggtgatgtgattgcc

2.2: POLYMERASE CHAIN REACTION OF REPORTER CONSTRUCT CASSETTES

In order to gauge the activity of BXBInt in *S. cerevisiae,* a two-plasmid strategy was applied. One plasmid expressed yeGFP,[58] which has been codon-optimized for

expression in *S. cerevisiae* and the other plasmid expressed the BXBInt enzyme. In order to create a valid fluorescent assay, plasmids were constructed via a DNA Assembler[59] based strategy as illustrated in Figure 2-1.

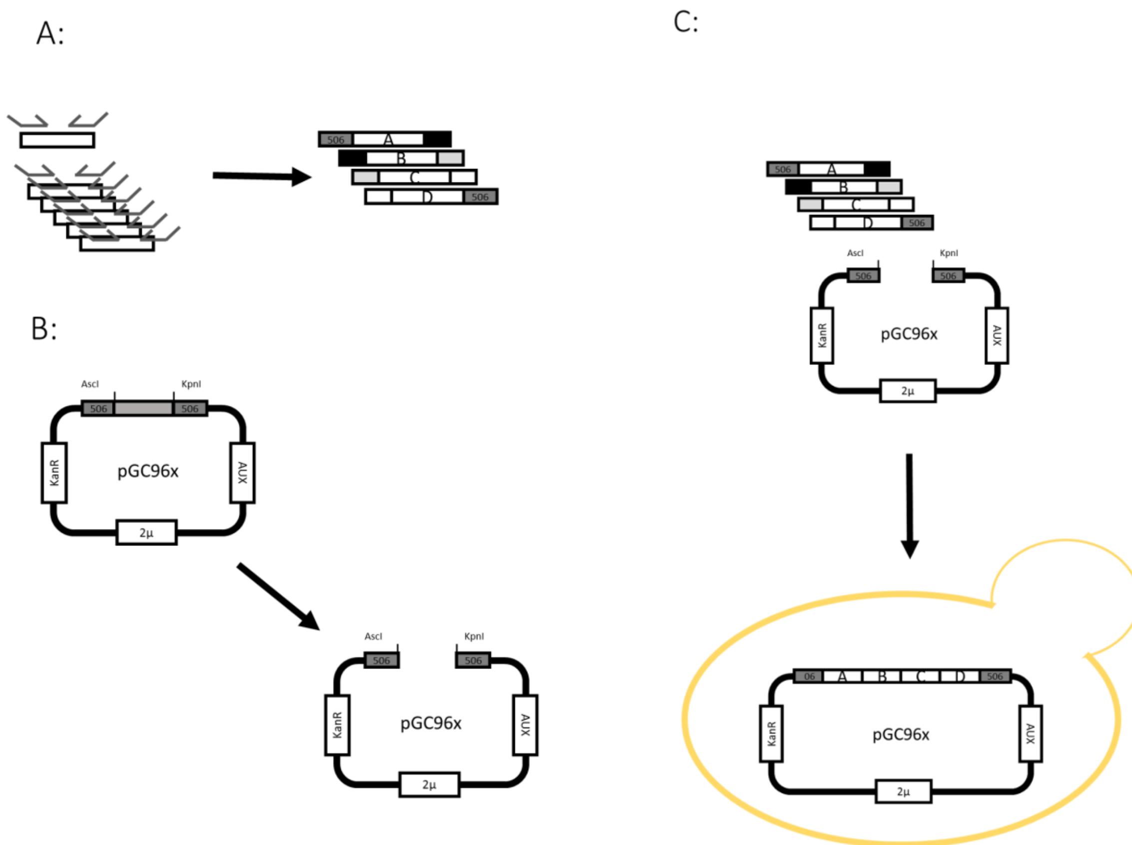


Figure 2-1: Schematic representation of DNA Assembler based cloning strategy. A: Insert fragments are amplified from numerous sources using primers synthesized such that they incorporate 40 bp sequences homologous to each adjacent amplicon and to the vector. B: the pGC966 and pGC967 vectors used in this study were prepared by digestion with *AscI* and *KpnI* which will expose 50 bp sequences compatible with homology-based cloning. C: Amplicons and linearized plasmids are co-transformed into a CEN.PK strain of yeast via lithium acetate/ssDNA mediated heatshock, and the Homologous Recombination pathway of the yeast assembles the amplicons and linearized backbone into a closed, circularized plasmid.

To construct the pTDH3YEGFP yeGFP expression cassette, promoter and terminator sequences were amplified using CEN.PK genomic DNA as a template, using primers **1** and **2** for the *TDH3* promoter and primers **3** and **4** for the *ENO2* terminator. These primer sets were designed to incorporate the ~40 bp 506 region of homology to

the vector. The vectors used in this study are based on the pGreg vector system,[60] modified to create unique *KpnI* and *AscI* restriction endonuclease sites to facilitate DNA Assembler cloning.[57] The yeGFP sequence was amplified by PCR using pGC966 as a template, and using primers **5** and **6**, which would add 40 bp of homology to the promoter and terminator amplicons.

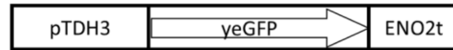
The cassette of interest for pYEGFPATTBP was constructed by amplifying the *TDH3* promoter and *ENO2* terminator from CEN.PK genomic DNA using **1** and **2** for the *TDH3* promoter and primers **3** and **4** for the *ENO2* terminator. The *TDH3* amplicon was used as a template for primers **1** and **11**, which would incorporate the sequence for AttB to the 3' end of the promoter sequence. The *ENO2* amplicon was used as a template for primers **4** and **10**, which would add the sequence for AttP to the 5' end of the terminator sequence. The yeGFP sequence was amplified using pGC966 as a template and primers **8** and **9**, which would add the AttP and AttB sequences to the forward-orientation yeGFP amplicon.

The cassette of interest for pYEGFPATTRL was constructed by amplifying the yeGFP ORF from pGC966 using primers **12** and **13**, which add the sequences for AttR and AttL to the 5' and 3' ends of the yeGFP amplicon. This amplicon was used as a template for a PCR reaction, which would add 40 bp of homology to the *TDH3* promoter and *ENO2* terminator using primers **34** and **35**. Promoter and terminator sequences were amplified from CEN.PK genomic DNA using primers **1** and **2** for the *TDH3* promoter and primers **3** and **4** for the *ENO2* terminator.

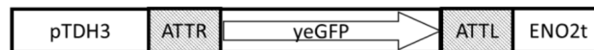
The cassette of interest for pRCGFPATTBP was constructed by amplifying the yeGFP ORF using pGC966 as a template, and using primers **16** and **17** to amplify the

yeGFP ORF in the reverse compliment orientation with respect to the AttB and AttP sequences included in the primer sequences. This amplicon was used as a template in a further amplification to add 40 bp of homology to the *TDH3* promoter and *ENO2* terminator using primers **10** and **11**. Promoter and terminator sequences were amplified using CEN.PK genomic DNA as a template, using primers **1** and **2** for the *TDH3* promoter and primers **3** and **4** for the *ENO2* terminator, which will amplify these regulatory elements with 40 bp linkers homologous to the 506 regions of pGC966. Schematic representations of the cassettes of interest from the aforementioned plasmids showing the location and orientation of regulatory elements with respect to the reporter ORF are shown in Figure 2-2.

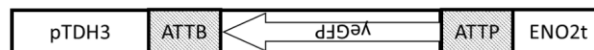
A



B



C



D

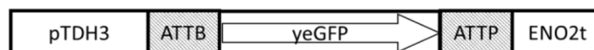


Figure 2-2: Schematic representations of the four constructs assembled to validate the activity of BXBInt in *S. cerevisiae.* A: Cassette from pTDH3YEGFP. B: Cassette from pYEGFPATTRL. C: Cassette from pRCGFPATTBP. D: Cassette from pYEGFPATTBP

Thermal cycling conditions for all reactions were optimized for the Phusion DNA Polymerase (New England Biolabs, Ipswich, MA, USA), with denaturing and annealing temperatures optimized for each template and primer set. PCR amplicons were subjected to electrophoresis in 0.8% (w/v) agarose gel with Tris/Acetate/EDTA (TAE) buffer. Amplicons of the correct size were excised from the gel and purified using the ThermoFisher GeneJET Gel Extraction kit (ThermoFisher Scientific, Waltham, MA, USA).

All plasmids were assembled via Transformation Associated Recombination (TAR) mediated by the Lithium Acetate/PEG/ssDNA method.[61] The pGC966 and pGC967 vectors used in this portion of the study differ from the original pGREG vectors from whence they derive in that they included two 50-bp sequences which are the sites of homologous recombination during TAR assembly, referred to as the “506” linkers of the vector in Fossati *et al.*[57]. The 506 linker sequences were exposed by digesting the vectors with *AscI* and *KpnI* (purchased from New England Biolabs, Ipswich, MA, USA). This digest was subjected to electrophoresis in 0.8% (w/v) agarose gel with TAE buffer. Linearized vector backbone was excised from the gel and purified using the ThermoFisher GeneJET Gel Extraction kit (ThermoFisher Scientific, Waltham, MA, USA).

Between 100 and 500 femtomoles of linearized vector and 300-1000 femtomoles of PCR products were pooled and introduced into yeast cells via lithium acetate mediated heat shock as described by Gietz & Schiestl.[61]

Reporter construct plasmids were used to transform CEN.PK auxotroph strains deficient in either the *LEU2* or *URA3* markers, such that successful recombination would complement these auxotrophies and allow for selection of assembled constructs by growth of the transformants on minimal media. A 1/5 and 1/50 dilution of each transformation reaction were plated YNB + 2% (w/v) glucose agar plates. In parallel, each CEN.PK strain was also transformed with only the linearized pGC966 vector as a negative control. Transformation plates were incubated at 30°C for 48-72 hrs, or until colonies became visible. Transformation reactions where a significantly greater (at least tenfold) number of colonies appeared on the transformation plates compared to the

negative control plates, putative transformant colonies were used to inoculate 5 mL cultures of liquid YNB + 2% (w/v) glucose in glass test tubes and incubated at 30°C with shaking at 200 RPM overnight.

Putative assembled vectors were purified from yeast cells by centrifuging 1-mL of overnight cultures at 16,000 RCF, decanting media, and suspending cell pellets with 50 μL of yeast lytic enzyme Zymolase 20-T (Amsbio, Cambridge, MA, USA product no. 120491-1) solution containing 50 uL of 100% β-mercaptoethanol per mL of lytic enzyme solution. Cell pellets suspended in lytic enzyme solution were incubated for 2-3 hrs at 37°C and plasmids purified using a Thermo Scientific GeneJet plasmid extraction kit modified for yeast plasmid purification by doubling the volume of reagents used in the protocol compared to bacterial plasmid purification.

Purified plasmids were used to transform *E. coli* strain DH5α via $CaCl_2$ mediated heat shock, and isolated colonies were grown in Luria Bertani liquid media with 100 μg/mL of ampicillin overnight and stored in 15% (v/v) glycerol at -80. Plasmids purified from *E. coli* stocks were sequenced by Eurofins Operon Genomics (Huntsville, AL, USA).

2.3: ASSEMBLY OF BXBINT PLASMIDS

The BXBInt expression cassettes were constructed by amplifying the BXBInt ORF using the International Genetically Engineered Machines (iGEM) 2011 distribution part BBa_I20284, whose sequence was obtained from[47] and can be found at (http://parts.igem.org/Part:BBa_I20284: Design as of February 2016). Primers **18** and **19** were used to amplify the ORF of BXBInt. This amplicon was used as a template for

further amplification to add the C-terminal tags to the BXBInt ORF and homologous sequences to terminator amplicons. Primers **27** and **28** were used to create the HA-tagged variant with homology to the *TDH2* terminator, primers **27** and **31** were used to create the SV40NLS tagged variant, and primers **27** and **29** were used to create the Ty1NLS tagged variant. Primer **27** included the consensus Kozak sequence (5'-aaaaaaatgagat-3') which would be incorporated immediately upstream of the BXBInt ORF. The sequence of primer **28** included a yeast codon-optimized DNA template for the HA tag amino acid sequence (YPYDVPDYA). The sequence of the Ty1 NLS included in primer **29** was a yeast codon optimized DNA template for the amino acid sequence (TTINSKKRSLEDNETEIKVSRDTW) as described in Moore *et al*.[62] The SV40NLS sequence included in primer **31** was a yeast codon-optimized DNA template for the amino acid sequence (PKKKRKV) as described in Makkerh *et al*.[63] The *GAL1* promoter was amplified using pGC966 as a template and using primers **23** and **24**. This amplicon was used as a template for primers **25** and **26** to amplify the *GAL1* promoter with the 506 linker and 40 base pairs of homology to the BXBInt ORF. The TDH2 terminator was amplified using CEN.PK genomic DNA as a template and using primers **20** and **21**. This amplicon was used as a template for further amplification to add linker sequences to the TDH2 terminator amplicon. Primers **20** and **22** were used to add the linker sequence homologous to the 506 vector linker to the 3' end of the terminator sequence. Primers **32** and **22** were designed to add the 506 linker and the SV40NLS sequence, Primers **30** and **22** were used to add 506 linker homology and Ty1 NLS sequence to the terminator amplicon.

2.4: QUANTIFICATION OF GFP FLUORESCENCE FROM TDH3YEGFP, YEGFPATTBP, AND YEGFPATTRL

The chosen structure of the BXBInt-mediated inversion cassette was designed to place AttB between the strong constitutive promoter *TDH3* and the yeGFP ORF, and to place AttP between the yeGFP ORF and the *ENO2* terminator. It was considered likely that this construct design may impact expression of the yeGFP reporter. To quantify what that effect may be, plasmids pTDH3YEGFP, which consisted of the forward orientation yeGFP under the power of *TDH3* without Att sites, was constructed to serve as a baseline for expression. pYEGFPATTBP and pYEGFPRL were constructed to contain a cassette where the pre-recombination AttB and AttP sites were interpolated between the ORF and regulatory elements and the post-recombination AttR and AttL sites in those same places, respectively. yeGFP fluorescence was quantified by flow cytometry.

Strains TDH3YEGFP, YEGFPATTBP, and YEGFPATTRL, were created via lithium acetate/PEG/ssDNA transformation by their respective plasmids (Table 2.1 and 2.2) and transformants selected on YNB + 2% (w/v) glucose agar plates. Three colonies from each plate were used to inoculate 5 mL of YNB + 2% (w/v) glucose in 25mm x 150mm borosilicate test tubes and all cultures were incubated at 30°C with shaking at 200 RPM on a rotary shaker for 18 hrs. Overnight cultures were diluted 1:10 in fresh media in preparation for flow cytometry.

Cells were analyzed by flow cytometry using an Accuri C6 cell analyzer from BD Biosciences (San Jose, CA, USA). This analyzer was equipped with a λ 488nm

excitation laser and λ 530/30 emission filter optimized for GFP fluorescence. A second detection filter channel at λ585/40 was also used to get a 2-dimensional map of green fluorescence and red autofluorescence. Gating of the cell analyzer data to get consistent measurements of only yeast cells in a similar stage of the cell cycle was as follows: Raw input from the detectors was normalized for Side Scatter Height (SSC-H) and Forward Scatter Height (FSC-H) to control for cell size. Events were gated along the line where a 1:1 SSC-Area: SSC-H ratio would be to ensure that only fluorescence from single cells is measured. The fluorescence from events, which were located within this gate was used to determine GFP fluorescence. GFP fluorescence was measured using the FL1 channel, corresponding to λ530/30 filtered detector. Measures of green fluorescence reported were the mean fluorescence from the FL1 channel, the percentage of events in the UL quadrant, and the mean fluorescence of events in the UR quadrant. Another graph of events was constructed consisting of a simple histogram of fluorescence from events from the FL1 channel, which could be gated around a non-fluorescent empty vector control. Mean fluorescence and the percentage of events whose FL1 values exceeded the gated value were recorded. Mean fluorescence values are reported in Chapter 3.1.

2.5: TESTING THE FUNCTION OF BXBINT IN S. CEREVISIAE VIA FLOW CYTOMETRY

The function of BXBInt to catalyze inversion of plasmid DNA in *S. cerevisiae* was assayed by generating the plasmids pBXBHA, pBXBTy1, and pBXBSV40, all of which contained the BXBInt ORF under the power of the galactose inducible reporter *GAL1*.

These plasmids were co-transformed alongside pRCGFPATTBP and pYEGFPATTBP, consisting of the reverse compliment sense yeGFP ORF and forward sense yeGFP ORF respectively, both flanked by AttB and AttP. Once BXBInt was expressed by growing the strains generated by this transformation in media containing galactose, inversion of the DNA on the yeGFP bearing plasmids could be inferred by flow cytometry looking for the appearance or disappearance of GFP fluorescence.

Strains RCGFPATTBP, BXBHA_YEGFPATTBP, BXBHA_ RCGFPATTBP, BXBTY1_ RCGFPATTBP, and BXBSV40_ RCGFPATTBP were created via lithium acetate/PEG/ssDNA transformation by their respective plasmids (Table 2.1) and transformants selected on YNB + 2% (w/v) glucose agar plates. For galactose-induction experiments, six colonies from each transformation plate were picked and used to inoculate 5-mL cultures of YNB media for overnight growth. Three colonies from each plate were used to inoculate YNB + 2% (w/v) glucose and three colonies were used to inoculate YNB + 2% (w/v) galactose, and all cultures were incubated at 30°C with shaking at 200 RPM on a rotary shaker for 18 hrs. Overnight cultures were diluted 1:10 in fresh media in preparation for flow cytometry. Flow cytometry was carried out as described in Section 2.4.

2.6: TIME COURSE

To determine the timing and heritability of BXBInt inversion, as well as to account for the relatively high stability of yeGFP in yeast cells when attempting to invert the pYEGFPATTBP yeGFP cassette, long-term incubation of strains RCGFPATTBP, BXBHA_RCGFPATTBP, and BXBHA_YEGFPATTBP in liquid culture was carried out.

For this experiment, strains RCGFPATTBP, BXBHA_RCGFPATTBP, and BXBHA_YEGFPATTBP were grown overnight in YNB + 2% (w/v) glucose in 25mm X 150mm test tubes with shaking at 200 RPM for 18 hrs. After the cultures had grown to significant density, cells were washed with sterile ddH_2O, and used to inoculate 5 mL of induction medium composed of YNB + 2% (w/v) galactose. In parallel, these cultures were used to inoculate fresh YNB + 2% (w/v) glucose as negative controls. Cultures were incubated as stated above for 72 hrs. Subsamples of each culture (100 µL) were analyzed by flow cytometry as described in Section 2.4 after being transferred to the induction medium. After 72 hrs of growth, those cultures grown in galactose were diluted 1/5000 in fresh media composed of YNB + 2% (w/v) glucose and analyzed by flow cytometry. Results are shown in Sections 3.2 and 3.3.

2.6: INTEGRATION OF THE REPORTER CONSTRUCT INTO YEAST CHROMOSOME14

In addition to constructing plasmids bearing the yeGFP reporter constructs, attempts were made to integrate these same reporter constructs into the genome of *S. cerevisiae* in order to determine whether BXBInt could invert genomic DNA. The site of the integration event was selected based on a study by Flagfeldt *et al.,*[64] wherein multiple genomic integration sites were characterized based on the level of gene expression of an integrated reporter cassette. Flagfeldt site 16 was selected based on the high expression levels of integrated reporter constructs found in this study, and based on the high quality of publicly available sequence information of this region of the yeast genome. Site 16 is located on the p arm of chromosome 14, at the YNRCΔ9 locus

(SGD ID S000007145), an intergenic Long Terminal Repeat (LTR) sequence from the Ty1 retrotransposon.[65] Integration of the reporter constructs from pTDH3YEGFP and pRCGFPATTBP was attempted by simple homologous recombination of a linear DNA construct of the reporter construct linked to a G418 resistance cassette and two ~600 bp regions of chromosomal homology upstream and downstream of the YNRCΔ9 LTR. A schematic representation of the chosen integration strategy is shown in Figure 2-3.

Integration:

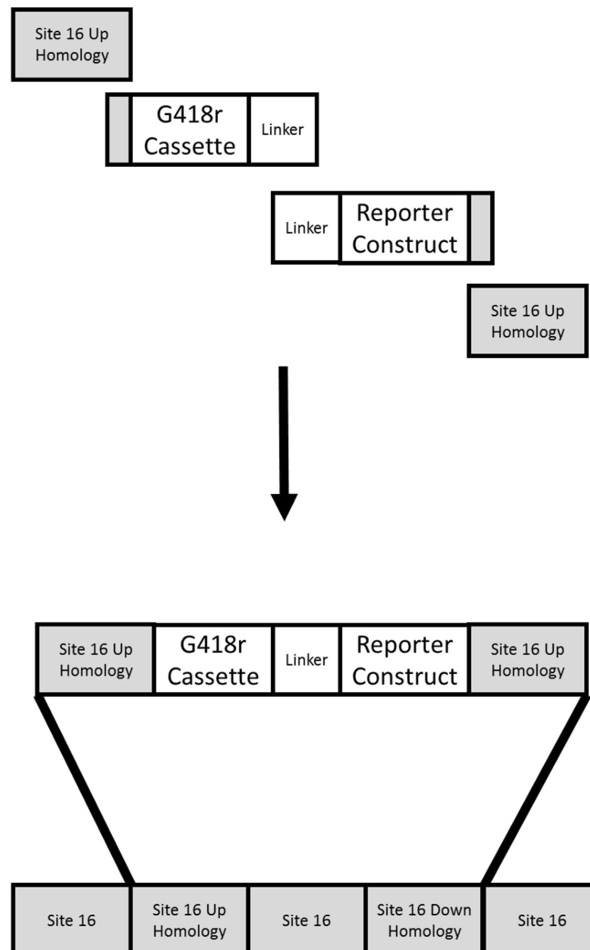


Figure 2-3: Schematic representation of the strategy implemented to integrate the RCGFPATTBP construct into Flagfeldt Site 16. The integration cassette was amplified in four parts, which were spliced together by SOE-PCR. Upon transformation of CEN.PK113-13D, the integration cassette was expected to recombine into Flagfeldt site 16. Integration carried out in this manner could be selected for using G418.

The Site 16 Upstream (16U) chromosomal homology regions were amplified from CEN.PK genomic DNA with primers **33** and **34**, which incorporated 40 base pairs of homology to both a G418 resistance cassette derived from the pUG6 plasmid. This G418 resistance cassette was amplified using pUG6 as a template and primers **35** and

36. Primer 36 incorporated a 40 bp linker sequence, called C1, which would facilitate homologous recombination or Splicing by Overlap Extension PCR (SOE-PCR) with the downstream amplicon. The TDH3YEGFP and RCGFPATTBP cassettes amplified using plasmids pTDH3YEGFP and pRCGFPATTBP as templates and using primers **39** and **40** in both cases. These primers would incorporate the C1 linker to the 5' terminus of the amplicon and the similarly designed 40 bp C6 linker sequence to the3' end of this amplicon. The Site 16 Downstream chromosomal homology region (16D) sequence was amplified using primers **37** and **38**, which would add the C6 linker to the 5' terminus of the amplicon. The two chromosomal homology regions, the G418 cassette, and reporter construct cassette were amplified into one fragment of DNA by SOE-PCR using a mix of the four amplicons described above (diluted 1/10, approximately 10 ng of each amplicon) as template and primers **33** and **38.** Thermal cycling conditions were optimized for the Phusion polymerase (New England Biolabs, Ipswich MA, USA). The spliced PCR product was subjected to electrophoresis in 0.8% (w/v) agarose gel with TAE buffer. Amplicons of the correct size were excised from the gel and purified using the ThermoFisher GeneJET Gel Extraction kit (ThermoFisher Scientific, Waltham, MA, USA) and used to transform strains of CEN.PK113-13D via the lithium acetate/PEG/ssDNA method. Transformants were plated on YPD + G418 agar plates and incubated at 30°C for three days. A negative control transformation containing only the Site 16 homology amplicons and the G418 resistance cassette was undertaken in parallel. In order to ensure that individual clones were isolated, colonies from the transformation reaction were picked and re-streaked on fresh YPD + G418 plates and incubated in an identical manner. Colonies from the first round re-streak plates were

also picked and re-streaked on fresh YPD + G418 plates and incubated a second time. Colonies from the second round re-streak plates were selected for further analysis.

Successful integration of reporter construct cassettes was verified by PCR. Genomic DNA was extracted from cells by growing putative integrants overnight in YPD + G418 liquid cultures, centrifuging 1 mL of cultures, which grew in the selective media to obtain cell pellets. Cells were lysed by suspending the cell pellets in 50 µL of a solution composed 0.3 mM lithium acetate and 20% (v/v) dimethyl sulfoxide in ddH_2O. Four volumes of 95% (v/v) ice-cold ethanol was added to the lysate, the solution was mixed thoroughly, and incubated at -20°C for 30 minutes in order to precipitate DNA. Samples were centrifuged at 16,000 RCF for 20 minutes at 4°C to pellet DNA, after which the supernatant was decanted and the DNA pellet washed with ice-cold 70% (v/v) ethanol. DNA pellets were dried and suspended in ddH_2O. One microliter of this DNA extraction was used as a template in a 50 µL PCR analysis using four primer sets: Primers **33** & **38**, **35** & **36**, **39** & **40**, and **39** & **38** as shown in Figure 2-4:

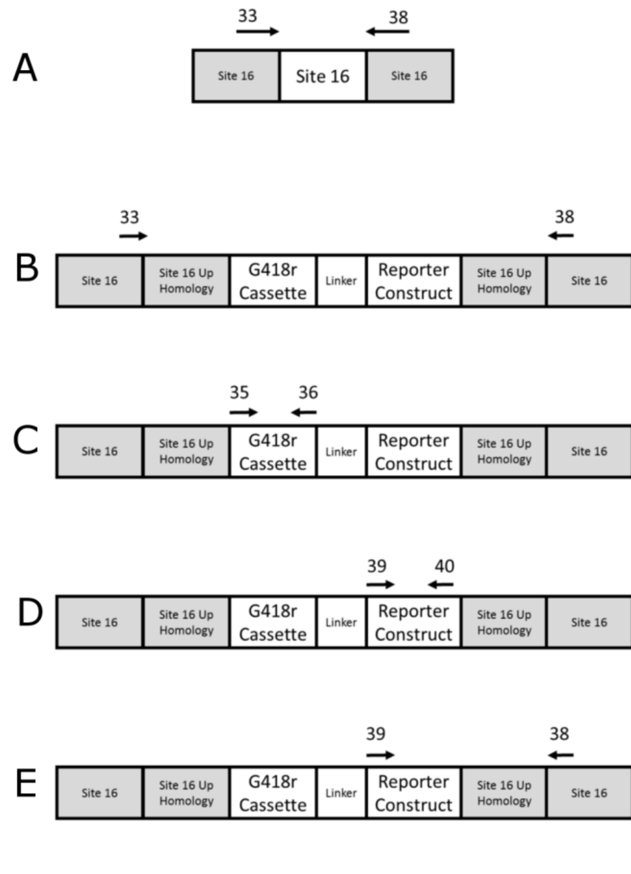


Figure 2-4: PCR reactions used to map the integration cassette from purified genomic DNA of G418 resistant colonies. Primers used are shown based on their numbers in Table 2-3. A: Primers **33** and **38** in the background strain CEN.PK113-13D would amplify a short, ~500bp fragment. B: Primers **33** and **38** would amplify a large ~5 kb fragment if integration was successful. C: primers **35** and **36** would amplify a 2 kb fragment. D: primers **39** and **40** would amplify a 2 kb fragment. E: Primers **39** and **38** would amplify a ~2.5 kb fragment.

Colonies that tested positive for features of the pRCGFPATTBP integration cassette were then transformed with the pBXBHA plasmid to test for inversion activity of chromosomally integrated DNA. Plasmid was transformed using the lithium acetate/ssDNA/PEG method described in Section 2.3 and selected on YNB + 2% w/v) glucose.

2.7: FLUOROMETRY OF PUTATIVE RCGFPATTBP INTEGRANTS

F16TDH3YEGFP, F16RCGFPATTBP, and F16RCGFPATTBP_BXBHA strains, which were transformed with the galactose inducible BXBInt plasmid pBXBHA were analyzed by microtiter plate fluorometry. Transformants were grown for 24 hrs in YNB + 2% (w/v) glucose (control) and YNB + 2% (w/v) galactose, to induce expression of BXBInt. Cultures were diluted in ddH_2O to an OD_{660} of 1. Fluorescence analysis of F16TDH3YEGFP, F16RCGFPATTBP, and F16RCGFPATTBP_BXBHA strains was carried out using a Tecan Infinite 200 plate reader instrument. Cultures were transferred to BD Falcon 96 well flat bottom transparent/black polyethylene terephthalate HTS FluoroBlok plates in aliquots of 100 µL. Fluorometric parameters were as follows: Excitation wavelength of 475/9, emission wavelength detected at 509/20, gain set to 100, 25 flashes per well. Plates were set to shake for 1 second with a linear amplitude of 1 mm at 100 RPM before each measurement. Biological triplicates of each strain were analyzed and average fluorescence was recorded, the results of which can be found in Section 3.4.

CHAPTER 3: RESULTS

3.1: EFFECT OF ATT SITES ON YEGFP EXPRESSION LEVELS

Quantitative flow cytometry was undertaken to determine whether the presence of Att sites between promoter and the reporter ORF would impact expression of that reporter, which would be an important factor to consider if this system of gene expression control is to be used for quantitative biological manipulations. Figure 3-1

shows a comparison of yeGFP expression from a CEN.PK1113-16B strain bearing a pGC966 derivative with a non-protein coding DNA fragment as a negative control, TDH3YEGFP, YEGFPATTBP, and YEGFPATTRL.

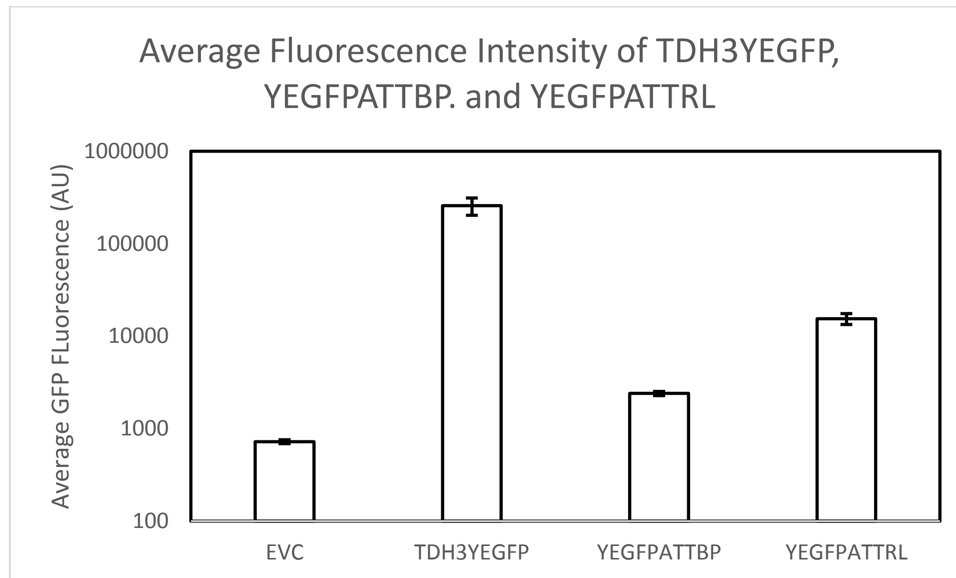


Figure 3-1: Flow cytometry of TDH3YEGFP, YEGFPATTBP, and YEGFPATTRL. Bars represent average fluorescence of three biological replicates, black lines indicate standard error. Interpolating the AttB and AttP sites between the ORF and promoter/terminator leads to a 100-fold decrease in fluorescence compared to the same construct without the Att sites, while interpolating the post-recombination AttR and AttL in the same locations leads to a 10-fold decrease in fluorescence compared to TDH3YEGFP. Also shown is the fluorescence from CEN.PK113-16B transformed with a non-fluorescent empty pGC966 vector (EVC)

The greatest amount of fluorescence comes from the construct without any Att sites separating the promoter region from the start codon of the yeGFP ORF. The construct where the AttB site is located between the promoter and coding sequence and the AttP site located between the end of the coding sequence and the terminator sequence produces approximately 1% of the fluorescence as the non-Att site bearing

control. The construct with the BXB AttR and AttL sites situated between the promoter, coding sequence, and terminator results in approximately 10% of the fluorescence of the non-Att site bearing control, ten times brighter than the AttB and AttP site bearing control.

3.2: SERINE INTEGRASE MEDIATED INVERSION OF PLASMID DNA

Figure 3-2 shows the results of flow cytometry of RCGFPATTBP, BXBHA, and BXBHA_RCGFPATTBP grown in minimal media containing glucose and galactose.

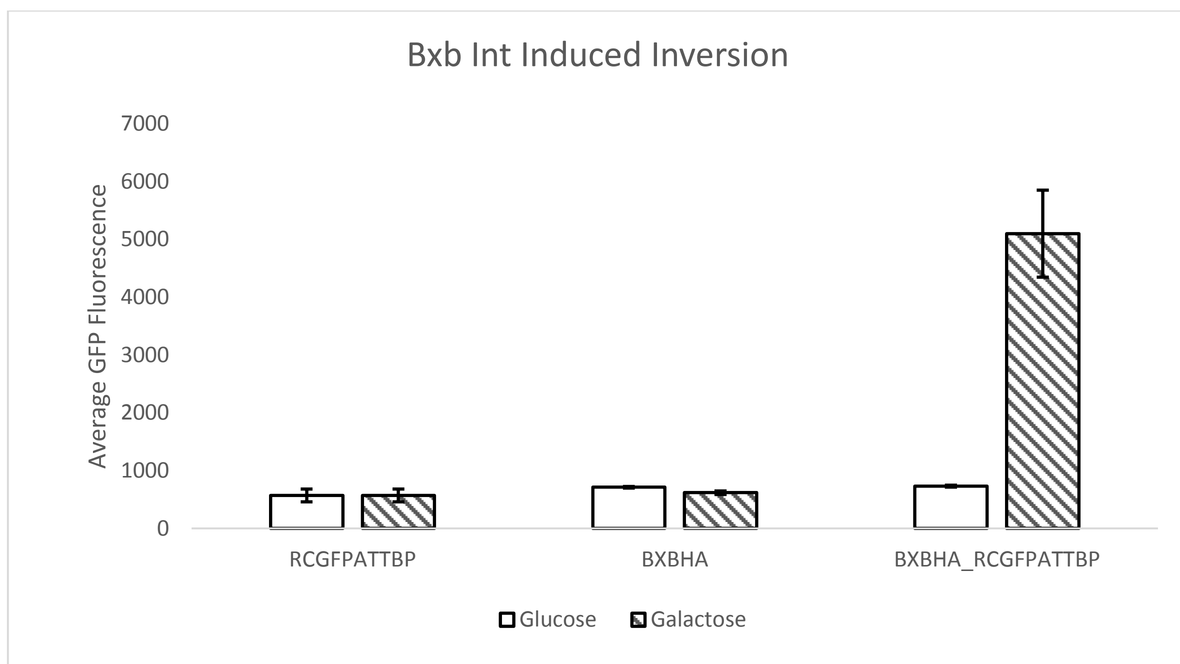


Figure 3-2: Flow cytometry of BXBInt mediated inversion. Yeast strains bearing either the pRCGFPATTBP plasmid, the galactose-inducible pBXBHA g plasmid, or BXBHA_RCGFPHA harbouring both plasmids, were grown for 24 hrs either in YNB + 2% (w/v) glucose or YNB + 2% (w/v) galactose. Bars represent the mean fluorescence of three biological replicates, black lines represent the standard error three biological replicates. Fluorescence intensity is reported in arbitrary units (AU). Fluorescence intensity only increases above background levels when both pBXBHA and pRCGFPATTBP are present in the cell, and BXBInt expression is induced with galactose, indicating that the reverse-compliment GFP ORF has been inverted into a translationally active orientation.

YeGFP expression was not observed in the presence of the reporter construct alone, ruling out the possibility that the pRCGFPATTBP cassette could be induced by galactose on its own. The absence of fluorescence from the pRCGFPATTBP plasmid transformed strain also indicates that yeGFP is not expressed prior to the expression of BXBInt. The lack of fluorescence from strain BXBHA precludes any contribution of the BXBInt plasmid or protein to fluorescence. Fluorescence is only observed in the strain harbouring both pRCGFPATTBP and pBXBHA when grown in galactose. These results make sense in light of the DNA inversion hypothesis, where induction of the BXBInt protein remodels the reporter construct such that the yeGFP ORF is in an orientation with respect to the *TDH3* promoter that permits translation and transcription of a functional yeGFP transcript and protein. This result provides strong evidence of the inversion activity of a serine integrase in yeast.

In order to determine the timing of serine integrase mediated inversion, the expression of yeGFP was measured over a 24 hrs post-galactose induction period (Figure 3-3). This experiment shows that yeGFP fluorescence becomes detectable by flow cytometry between 1 and 3 hrs, and seems to peak at 5 hrs.

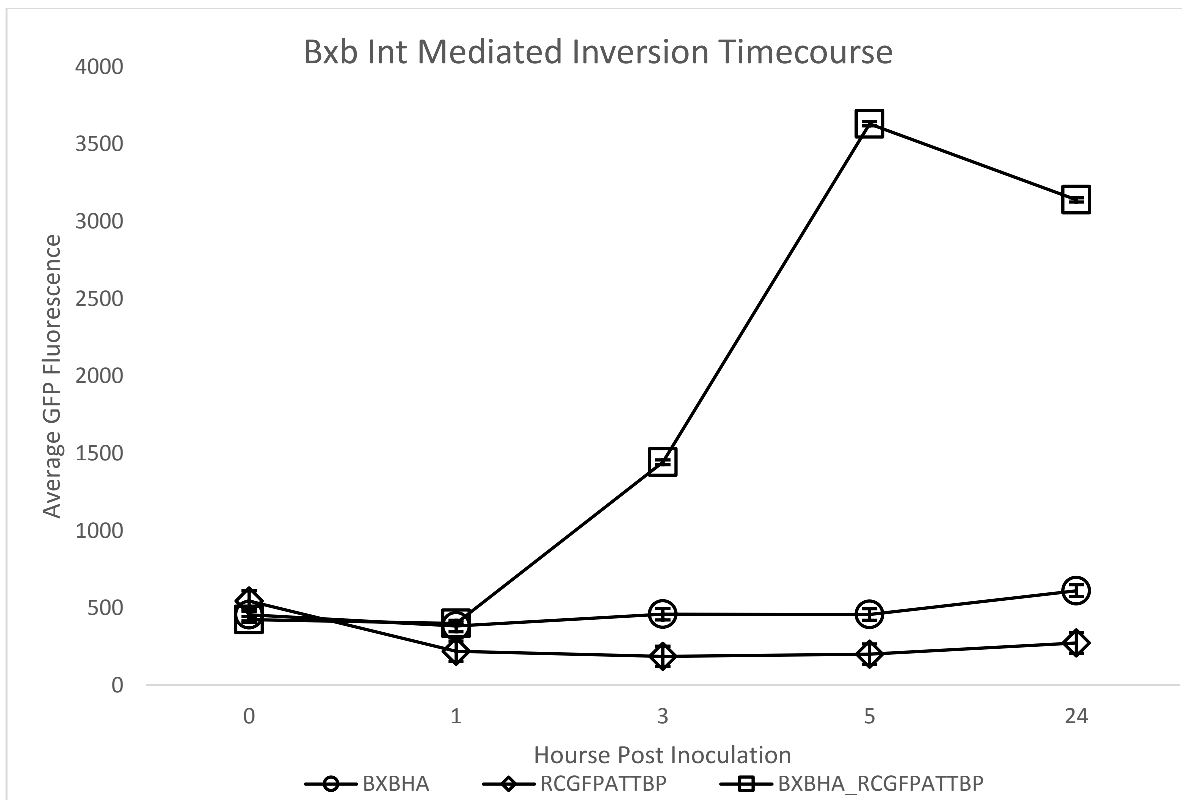


Figure 3-3: Time course of BXBInt inversion. Yeast strains are BXBHA (circles), RCGFPATTBP plasmid (diamonds) and BXBHA_RCGFPATTBP (squares). Expression of the inverted yeGFP reporter peaks at 5 hrs and slightly decreases after 24 hrs. Points represent average ungated fluorescence of three biological replicates and black lines indicate standard error of three biological replicates.

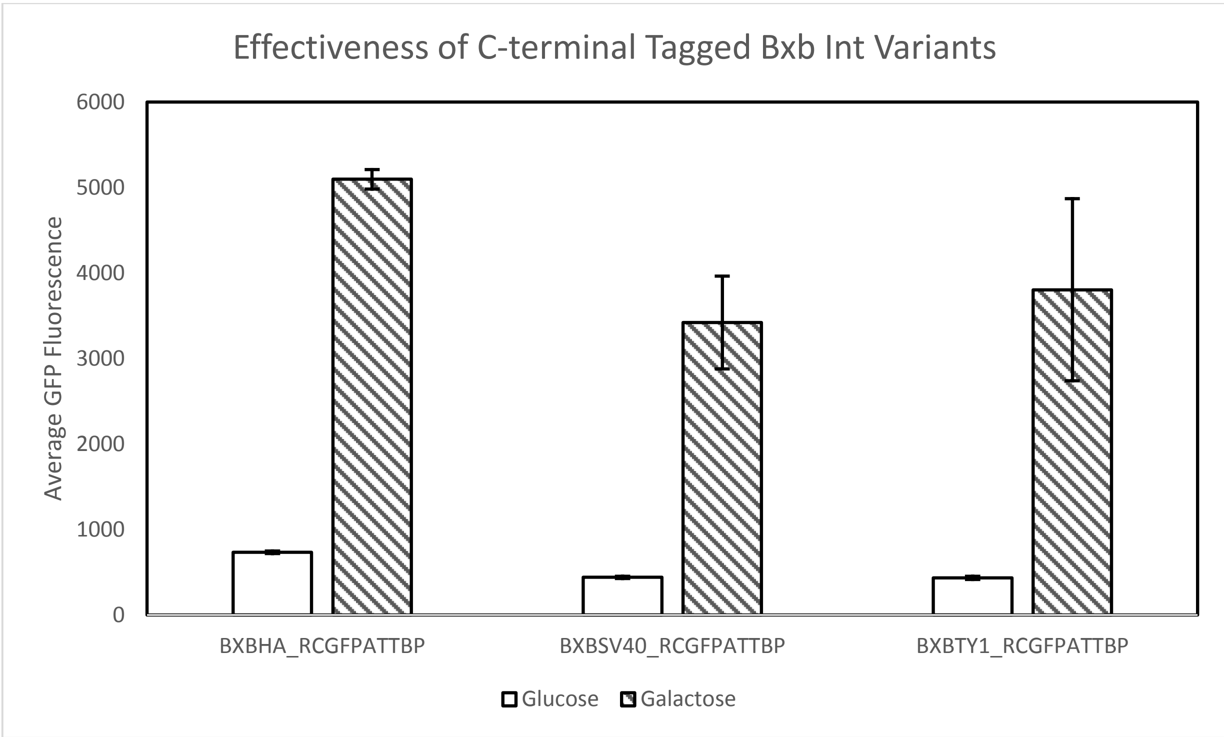


Figure 3-4: Effectiveness of C-terminal tagged BXBInt variants. Flow cytometry of yeast strains bearing the pRCGFPATTBP plasmid, as galactose inducible BXBInt bearing vectors with different C-terminal tagged BXBInt variants. Categories represent average ungated fluorescence of three biological replicates. Black lines represent standard error of three biological replicates.

Variants of the BXBInt protein bearing C-terminal nuclear localization sequences were constructed to determine whether nuclear localization would have an effect on GFP fluorescence. It is possible that when BXBInt is synthesized in the cytosol of the cell, the nucleus might act as a barrier preventing the integrase from accessing the plasmid bearing the reporter construct. Figure 3-4 shows fluorescence from strains BXBHA_RCGFPATTBP, BXBTY1_RCGFPATTBP, and BXBSV40_RCGFPATTBP. There is no difference in the average fluorescence of strains expressing the different C-terminal tags, indicating that nuclear localization is not a significant obstacle towards BXBInt inversion of plasmid DNA. Figure 3-5 shows the fluorescence of strains BXBHA_RCGFPATTBP, RCGFPATTBP, and BXBHA grown for three days in induction

media followed by 24 hrs in glucose. There is no significant decrease in fluorescence of BXBHA_RCGFPATTBP when grown in galactose for 24 hrs and 72 hrs, or when back diluted into glucose containing media. This indicates that the inversion of DNA mediated by BXBInt is heritable in yeast plasmids.

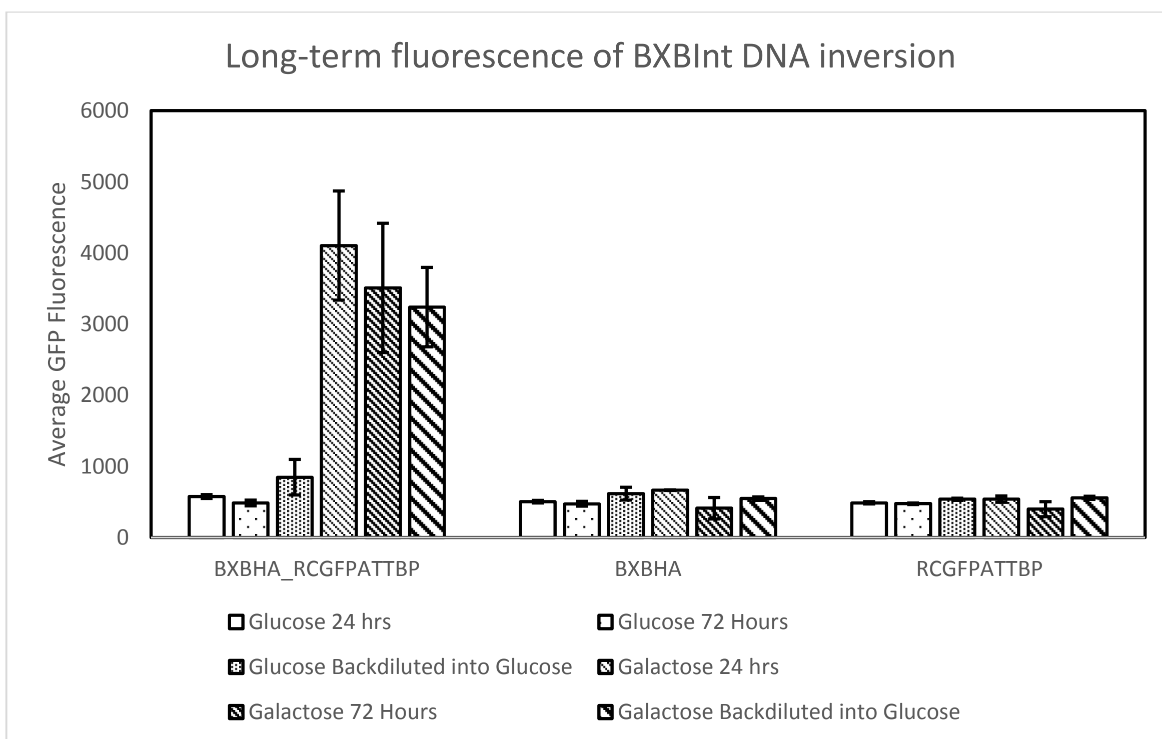


Figure 3-5: Long-term expression of BXBInt mediated inversion. Bars show average fluorescence of three biological replicates and black lines show standard error. Cultures were grown in YNB + 2% (w/v) glucose and YNB + 2% (w/v) galactose for 72 hrs, after which they were backdiluted into fresh YNB + 2% glucose.

3.3: FYEGFPATTBP: TURNING OFF GENE EXPRESSION USING BXBINT

The proposed mechanism of serine integrase mediated DNA inversion implies that the initial orientation of a DNA sequence flanked by Att sites should be irrelevant so

long as the orientation of the Att sites favours recombination, and that DNA inversion should occur regardless of the initial state of the intervening DNA. To test this hypothesis about the BXBInt system an alternate reporter construct, YEGFPATTBP, was constructed consisting of the yeGFP ORF flanked by the BXB AttB and BXB AttP sites, cloned in the transcriptionally permissible “forward” orientation with respect to the *TDH3* promoter and *ENO2* terminator. Flow cytometry of yeast cells bearing this reporter construct plasmid shows that there is significant GFP fluorescence in the magnitude expected of a GFP ORF separated from the *TDH3* promoter sequence by a DNA sequence as shown in Figure 3-6, further demonstrating that this reporter construct was assembled as expected.

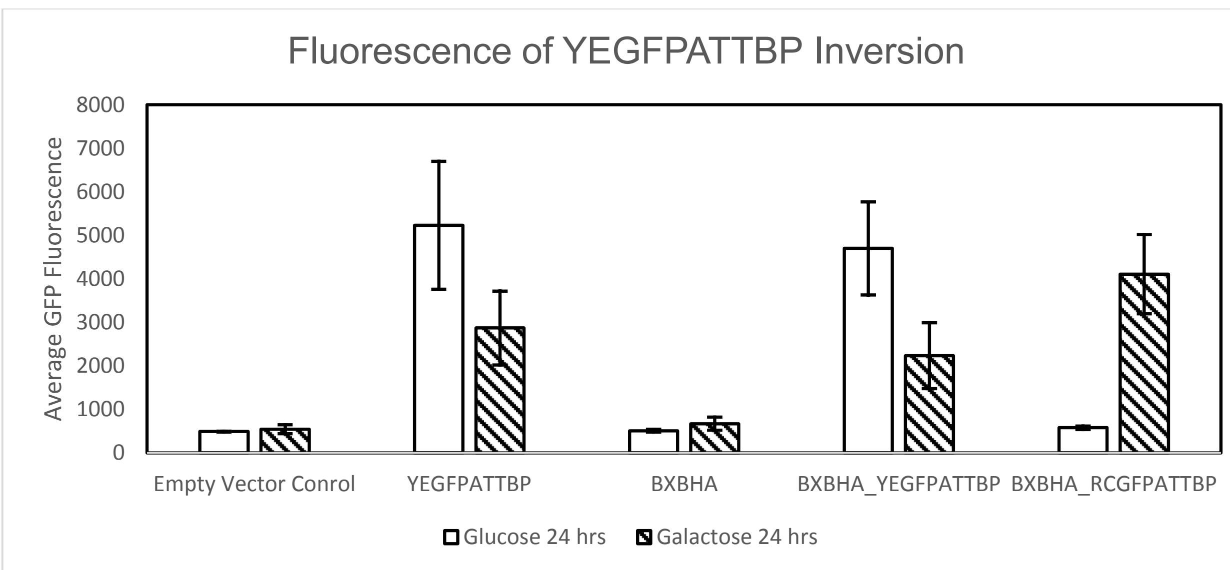


Figure 3-6: Fluorescence of YEGFPATTBP inversion. Flow cytometry of yeast cultures bearing pYEGFPATTBP, pBXBHA, and both plasmids plasmid after 24 hrs of growth in YNB + 2% (w/v) glucose and YNB + 2% (w/v) galactose.

Co-transformation of the pYEGFPATTBP plasmid and the pBXBHA plasmid also proceeded as expected of a successful co-transformation. A uracil and leucine auxotrophic strain (CEN.PK 17A) subjected to lithium acetate mediated heat shock in the presence of both plasmids yielded viable colonies when plated on YNB + 2% (w/v) glucose agar plates, where the same double auxotrophic strain transformed with the pYEGFPATTBP plasmid alone yielded no viable colonies. Growing the viable colonies from the double transformation experiment in YNB + 2% (w/v) glucose and YNB + 2% (w/v) galactose yields an interesting result. After 24 hrs of logarithmic growth from a single colony to a high-density liquid culture, there is no difference in the GFP fluorescence of the yeast bearing both the YEGFPATTBP and BXBHA_YEGFPATTBP strains grown in glucose and grown in galactose. This pattern was also observed after 72 hrs of growth, and also after back diluting the cultures into fresh induction media, as is shown in Figure 3-7. This back dilution was performed so as to allow for the yeast to divide for several generations in the induction media in order to eliminate any effect that the stability of the yeGFP protein would have on the observations.

Figure 3-8 shows the percentage of flow cytometry events that show GFP fluorescence exceeding the level gated around a non-fluorescent empty vector control. This graph shows a similar pattern to that shown in Figure 3-7, where there is a decrease in fluorescence observed in those cells grown in galactose compared to glucose for both YEGFPATTBP and BXBHA_YEGFPATTBP over the course of 72 hours. There appears to be no significant difference in GFP fluorescence over the course of 72 hours when you compare these two strains to each other when grown in galactose.

Back dilution into fresh media does appear to reveal a difference in the proportion of GFP fluorescing cells between YEGFPATTBP and BXBHA_YEGFPATTBP. YEGFPATTBP grown in glucose for 72 hours and back diluted into glucose shows 66.7% of cells fluoresce above the background, where YEGFPATTBP grown in galactose for 72 hours and back diluted into glucose shows 53.4% of cells fluoresce above the background level. BXBHA_YEGFPATTBP grown in glucose for 72 hours and back diluted into glucose shows 66.4% of cells fluoresce above the background, where BXBHA_YEGFPATTBP grown in galactose for 72 hours and back diluted into glucose shows 32.4% of cells fluoresce above the background level. Also shown is the percentage of BXBHA_RCGFPATTBP cells that fluoresce above background level when grown in glucose and galactose over the same time scales and with the same back dilution.

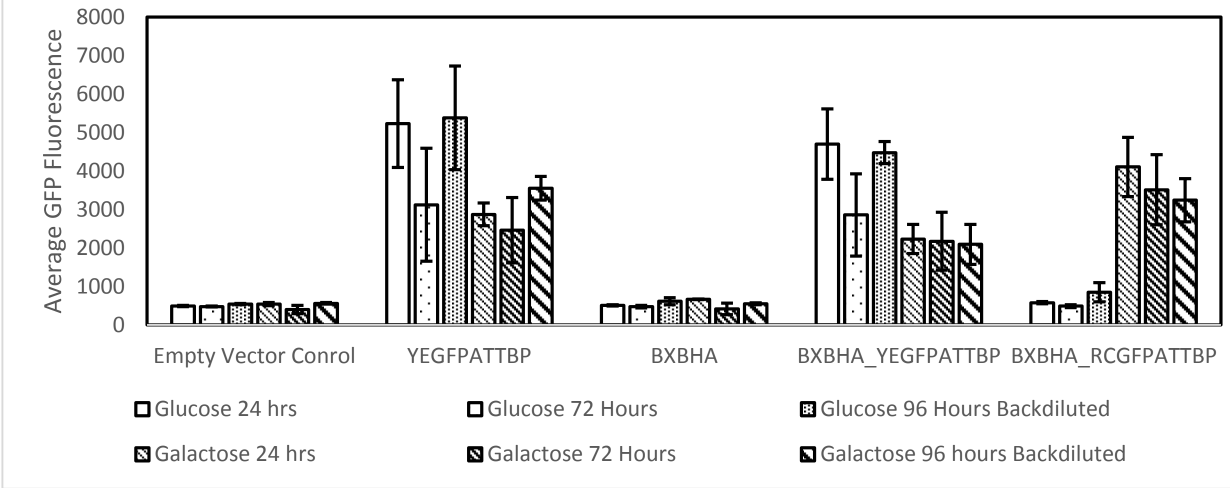


Figure 3-7: Fluorescence of YEGFPATTBP after 24, 72, and 96 hrs of induction. Attempted inversion of pYEGFPATTBP monitored for fluorescence at 24 and 72 hrs post induction, and following re-inoculation 1/5000 in fresh media followed by 24 hrs of exponential growth.

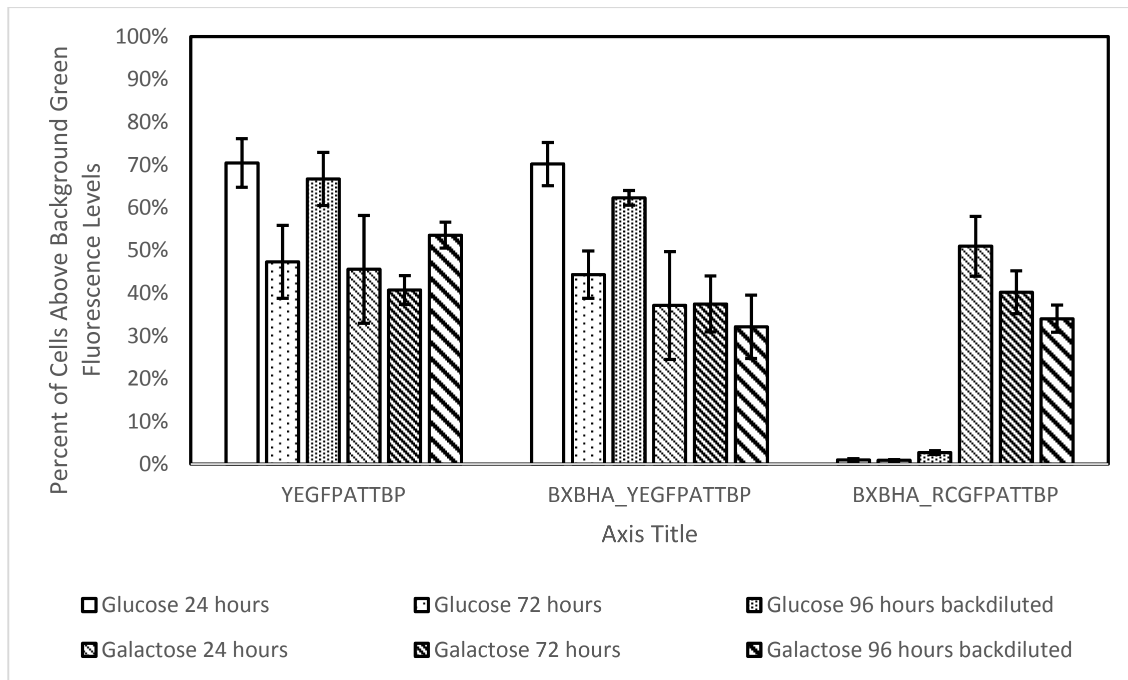


Figure 3-8: The proportion of cells whose green fluorescence exceeds the level of the gate set around an empty vector control. Fluorescence of YEGFPATTBP after 24, 72, and 96 hrs of induction. Attempted inversion of pYEGFPATTBP monitored for fluorescence at 24 and 72 hrs post induction, and following re-inoculation 1/5000 in fresh media followed by 24 hrs of exponential growth

3.4: INTEGRATION OF REPORTER CONSTRUCT INTO CHROMOSOME 14

Literature characterizing the activity of LSTRs in microbial hosts does not include published attempts to invert chromosomal DNA.[52] To this end, the pRCGFPATTBP cassette was integrated into chromosome 14 at Flagfeldt site 16. An integration cassette was constructed in four parts to be assembled via TAR and integrated via homologous recombination. The four parts were a 627 bp upstream region of homology to Site 16 (Site 16 Up), a 1601 bp G418 resistance cassette from the pUG vector (G418r), the 1800 bp reporter construct cassette from the previously described pRCGFPATTBP or pTDH3YEGFP plasmids, and a 673 bp region of downstream homology to Flagfeldt site 16 (Site 16 Down). These parts were amplified using primers that added ~40 bp of overlap to their adjacent parts. When 500 ng of each part was used to transform the uracil auxotrophic strain CEN.PK 13D via lithium acetate/PEG/ssDNA mediated heat shock and plated on YPD + G418 agar plates, dozens of colonies appeared on plates inoculated with the yeast transformed with all four parts, and no colonies appeared on plates inoculated with yeast transformed with only the Site 16 Up, G418R parts, and Site 16 Down parts. Colonies from the 4-part transformation reaction were picked and streaked onto fresh YPD + G418 plates, resulting in successful growth of these colonies on fresh plates. Inoculating fresh YPD + G418 plates using a pipette tip which had been dragged across the surface of the plate which had been inoculated with the 3-part transformation mixture did not result in the growth of any colonies. This technique for re-streaking the putative integrants was repeated a second time with identical results. It can be inferred from this that the G418r

cassette was integrated successfully and could only be integrated in the presence of all four parts of this integration cassette, further suggesting that all four parts were integrated successfully. To confirm this integration, genomic DNA was extracted from multiple colonies corresponding to unique colonies from the initial transformation plate. PCR was undertaken to confirm the presence of the integration construct within Site 16. Primers selected to amplify the entirety of the Site 16 region, external to the homology sequence of the integration cassette, were expected to amplify a ~1500 bp region of Site 16 in unintegrated CEN.PK 13D, and amplify a ~3500 bp region of Site 16 if integration of the complete construct was successful. As is shown in Figure 3-8, although a clean 3500 bp band could not be obtained from the putative integrants, they showed the same streaked pattern of amplification as the flow-cytometry verified integrant F16TDH3YEGFP, with the size amplified DNA appearing to have an upper limit of ~3500 bp. Most notably, the same primers were able to amplify a ~1500 bp band from the background strain CEN.PK 13D. The ~1800 bp G418r cassette was also amplified from these genomic DNA extractions in those colonies selected for further testing, as well as the ~2000 bp pRCGFPATTBP cassette. One of the integrants gives an inconsistent PCR amplification profile, where the G418r cassette and reporter construct amplifications work, but the reactions using primers **33** and **38**, and **38** and **39** failed. This is likely due to a failure of the PCR amplification due for practical regions, not necessarily because the integrant contained two of the targeted sequences but not the other two. In any case, five of the six integrants shown are positive for all four of the diagnostic PCR amplifications.

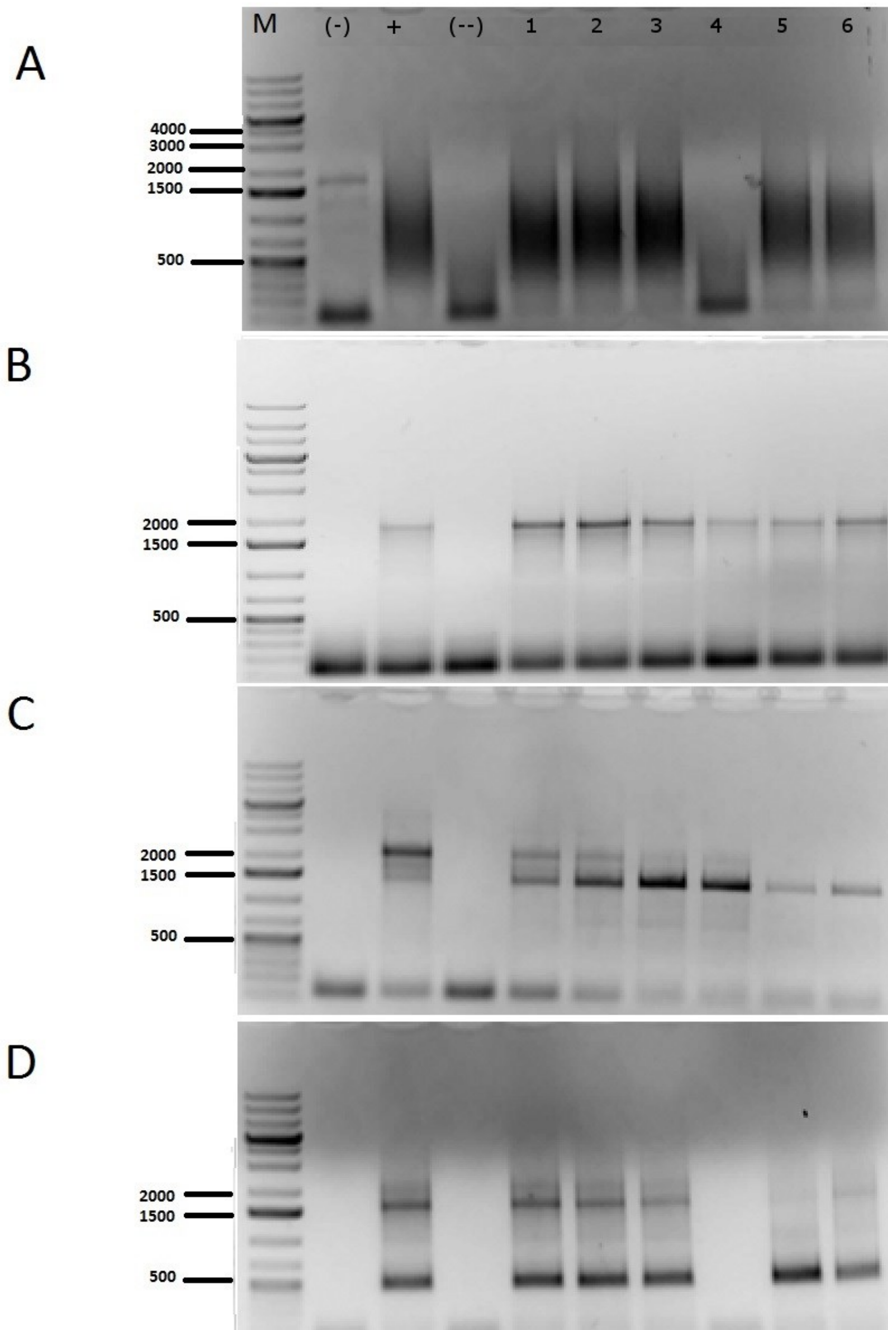


Figure 3-9: Ethidium bromide stained agarose gel images of Site 16 integration PCR characterization. Wells are, from left to right, M: Generuler 1 Kb Plus DNA marker, (-): CEN.PK 13D background strain (Negative control), (+): F16TDH3YEGFP (Positive Control), (--): No Template Control (sterilized MilliQ water used as template), 1-6: Putative F16RCGFPATTBP integrants. A: Primers **33** and **38** amplifying anything between the borders of Site 16. B: Primers **35** and **36** amplifying the G418 resistance cassette**.** C: Primers **39** and **40** amplifying the RCGFPATTBP cassette. D: Primers **39** and **38** amplifying the region between the RCGFPATTBP cassette and the external border of Site 16**.**

Fluorometry of F16TDH3YEGFP, BXBHA_RCGFPATTBP, CEN.PK113-13D, and three colonies of F16RCGFPATTBP_BXBHA show that although strong GFP fluorescence was detectable from F16TDH3YEGFP integrant, and more modest fluorescence could be detected from BXBHA_RCGFPATTBP grown in galactose using this instrument, no fluorescence could be detected from any of the tested F16RCGFPATTBP_BXBHA clones grown in glucose or galactose, and no evidence for inversion of the integrated RCGFPATTBP construct could be detected. The results of this fluorometry analysis is shown in Figures 3-9 and 3-10.

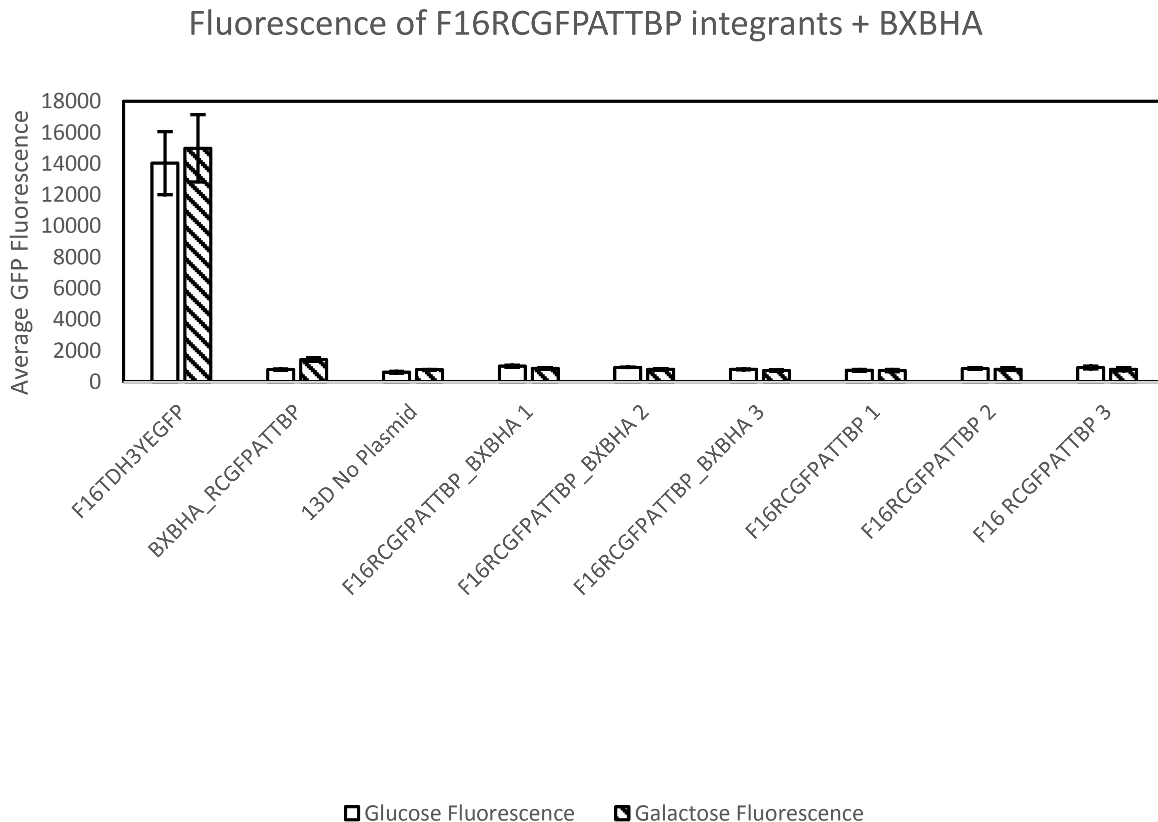


Figure 3-10: Fluorescence of F16RCGFPATTBP integrants + BXBHA. Microtiter plate fluorometry of CEN.PK 13D integrated with RCGFPATTBP as well as those putative integrants also bearing the *GAL1*p-BXBInt plasmid. Shown also are CEN.PK 13D integrated with a *TDH3* regulated yeGFP without

Att sites, as well as TECAN fluorometry of plasmid borne RCGFPATTBP + BXBInt.

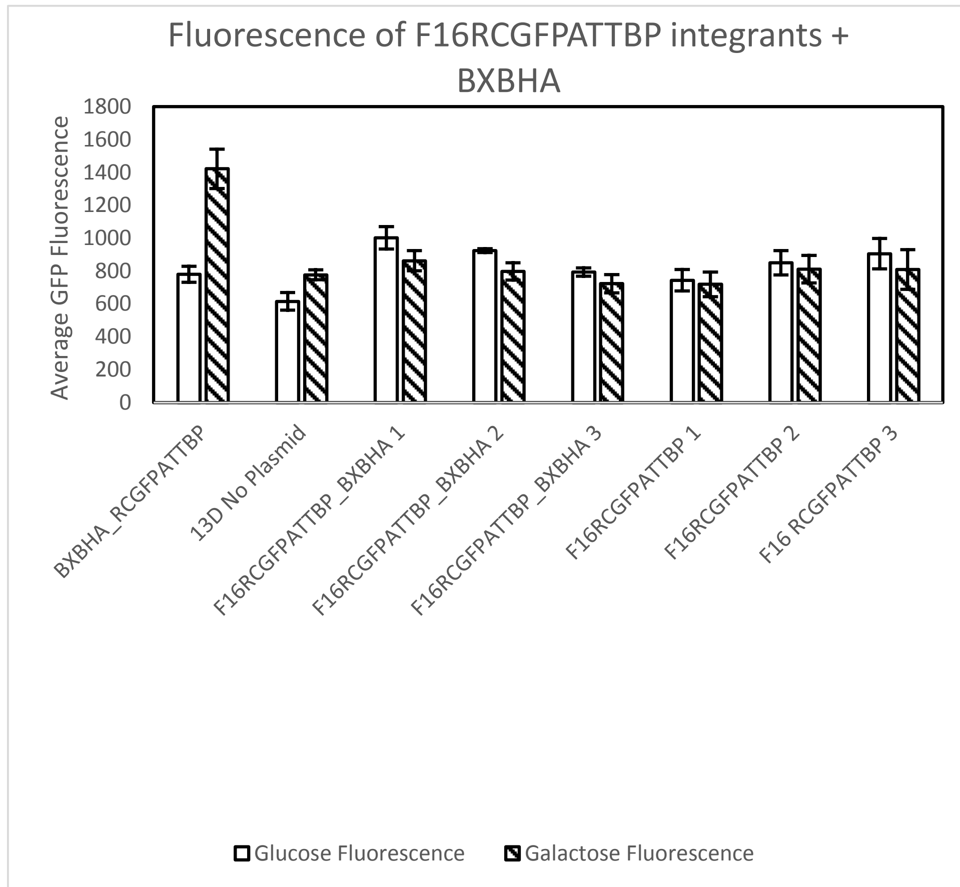


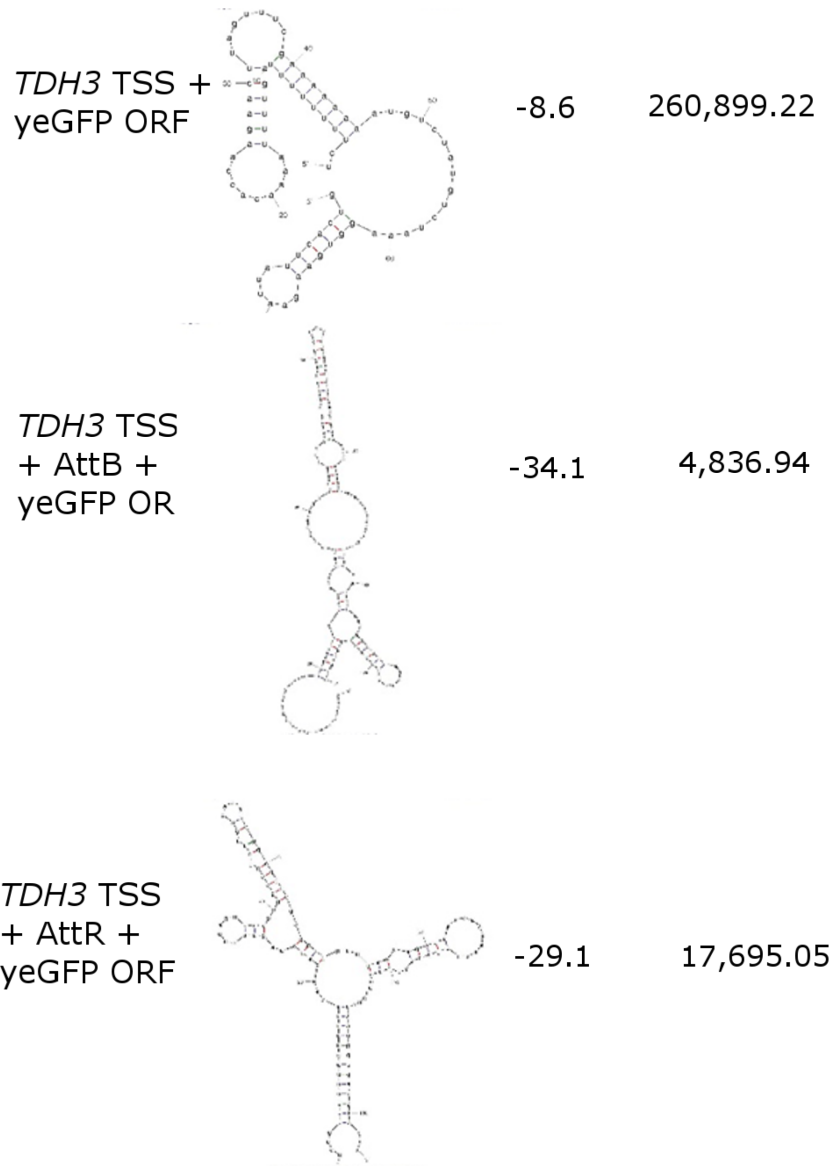
Figure 3-11: Fluorometry of integrated RCGFPATTBP cassette compared to plasmid borne RCGFPATTBP cassette. TECAN fluorometry of CEN.PK 13D integrated with pRCGFPATTBP as well as those putative integrants also bearing the pBXBHA plasmid. Shown also is TECAN fluorometry of plasmid borne pRCGFPATTBP and pBXBHA.

CHAPTER 4: DISCUSSION

4.1: IMPACT OF ATT SITES ON YEGFP EXPRESSION

If DNA inversion is to be a practical method for gene expression control for the purposes of synthetic biology or metabolic engineering, it would be ideal if the sites themselves didn't interfere with the production of the gene that they are meant to regulate. Unfortunately in the case of the BXBInt Att sites as they are configured in the

assays of this thesis, there is such interference. The reduction of expression that comes from introducing recombination sites between the coding sequence and the regulatory elements is not surprising in the light of the scanning model of eukaryotic translation, reviewed in Kozak.[66] In brief, translation of a gene is initiated when eukaryotic initiation factors (eIFs) bind to the 5’ 7-methyllguanosine cap and sometimes the 5’ Untranslated Region (5’ UTR) of the mRNA transcript, thus recruiting the ribosome to the 5’ UTR where it will “scan” the transcript until it encounters a start codon which will truly initiate protein synthesis.[66] It would stand to reason that adding extraneous (from the perspective of protein synthesis) sequence between the Translation Start Site (TSS) and the start codon may interfere with the rate of protein synthesis from a transcript. It is hypothesized that the low levels of fluorescence deriving from the YEGFPATTBP and YEGFPATTRL strains is the result of increased stability of RNA hairpins between the TSS and the start codon. Analysis of the secondary structure for the TSS of *TDH3*, the AttB and AttR sites, and the first 80 bases of the yeGFP ORF was undertaken using the mFold RNA folding prediction software (unafold.rna.albany.edu). Figure 4-1 shows that the AttB site forms a highly stable hairpin structure, and the AttR site is also predicted to form a slightly less energetically favourable hairpin, although still significantly greater than solely the *TDH3* TSS and first 80 bp of the predicted pTDH3YEGFP transcript.



Construct	*mFOLD predicted Structure*	*ΔG of Structure Formation*	*Average GFP fluorescence (AU)*
TDH3 TSS + yeGFP ORF		-8.6	260,899.22
TDH3 TSS + AttB + yeGFP OR		-34.1	4,836.94
TDH3 TSS + AttR + yeGFP ORF		-29.1	17,695.05

Figure 4-1: RNA secondary structure of Att Sites. mFold analysis and free energy of the predicted 5' Untranslated Region (UTR) and first 40 nt of the ORF of the pTDH3YEGFP, pYEGFPATTBP, and pYEGFPATTRL predicted transcripts. There is a correlation between hairpin stability and decreased yeGFP fluorescence, which may explain the results obtained in Section 3.1.

Previous work with engineered expression systems shows that interposing sequences with a high degree of secondary structure between promoter elements and ORFs in the 5' UTR of *S. cerevisiae* transcripts attenuates the activity of reporter constructs dramatically irrespective of transcript abundance,[67] in relation to the thermodynamic stability of the secondary structure formed.[68] Increasing the secondary structure of sequences immediately upstream of the AUG start codon have the greatest effect on protein abundance, as these secondary structures are likely to impede ribosomal scanning and initiation of peptide synthesis.[69]

The secondary structures of the AttB and AttP recognition sequences of the BXB integrase may be problematic for its use as a tool for metabolic engineering. The attenuation of protein levels compared to the pTDH3YEGFP control is significant, measured at a ten-fold reduction in GFP fluorescence for AttR/L and 100-fold reduction of GFP expression for AttB/P. One might easily consider based on classical Michaelis-Menten kinetics that lowering the concentration of an enzyme within a cell would lower the V_{max} of the reaction which that enzyme catalyzes, and thus reducing the overall rate of that reaction. This could be problematic for metabolic engineering, which aims to increase the rate of formation of some valuable metabolic product. Adding Att sites immediately upstream of a weak promoter might be even more problematic, as the abundance of protein may be reduced to levels too low for efficient enzymatic

conversion. However, optimizing the rate of formation of the product of a multi-enzyme pathway does not always require high rates of reaction at every step in that pathway. Reactions which form toxic by-products or which are susceptible to product inhibition need not be upregulated. Metabolic engineering depends on having a wide range of regulatory elements available to optimize carbon flux by fine-tuning the expression levels of multiple enzymes within a pathway. Adding the BXB integrase Att sites immediately upstream of a promoter may not be appropriate in situations where very high levels of protein are required for optimization, but may not have a deleterious effect if the BXB-regulated step does not require a high rate of reaction. It may also be possible that taking into account the attenuation of expression that comes from interposing Att sites between promoter and ORF would allow for predictable metabolic engineering. It may even be desirable. It has been suggested that increasing mRNA secondary structure immediately upstream of an ORF by adding leader sequences would be a method to predictably fine-tune gene expression, as modulating the stability of mRNA secondary structure would allow for more predictable expression modulation than saturation mutagenesis of promoter elements.[70] If the effects of mRNA secondary structure on protein levels were quantified for a particular Att site, then experiments which depend on quantifying protein concentrations might not necessarily be incompatible with serine integrase-mediated gene control.

It may be possible to recover the ablation of expression caused by interpolating the Att sites between the *TDH3* promoter and the following coding sequence. The argument put forward in this thesis and in other work involved with inserting sequences between promoters and reporter coding sequence initiation sites[68] is that this inhibition

of gene expression occurs at the level of translation. This hypothesis would state that increased secondary structure in the 5'UTR immediately upstream of the translation initiation site interferes either with the translation pre-initiation complex scanning the transcript and recognizing translational initiation sites, or interferes with the ribosome binding to the transcript at the correct initiation site. In either case, gene expression levels may be recovered to the same level as the wild-type promoter by adding an additional Internal Ribosome Entry Site (IRES) such as the YAP1 yeast IRES[71] or by adding additional Kozak sequences between the Att site and the translation initiation site so that the linear order of regulatory elements would be Promoter-Att Site-IRES/Kozak sequence-Coding Sequence rather than Promoter-Att Site -Coding Sequence. If this supposition about the inhibitory effects of Att sites immediately upstream of the GFP coding sequence in the pRCGFPATTBP construct is correct, it may be possible to utilize serine integrase mediated inversion *in vivo* without reducing gene expression levels of the regulated gene. While this experiment raises some troubling questions about the use of BXBInt as a tool for gene expression control and metabolic engineering, the levels of fluorescence from the YEGFPATTRL strain indicate that it is possible to detect fluorescence from an inverted yeGFP ORF as would occur if pRCGFPATTBP was inverted by BXBInt.

4.2: ACTIVATION OF EXPRESSION BY DNA INVERSION

Gene activation mediated by DNA inversion is one of the main criteria for success of this enzyme's function in *S. cerevisiae.* Developing a system of gene expression control based on genetic memory requires this function. It was shown in

Section 3.4 that the expression of a yeGFP reporter cloned into a plasmid in the reverse compliment orientation with respect to a promoter could be activated upon induction of expression of BXBInt in *S. cerevisiae*. This indicates that the DNA of the yeGFP ORF was successfully inverted in the plasmid borne construct, which agrees with the findings from a similarly assembled construct expressed in *E. coli.*[30] This is a welcome result for the prospect of using this recombinase class in eukaryotic cells, and its success was not necessarily a foregone conclusion. The main possible barrier to the function of BXBInt in *S. cerevisiae* is the nucleus of the cell. Given that maintenance and replication of plasmid DNA is localized to the nucleus and protein synthesis occurs in the cytoplasm, it may have been possible that BXBInt would not have access to plasmid borne Att sites sequestered in the nucleus. While experiments have shown that BXBInt could catalyze integration into and deletion from chromosomal DNA in other eukaryotic species,[42,72] DNA inversion catalyzed by BXBInt has not been previously demonstrated. The fact that there is no significant increase in the capacity for BXBInt to catalyze inversion when given C-terminal Nuclear Localization Sequences further supports the idea that the nucleus provides no barrier between this recombinase and its DNA substrates.

The results of the time course experiment shown in Figure 3-4 provide some interesting insights into the kinetics of BXBInt's inversion activity in yeast. Expression of GFP is undetectable by flow cytometry an hour after induction, with FL1 values being equal to both BXBHA strain and the RCGFPATTBP strain. The expression of GFP is detectable at 3 hrs, and reaches a maximum value at 5 hrs. Twenty-four hrs after transfer to the induction medium the value of detectable GFP fluorescence decreases slightly. This decrease in fluorescence is likely insignificant and could be attributed to

loss of the plasmid amongst some individuals within the culture as opposed to any alteration of the expression levels of the reporter construct itself. Results in Figure 3-5 shows the same experiment over a longer timescale, where GFP levels were measured at 24 and 72 hrs of batch culture, after which cultures were back diluted 1/5000 in fresh media and allowed to grow for a further 24 hrs. The fluorescence of the galactose induced BXBHA_RCGFPATTBP culture shows a slight reduction in GFP expression at each time point, although the average GFP levels of each time point fall within the standard deviation of the measured replicates. It can be concluded that the expression level of GFP of inverted reporter constructs is stable over the course of a multiple-day small-scale batch culture. The fact that GFP fluorescence remains constant after back dilution into fresh media and overnight growth also shows that the inverted plasmid can be maintained in its inverted state over the course of several generations in the presence of the inducer. This is a demonstration of the heritability trait of the serine integrase inversion system, that there is no alteration in expression of inverted DNA after being allowed to replicate and pass from mother to daughter cell.

4.3: DEACTIVATION OF GENE EXPRESSION BY BXBINT MEDIATED INVERSION OF YEGFPATTBP

If DNA Inversion is to function as a truly bi-directional and multipurpose tool for gene expression control, it should be possible to activate gene expression using DNA inversion and also to completely de-activate it. DNA inversion should proceed in the same manner regardless of what the starting orientation of the DNA is. To test for gene de-activation, constructs were developed which would test the ability of BXBInt

inversion to de-activate gene expression. The results from Section 3.3, showed that expression of a yeGFP coding sequence regulated by the BXBInt AttB and AttP sites could not be turned off to any significant degree by BXBInt mediated inversion. It should be noted that this form of inversion has not been demonstrated in other works describing the activities of LSTRs. Even after a multiple day induction of the BXBInt plasmid, yeGFP expression was maintained at the same level as strains lacking the BXBInt plasmid and strains that were not induced with galactose. Moreover, average yeGFP were maintained even after cultures were back diluted into fresh media and were allowed to grow exponentially for 24 hrs, excluding the possibility that yeGFP expression had been turned off but that the yeGFP protein had simply not degraded over the course of the experiment. Looking at the percentage of cells that display GFP fluorescence above the background level shows a small decrease in the number of cells which have no significant GFP fluorescent compared to the same reporter construct without the integrase. This could be seen as a decrease of gene expression caused by the inversion of the YEGFPATTBP reporter construct, but it is not nearly as clear cut as an effect as is seen when gene expression is turned on by DNA inversion as is seen in the strain BXBHA_RCGFPATTBP. The latter shows an increase in the number of fluorescent cells from zero percent to over fifty percent when integrase expression is induced. This means that over half of the cells that were “Off” in terms of GFP expression became “On” due to the integrase. It would be expected that for the strain BXBHA_YEGFPATTBP at least half of the cells that start out as “On” be turned “Off” due to DNA inversion. It is hard to say whether this effect was observed based solely on GFP fluorescence due to the large variation between biological replicates and

considering the confounding variable of GFP fluorescence being less in the YEGFPATTBP strain (which lacks the integrase plasmid) when grown in galactose compared to glucose.

If DNA inversion does occur in the strain BXBHA_YEGFPATTBP, the assay presented in this thesis is not capable of observing it. A logical next step to detect DNA inversion would be to look at the physical structure of the reporter construct plasmid. It would be possible to design an assay based on DNA sequencing, quantitative PCR, restriction mapping, or Single Nucleotide Polymorphism analysis of the reporter constructs would allow for quantification of physical DNA inversion events in the case where only a minority of a cellular population actually experiences DNA inversion. Without the data from these assays, all that can be said now is that the collective effect on the gene expression of a population of cells, is much less when DNA is inverted from “On” to “Off” than it is when DNA is inverted from “Off” to “On”, and this asymmetry of function needs an explanation. It is the argument of the author that this explanation is that the inversion mediated by BXBInt does not occur to the same extent when activating and de-activating gene expression.

It could be a possibility that DNA inversion is only occurring in a small proportion of the cells being tested in these assays. With this in mind Figure 3-8 illustrates the percentage of individual cells that show GFP fluorescence above background levels. This figure shows the same pattern as that seen in Figure 3-7, showing the average GFP fluorescence of the population of cells. There appears to be a general decrease in GFP fluorescence when YEGFPATTBP is grown in galactose for 72 hours and then grown for an additional 14 hours in glucose (66.7% to 53.54%, a 13% decrease) when

compared to that strain grown glucose for 72 hours and then back diluted into glucose. The decrease in fluorescence is also seen in BXBHA_YEGFPATTBP (62.3% to 32.1%, a 30% decrease), and the decrease is slightly more than that seen in YEGFPATTBP. The difference in the decrease of fluorescing cells is around 17%. It is difficult to say whether this represents an actual incidence of DNA inversion causing the de-activation of GFP expression. The variation in the proportion of fluorescent cells between replicates is quite large, such that the gap between the upper error bar of BXBHA_YEGFPATTBP Galactose 96 hour back diluted and the lower error bar of YEGFPATTBP Galactose 96 hour back diluted is 11%. This would be the minimum proportion of cells whose decrease in fluorescence could be attributed to DNA inversion mediated by BXBInt. This figure seems low, considering that when BXBHA_RCGFPATTBP is grown in glucose and galactose, the proportion of cells that fluoresce above background goes from 0% to over 50% within 24 hours. This latter fact seems to indicate that DNA inversion occurs in at least 50% of the cells that bear the integrase and reporter construct plasmids. If DNA inversion occurred at the same frequency in the strain BXBHA_YEGFPATTBP, the difference in the decrease of fluorescing cells should be at least half of the number of cells that fluoresce above background without the presence of BXBInt. In other words, if 70% of YEGFPATTBP cells fluoresce when grown in glucose under ideal conditions, the difference between any decrease in fluorescence between YEGFPATTBP and BXBHA_YEGFPATTBP should be at least 35%. Since the observed difference is between 11% and 17% over the course of 96 hours, it would be premature to conclude that any decrease in fluorescence was caused by BXBInt expression. In any case, looking at the average

GFP fluorescence in figure 3-7 shows that the bulk effect of any DNA inversion is minimal at best.

The most likely explanation for the lack of conclusive is that the pYEGFPATTBP plasmid was unmodified by BXBInt in this experiment. This finding is surprising given the previous research into BXBInt and other LSTRs, which have identified the presence of AttB and AttP sites as sufficient conditions for inversion. There are a number of possible explanations as to why this may have happened. Firstly, there may be some experimental error preventing BXBInt mediated inversion. Although the pYEGFPATTBP plasmid was sequenced prior to being transformed into the double auxotroph alongside the BXBInt plasmid, and the BXBInt plasmid was used to transform the pRCGFPATTBP plasmid bearing strain as a positive control for plasmid inversion, it is possible that a rare mutation altered one of the Att sites. It is possible that recombination occurred between the two plasmids, which may have altered the *GAL1* regulated BXBInt cassette. Random mutation seems unlikely, as the same phenomenon was observed in multiple biological replicates of BXBHA_YEGFPATTBP, and lack of expression of the BXBInt protein also seems unlikely, as the same plasmid was able to cause recombination in the pRCGFPATTBP bearing strain. There may be a mechanistic reason why this inversion was unable to occur. It could be that elements of transcription initiation factors and RNA polymerase binding to plasmid DNA interferes with BXBInt binding to the Att sites in such a way that only occurs when AttB is directly proximal to the start codon of yeGFP. Inversion occurs when AttB is distant from any possible start codon, as is the case when the reporter cassette of pRCGFPATTBP is inverted, and transcription and translation of the reporter ORF occurs when AttR is proximal to the

start codon, as is the case when RCGFPATTBP has been inverted. However, any proteins that bind to the pYEGFPATTBP construct in the region of the Att Sites in order to inhibit BXBInt, such as transcription factors, which bind to the *TDH3* promoter or *ENO2* terminator would also bind to those regulatory elements of the RCGFPATTBP reporter cassette without inhibiting inversion. This molecular crowding hypothesis might not provide a completely coherent explanation for the asymmetry between the two kinds of inversion activity, but it is clear that if the inability of BXBInt to invert DNA that is being actively transcribed is not an experimental error or caused by a rare mutation, that it is a major drawback for the use of this enzyme as a universal tool for genetic engineering.

4.4*: INTEGRATION OF RCGFPATTBP*

Integration of the reporter cassette from pRCGFPATTBP into the yeast chromosome also poses difficulties for the use of BXBInt as a genetic engineering tool. It would be a useful tool for synthetic biologists and metabolic engineers to re-arrange the structure of chromosomal DNA *in vivo* in a dynamic manner, and possibly control gene expression through this mechanism. There are a number of possible explanations as to why BXBInt mediated inversion of chromosomally integrated DNA could not be observed. The first plausible explanation is failure to successfully integrate the F16RCGFPATTBP integration construct into Flagfeldt site 16. There is evidence to suggest that the RCGFPATTBP cassette was integrated successfully, but this evidence is not absolutely conclusive. The design of the integration cassette was based on homologous recombination of four amplified fragments of DNA, which had 40 base-pair

regions of homology to their adjacent fragments. The sequence of amplicons is structured thusly: Site 16 homology region Up-G418 cassette-RCGFPATTBP cassette-Site 16 homology region Down. Negative control transformations which contained three of these four amplicons were sensitive to G418, indicating that integration cannot occur unless all four amplicons are present in order to bridge the two chromosomal homology regions into one cohesive integration construct. Moreover, a parallel integration where the pRCGFPATTBP cassette was replaced by the pTDH3YEGFP expression cassette with the same linkers joining it to the G418 cassette and chromosomal homology region was resulted in G418 resistant colonies, and fluorometry of these colonies showed high levels of yeGFP expression after two rounds of re-streaking in a large proportion of colonies tested. Thus there is circumstantial evidence that acquiring the G418r cassette also means that the pRCGFPATTBP cassette has also been integrated. This, alongside the pattern of amplicons obtained from PCR analysis of the integration site being what would be expected from a successful integration of the pRCGFPATTBP construct, provides strong evidence to suggest that the construct integrated successfully but failed to invert.

The lack of evidence for inversion of chromosomally integrated Att sites is a major impediment to the use of the BXB Integrase as a mechanism for genetic memory in yeast. This result, if correct, would exclude the use of this enzyme from being used to establish a genetic memory based expression system on the chromosome, which would be the ideal location for experiments seeking to establish stable genetic expression for the purposes of metabolic engineering, pathway optimization, or large-scale fermentation. Combined with the selective activity towards plasmid based DNA, the

options towards which this enzyme could be used in *S. cerevisiae* appear to be more limited than they are for bacterial strains.

CHAPTER 5: CONCLUSIONS

ACTIVITY OF BXBINT IN SACCHAROMYCES CEREVISIAE.

The expression of BXBInt in *S. cerevisiae* was shown to catalyze dynamic, *in vivo* inversion of the yeGFP reporter cassette in this study. This is a useful result for those interested in utilizing this enzyme for genetic manipulation in *S. cerevisiae,* but there are certainly questions that remain to be answer regarding the heterologous expression of serine recombinases. The mycobacteriophage BXB1 integrase was selected for this study in part due to the fact that it had been previously shown to catalyze integration when expressed heterologously in eukaryotic organisms, [39,40,42] and inversion of DNA when expressed heterologously in *E. coli.*[30] However, Xu *et al*. made the observation that although they screened 15 large serine-type recombinases for integration activity in human cell lines, not all of these recombinases showed the same rate of recombination or specificity of integration. While the site-specific and orthogonal nature of Integrase/Att site recombination provide the promise for multiplex simultaneous recombination events in *S. cerevisiae,* this thesis failed to investigate other members of this diverse family of enzymes to compare their ability to catalyze DNA inversion, or investigate the possibility of multiplex simultaneous *in vivo* inversion invents. This property of serine integrases remains to be exploited to its full potential.

RESTRAINTS ON BXBINT INVERSION

This thesis provides some promising insights into the activity of BXBInt in *S. cerevisiae*, but also highlights some significant restraints on its potential use for genetic engineering of protein expression systems. Serine integrase Att sites can be easily adapted to homology based cloning platforms such as DNA Assembler or Gibson Assembly. It seems as though nuclear localization provides no barrier to the binding of the enzyme to its cognate Att sites. And it was demonstrated that the activity of DNA inversion could occur through dynamic induction in living cells, with no noticeable background expression of the Att-regulated reporter. These properties of the BXBInt inversion system constructed for this thesis are desirable for synthetic biologists interested in *in vivo* genetic manipulation of eukaryotes

However, there are detriments of this system, which must be addressed in order to optimize serine integrase based gene expression control. Quantification of the expression of a yeGFP reporter regulated by the BXBInt Att sites immediately adjacent to the reporter ORF showed a significant reduction of expression, probably caused by increased secondary structure of the mRNA transcript of this reporter preventing efficient translation. This is a significant drawback to the use of this recombination system as it was constructed in this study, as any application, which relied on expression of an integrase-regulated protein would obviously be negatively impacted. It may be possible that by adjusting the location of Att sites with respect to other elements of an expression cassette may restore the expression levels of the regulated protein. If the proposed hypothesis that increased secondary structure in mRNA transcripts prevents efficient recognition of start codons by the ribosome is correct, it would be

reasonable to suggest that adding spacer sequences between the Att site and start codon would increase the amount of functional protein within the cell. Studying the effects of Att site location within a reporter cassette would be a reasonable next step in the development of BXBInt as a tool for gene expression control

This thesis also made an interesting observation as to the directionality of DNA inversion catalyzed by BXBInt. While it was demonstrated that gene expression could be activated by DNA Inversion, deactivation of gene expression catalyzed by DNA inversion of a yeGFP cassette from a translationally active orientation to the reverse compliment orientation was not observed over the course of several days of exponential growth. This could be an indication that the context of the sequences adjacent to the Att sites are relevant to the activity of the recombinase. There may be some element of the ORF used in this study that inhibits recombination between AttP and AttB when placed in the “forward” orientation with respect to the promoter and these two sequences. The structure of the RCGFPATTBP reporter cassette and the proposed mechanism for mRNA transcription in yeast implies that is that the yeGFP mRNA is still transcribed when in the reverse compliment orientation, but not in an orientation that would result in translation of a functional yeGFP protein. After inversion, the yeGFP mRNA is transcribed with the ORF in a translationally active orientation. For the FYEGFPATTBP plasmid, it is clear that the yeGFP reporter is both transcribed and translated to form an active yeGFP, and would potentially transcribe an mRNA after inversion with a translationally inactive reverse compliment yeGFP ORF. There should be no difference in the ability of BXBInt to access the AttP and AttB sites between both plasmids. It could be that the forward orientation sequence of the yeGFP ORF adjacent to the Att sites

recruits some DNA binding proteins to the plasmid which exclude the Integrase from accessing the Att sites which do not bind to the reverse compliment yeGFP cassette adjacent to the Att sites. Using alternate reporters in place of yeGFP may support or refute this hypothesis, or attempting to co-precipitate these DNA sequences and DNA binding proteins. In any case, it can be said that DNA inversion by BXBInt may be dependent on the context of the sequence in which the Att sites are situated. More investigation is needed in order to develop a BXBInt as a truly versatile tool for dynamic gene expression or knockout.

INVERSION OF CHROMOSOMAL DNA

The lack of evidence for inversion of chromosomally integrated Att sites is a further impediment for the development of BXBInt as a tool for genetic engineering. The higher-ordered structure and binding of structural proteins of genomic DNA compared to plasmid DNA functioning to prevent the efficient formation of an invertasome-like loop of DNA during DNA inversion may be a reasonable explanation for the lack of BXBInt activity. Engineering chromosomal DNA is an attractive strategy for large scale biotechnology applications, as the stability of integrated constructs is much greater than plasmid-borne constructs in large scale fermentation. These considerations together would significantly undermine the stated goal of this thesis, of developing a BXBInt-mediated DNA inversion system as a form of genetic memory for the induction of genes in large-scale fermentation. However, if the hypothesis around the lack of inversion observed in the experiments described in Section 3.4 were to be shown invalid, then it might be possible to engineer DNA inversion of chromosomal DNA in *S. cerevisiae.*

OVERALL CONCLUSIONS

The experiments and observations described in this thesis highlight some limitations for the use of the BXB1 Integrase protein for synthetic biology and genome engineering in *S. cerevisiae.* This thesis represents the first published attempts to both activate and de-activate gene expression via DNA inversion, as well as the first attempts to invert chromosomal DNA of a eukaryotic organism, both of which were unsuccessful. It can be concluded from this these experiments that the universal application of serine integrases to perfectly perform complex, dynamic DNA manipulations may not be straightforward, and may require significant additional study and engineering if the goal of using DNA inversion as a form of genetic memory or as a device for synthetic biology is to be achieved using this enzymes or others like it. However, these experiments also suffer from significant limitations, namely that only one reporter and Att site construct was tested and only one Large Serine-type Recombinase family member was used to catalyze inversion. It is therefore quite difficult to draw definite conclusions about the activity, or lack thereof, of LSTRs and DNA inversion. The observation of activated gene expression of plasmid-based DNA in yeast is an encouraging result, but it is not enough to unreservedly recommend the use of BXBInt as a tool for multiplex dynamic gene expression control in this organism. Hopefully, the results of the experiments described in this thesis will serve to provide a starting point for further investigation into the use of this fascinating class of enzymes to enable DNA inversion in a greater number of contexts and applications.

REFERENCES

1. Lu, T. K., Khalil, A. S. & Collins, J. J. Next-generation synthetic gene networks. *Nat. Biotechnol.* **27**, 1139–1150 (2009).
2. Stricker, J. et al. A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–9 (2008).
3. Daniel, R., Rubens, J. R., Sarpeshkar, R. & Lu, T. K. Synthetic analog computation in living cells. *Nature* **497**, 619–623 (2013).
4. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J. & Dueber, J. E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.* **41**, 10668–10678 (2013).
5. da Silva, N. A. & Srikrishnan, S. Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **12**, 197–214 (2012).
6. Haynes, K. A. & Silver, P. A. Eukaryotic systems broaden the scope of synthetic biology. *J. Cell Biol*. **187**, 589–596 (2009).
7. Verwaal, R. et al. Heterologous carotenoid production in *Saccharomyces cerevisiae* induces the pleiotropic drug resistance. *Yeast* **27**, 983–998 (2010).
8. Ro, D. et al. Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. *BMC Biot* **14**, 1–14 (2008).
9. Jensen, P. R. & Hammer, K. Artificial promoters for metabolic optimization. *Biotechnol. Bioeng.* **58**, 191–195 (1998).
10. Weinhandl, K., Winkler, M., Glieder, A. & Camattari, A. Carbon source dependent promoters in yeasts*. Microb. Cell Fact*. **13**, 5 (2014).
11. Santos, C. N. S. & Stephanopoulos, G. Combinatorial engineering of microbes for optimizing cellular phenotype. *Curr. Opin. Chem. Biol*. **12**, 168–176 (2008).
12. Cox, R. S., Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3**, (2007).

13. Ramon, A. & Smith, H. O. Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol. Lett.* **33**, 549–555 (2011).
14. Kim, B., Du, J., Eriksen, D. T. & Zhao, H. Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels. *Appl. Environ. Microbiol*. **79**, 931–941 (2013).
15. Nowroozi, F. F. et al. Metabolic pathway optimization using ribosome binding site variants and combinatorial gene assembly. *Appl. Microbiol. Biotechnol.* **98**, 1567–1581 (2014).
16. Yuan, J. & Ching, C. B. Combinatorial engineering of mevalonate pathway for improved amorpha-4,11-diene production in budding yeast. *Biotechnol. Bioeng.* **111**, 608–617 (2014).
17. Venayak, N., Anesiadis, N., Cluett, W. R. & Mahadevan, R. Engineering metabolism through dynamic control. *Curr. Opin. Biotechnol*. **34**, 142–152 (2015).
18. Michener, J. K., Thodey, K., Liang, J. C. & Smolke, C. D. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab. Eng.* **14**, 212–222 (2012).
19. Stephanopoulos, G. & Simpson, T. W. Flux amplification in complex metabolic networks. *Chem. Eng. Sci*. **52**, 2607–2627 (1997).
20. Farmer, W. R. & Liao, J. C. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol* **18**, 533–537 (2000).
21. Callura, J. M., Cantor, C. R. & Collins, J. J. Genetic switchboard for synthetic biology applications. *Proc. Natl. Acad. Sci*. **109**, 5850–5855 (2012).
22. Zhang, F. & Keasling, J. Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol*. **19**, 323–329 (2011).
23. Pfleger, B. F., Pitera, D. J., Newman, J. D., Martin, V. J. J. & Keasling, J. D. Microbial sensors for small molecules: Development of a mevalonate biosensor. *Metab. Eng*. **9**, 30–38 (2007).
24. Rogers, J. K. et al. Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res.* 1–13 (2015).

25. Mahr, R. et al. Biosensor-driven adaptive laboratory evolution of L -valine production in *Corynebacterium glutamicum*. *Metab. Eng.* **32**, 184–194 (2015).
26. Mattozzi, M. D., Ziesack, M., Voges, M. J., Silver, P. a. & Way, J. C. Expression of the sub-pathways of the *Chloroflexus aurantiacus* 3-hydroxypropionate carbon fixation bicycle in *E. coli*: Toward horizontal transfer of autotrophic growth. *Metab. Eng*. **16**, 130–139 (2013).
27. DeLoache, W. C. et al. An enzyme-coupled biosensor enables (*S*)-reticuline production in yeast from glucose. *Nat. Chem. Biol*. **11**, 465–471 (2015).
28. Shiue, E. & Prather, K. L. J. Improving d-glucaric acid production from myo-inositol in *E. coli* by increasing *MIOX* stability and myo-inositol transport. *Metab. Eng.* **22**, 22–31 (2014).
29. Brockman, I. M. & Prather, K. L. J. Dynamic knockdown of *E. coli* central metabolism for redirecting fluxes of primary metabolites. *Metab. Eng*. **28**, 104–113 (2015).
30. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol*. **31**, 448–452 (2013).
31. Yang, L. et al. Permanent genetic memory with >1-byte capacity. *Nat. Methods* (2014).
32. Boyle, P. M. et al. Amplifying genetic logic gates. *Science*. **340**, 599–603 (2013).
33. Groth, A. C. & Calos, M. P. Phage Integrases: Biology and Applications. *J. Mol. Biol.* **335**, 667–678 (2004).
34. Hallet, B. & Sherratt, D. J. Transposition and site-specific recombination: Adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.* **21**, 157–178 (1997).
35. Ballantyne, J., Henry, D. L. & Marcu, K. B. Antibody class switch recombinase activity is B cell stage specific and functions stochastically in the absence of ‘targeted accessibility’ control. *Int. Immunol*. **9,** 963–974 (1997).
36. Smith, M. C. M. & Thorpe, H. M. Diversity in the serine recombinases. *Mol. Microbiol.* **44**, 299–307 (2002).

37. Brown, W. R. A., Lee, N. C. O., Xu, Z. & Smith, M. C. M. Serine recombinases as tools for genome engineering. *Methods* **53**, 372–379 (2011).
38. Pope, W. H. et al. Expanding the diversity of mycobacteriophages: Insights into genome architecture and evolution. *PLoS One* **6**, e16329 (2011).
39. Xu, Z. et al. Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol*. **13**, 87 (2013).
40. Huang, J., Ghosh, P., Hatfull, G. F. & Hong, Y. Successive and targeted DNA integrations in the *Drosophila* genome by Bxb1 and phiC31 integrases. *Genetics* **189**, 391–5 (2011).
41. Hou, L. et al. An open-source system for *in planta* gene stacking by Bxb1 and Cre recombinases. *Mol. Plant* **7**, 1756–1765 (2014).
42. Blechl, A., Lin, J., Shao, M., Thilmony, R. & Thomson, J. The Bxb1 recombinase mediates site-specific deletion in transgenic wheat. *Plant Mol. Biol. Report*. **30**, 1357–1366 (2012).
43. Shao, M., Kumar, S. & Thomson, J. G. Precise excision of plastid DNA by the large serine recombinase Bxb1. *Plant Biotechnol. J*. **12**, 322–329 (2014).
44. Thomson, J. G., Chan, R., Thilmony, R., Yau, Y.-Y. & Ow, D. W. PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the *Arabidopsis* genome. *BMC Biotechnol*. **10**, 17 (2010).
45. Russell, J. P., Chang, D. W., Tretiakova, A. & Padidam, M. Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *Biotechniques* **40**, 460–464 (2006).
46. Kim, A. I. et al. Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis* groEL1 gene. *Mol. Microbiol*. **50**, 463–473 (2003).
47. Mediavilla, J. et al. Genome organization and characterization of mycobacteriophage Bxb1. *Mol. Microbiol*. **38**, 955–970 (2000).
48. Ghosh, P., Kim, A. I. & Hatfull, G. F. The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol. Cell* **12**, 1101–1111 (2003).

49. Ghosh, P., Pannunzio, N. R., Hatfull, G. F. & Gottesman, M. Synapsis in phage Bxb1 integration: Selection mechanism for the correct pair of recombination sites. *J. Mol. Biol*. **349**, 331–348 (2005).
50. Singh, S., Ghosh, P. & Hatfull, G. F. Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination. *PLoS Genet*. **9**, e1003490 (2013).
51. Ghosh, P., Wasil, L. R. & Hatfull, G. F. Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol*. **4**, 0964–0974 (2006).
52. Thomson, J. G. & Ow, D. W. Site-specific recombination systems for the genetic manipulation of eukaryotic genomes. *Genesis* **44**, 465–476 (2006).
53. Bruist, M. F. & Simon, M. I. Phase variation and the Hin protein: *In vivo* activity measurements, protein overproduction, and purification. *J. Bacteriol.* **159**, 71–79 (1984).
54. Johnson, R. C. & Bruist, M. F. Intermediates in Hin-mediated DNA inversion: a role for Fis and the recombinational enhancer in the strand exchange reaction. *EMBO J*. **8**, 1581–1590 (1989).
55. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci.* **109**, 8884–8889 (2012).
56. Oberdoerffer, P., Otipoby, K. L., Maruyama, M. & Rajewsky, K. Unidirectional Cre-mediated genetic inversion in mice using the mutant loxP pair lox66/lox71. *Nucleic Acids Res*. **31**, e140 (2003).
57. Fossati, E. et al. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in *Saccharomyces cerevisiae*. *Nat. Commun.* **5**, 3283 (2014).
58. de Wulf, P., Brambilla, L., Vanoni, M., Porro, D. & Alberghina, L. Real-time flow cytometric quantification of GFP expression and GFP-fluorescence generation in *Saccharomyces cerevisiae*. *J. Microbiol. Methods* **42**, 57–64 (2000).

59. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* **37**, 1–10 (2009).
60. Jansen, G., Wu, C., Schade, B., Thomas, D. Y. & Whiteway, M. Drag&Drop cloning in yeast. *Gene* **344**, 43–51 (2005).
61. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc*. **2**, 31–34 (2007).
62. Moore, S. P., Rinckel, L. A. & Garfinkel, D. J. A Ty1 integrase nuclear localization signal required for retrotransposition. *Mol. Cell. Biol.* **18**, 1105–1114 (1998).
63. Makkerh, J. P. S., Dingwall, C. & Laskey, R. A. Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. *Curr. Biol*. **6**, 1025–1027 (1996).
64. Flagfeldt, D. B., Siewers, V., Huang, L. & Nielsen, J. Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* **26**, 545–551 (2009).
65. Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A. & Voytas, D. F. Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.* **8**, 464–478 (1998).
66. Kozak, M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**, 1–34 (2002).
67. Laso, M. R. V et al. Inhibition of translational initiation in the yeast *Saccharomyces cerevisiae* as a function of the stability and position of hairpin structures in the mRNA leader. *J. Biol. Chem*. **268**, 6453–6462 (1993).
68. Oliveira, C. C., van den Heuvel, J. J. & McCarthy, J. E. Inhibition of translational initiation in *Saccharomyces cerevisiae* by secondary structure: the roles of the stability and position of stem-loops in the mRNA leader. *Mol. Microbiol*. **9**, 521–532 (1993).

69. Dvir, S. et al. Deciphering the rules by which 5’-UTR sequences affect protein expression in yeast. *Proc. Natl. Acad. Sci. U. S. A*. **110**, E2792–E2801 (2013).
70. Lamping, E., Niimi, M. & Cannon, R. D. Small, synthetic, GC-rich mRNA stem-loop modules 5′ proximal to the AUG start-codon predictably tune gene expression in yeast. *Microb. Cell Fact*. **12**, 74 (2013).
71. Zhou, W., Edelman, G. M. & Mauro, V. P. Transcript leader regions of two *Saccharomyces cerevisiae* mRNAs contain internal ribosome entry sites that function in living cells*. Proc. Natl. Acad. Sci. U. S. A.* **98**, 1531–6 (2001).
72. Huang, J., Ghosh, P., Hatfull, G. F. & Hong, Y. Successive and Targeted DNA Integrations in the *Drosophila* Genome by Bxb1 and C31 Integrases. *Genetics* **189**, 391–395 (2011).