Automated BioPart Characterisation for Synthetic Biology

Christopher David Hirst

Imperial College London

Department of Bioengineering

Centre for Synthetic Biology and Innovation

Submitted for Degree of Doctor of Philosophy

February 2014

Declaration

All work and data contained within this thesis is my own except where referenced.

Acknowledgments

I would like to thank my primary supervisor Richard Kitney for the many helpful discussions and support during this project, allowing me to grow as a researcher. His tireless belief in the future for engineering biology has been an inspiration. I would particularly like to thank him for the many opportunities including conferences, events and experiences that I have benefited so much from but most importantly the opportunity to work on this project.

I also greatly appreciate the guidance of my other supervisors Geoff Baldwin and Paul Freemont for many useful discussions about the technology in development and overall direction of the project. For my supervisors collectively I must also express my thanks for selecting me to compete as part of the Imperial iGEM team in 2008 and supervise the team in 2010, those experiences have taught me much about myself, the communication of science and re-enforced my decision to become a researcher.

I appreciate the many useful discussions with Dr. Tom Ellis regarding the initial design for the datasheets. Special thanks are also reserved for Kirsten Jensen for her many years of imparting knowledge and useful discussions and to Dr. Matthieu Bultelle for continually challenging my assumptions and always requiring more (either data or biological knowledge).

To my colleagues in the CSynBI and the Department of Bioengineering, I thank you for all the interesting debates and scientific discussions as well as helping me to cope with the many ups and downs we all share as scientists as part of our projects. My appreciation in particular goes out to the members of the CSynBI who used the automated platform while it was being set up and were tolerant of its occasional bugs and errors and always pushed what was required from the platform. Those occasional failed runs were vital in the identification of the small bugs that would otherwise have remained a mystery.

I would like to thank the EPSRC for providing the scholarship funding to carry out this research. Additionally I would like to thank the SynBioStandards Network for the travel grant that allowed me to attend the SB5.0 conference and meet with the members of the BioFAB in Stanford.

To my family and friends collectively, the deepest appreciation goes out to you for being a constant source of support and inspiration during the length of this project.

1 Abstract

Synthetic Biology is an approach to the development of biological systems based on engineering principles. By using concepts from other engineering disciplines such as abstraction it is possible to break down complicated biological functions into components termed 'BioParts'. BioParts can be assembled collectively into modules and systems to carry out advanced functions, designed from the bottom up. A key part of this approach is the standardisation of BioParts and practices to aid design, testing and implementation.

An automated characterisation methodology focused primarily on promoter BioParts has been developed which is potentially scalable to other BioPart families. The standardised workflow is optimised to enable BioPart characterisation under highly reproducible growth conditions, reliably producing high quality data. It has been designed around automation equipment which should ensure accurate reproduction of experiments at other sites.

The automated characterisation workflow has been demonstrated to produce high quality data for both constitutive and inducible promoters. The entire Anderson promoter collection has been characterised and high details results for all library members are available for the first time. A pair of inducible promoter BioPart were characterised to obtain a deep data level regarding their activity in response to inducer over time. To allow the characterisation of more inducible BioParts in a shorter period of time, promoter engineering was also used to generate novel promoters which are induced by xylose.

The development of the automated workflow should be a step towards the standardisation of characterisation protocols and production of large numbers of BioParts with associated high quality, reproducible characterisations. Standardisation will further aid the comparison of the data sets produced, potentially shining light on unknown interactions between BioParts and their environment and improving the ability of Synthetic Biologists to design novel biological systems from the ground up.

Contents

List of tables

List of figures

2 Introduction

2.1 Lay Summary

The Royal Academy of Engineering (2009) suggested a definition for Synthetic Biology to be:

'Synthetic Biology aims to design and engineer biologically based parts, novel devices and systems as well as redesigning existing, natural biological systems.'

To enable Synthetic Biology to reach these aims concepts common to many other engineering fields have been applied. One of the most critical differences between Synthetic Biology and previous attempts to build biological systems has been the application of the engineering principles of modularity, characterisation and standardisation and the emphasis on systematic design. Many efforts worldwide have been carried out to test and document the basic 'parts' from which biological systems can be composed in an attempt to replicate the standardisation of other engineered parts such as bolts and screw threads to allow their re-use. However much of the work to understand how these basic parts function has not being carried out in a standard way or using standard materials.

While the term 'part' has often been used short hand for the term biological parts, the term 'BioPart' was used in this thesis as an aid to help differentiate the DNA encoded parts from the other elements of a biological system which could also be termed 'parts' (such as a chemical, part of a plasmid or the chassis, i.e. host organism, itself). These BioParts encode the simplest biological functions and can be combined to carry out more complex activities in the same way that a radio can be built from simple electronic components.

Developing methods to measure the activity of these basic BioParts in a standard format would provide many benefits to the field of Synthetic Biology, in particularly making it easier to measure and understand BioPart activity and in turn making it easier to choose BioParts to fit a design. Additionally while large repositories of BioParts have slowly been generated the quality of information associated with these BioParts and their activity has not yet reached a suitable standard.

Application of automation technologies would help to push the field forward by allowing the functions of more BioParts to be determined in a standard scenario and better documented in a more useful format.

2.2 Synthetic Biology

Synthetic Biology has become one of the scientific fields showing the most promise to provide new solutions to a huge range of modern problems. Revolutions in the fields of mechanical engineering, chemistry and electronic engineering have triggered large jumps forward for society. Synthetic Biology is in essence the latest step in biological engineering and has the potential to advance society further again with applications in a range of areas. When the National Academy released its Grand Challenges document Biological Engineering has been identified by some authors as one of the technologies that could solve some of these challenges (Yoon and Riley, 2009). Synthetic Biology encompasses a combination of approaches to the design, building and implementation of biological designs, systems and machines. Many of these approaches have taken inspiration and learn from other engineering disciplines to find ways to make it quicker, easier and more reliable to make machines or products from biological designs or systems.

It was a pair of significant breakthroughs in the engineering of biological systems which kick started the field almost 15 years ago. The demonstration of a functioning oscillator (Elowitz and Leibler, 2000) composed of genetic elements working in a biological chassis was an indication to scientists worldwide that something more was possible. The production of a toggle switch (Gardner et al., 2000) shortly after illustrated that engineering of biological systems in this manner was no fluke and was a new technology worth investigating. Oscillators and toggle switches are very simple engineered devices but from these humble beginnings the inspiration has been drawn to develop much more complicated systems, from simple logic gates such as an AND gate (Wang et al., 2011) or NOR gate (Tamsir et al., 2011) to the more complicated edge detector design (Tabor et al., 2009). Each passing year the systems that are built have been becoming more and more complicated with recent systems including a completely re-engineered oscillator with population synchronisation (Danino et al., 2010) and an entire re-design of the nitrogen metabolic pathway of *Klebsiella oxytoca* (Temme et al., 2012b). Some of the early efforts in Synthetic Biology that preceded these were initiated by the undergraduate iGEM competition[1] and the first repository for Synthetic Biology parts – the Parts Registry (n.d). The iGEM competition and Parts registry supplied many of the earliest BioParts flanked by restriction sites in a format which was known as a BioBrick. These BioBricks were given letter and number identifiers which will be used throughout this document (e.g. J23101).

[1] www.igem.org

Underpinning all these advances have been core engineering ideals that make it possible to see beyond the complexity that is endemic to biological systems. It has been noted that key engineering principles can be applied to biological engineering, in particular Abstraction, Modularity and Standardisation (Endy, 2005), though there have been slight differences in the approaches to this abstraction(Andrianantoandro et al., 2006). The engineering design cycle has also played a key role in developments as various groups have worked to accelerate the rate at which designs can proceed through the cycle. It used to be very slow process to first design then assemble (both *in silico* and physically) and finally test a prototype biological system. New tools have sped up the design and modelling steps of the process with programmes such as the Proto Biocompiler (Beal et al., 2011) turning specifications into a design and BioCAD software like Tinkercell (Chandran et al., 2009) which can be used to design and model biological systems. The next stage of the cycle (assembly) has seen particularly strong development with a host of recently developed techniques such as Gibson (Isothermal) assembly (Gibson, 2011), golden gate (Engler et al., 2008) and CPEC (Quan and Tian, 2009) rapidly being taken up and used by the field. Assembly processes have been further improved by the development of software such as j5 which was designed to optimise assembly strategies based on the design, cost and available technologies (Hillson et al., 2012). The final step before a new cycle begins is testing which has often been the slowest step in the process but new technologies such as the latest generation transcription translation technologies from Sun et al. (2013) have been developed to help speed up this part of the process.

While many of the *in silico* tools to speed up the designing and building of biological systems have been rapidly improving and the methods to physically piece together the DNA have advanced at a rapid rate there are still a few areas where improvement can yield significant results. One of the areas with the potential to reap strong rewards for such improvements has been in the characterisation of BioParts to be used for engineering biological systems.

2.3 Rational design of biology

The abstraction hierarchy and individual parts are common to many fields of engineering. The principle is that any system can be broken down into layers of ever decreasing complexity until the level of the simplest functions – parts (or BioParts) – is reached. If these parts are of sufficient quality they can be combined to form modules, which carry out more advanced functions(Endy, 2005). These modules in turn can be combined to form systems which then perform the task that

was required. This approach relies on standardisation and part inter-operability to ensure that the parts to be used are compatible. It has been suggested that the design process that forms part of this process can be further simplified by defining BioParts only in terms of their inputs and outputs (Canton et al., 2008).

The objective of an engineering discipline is to refine this process enough that it should be possible to predict how a design will work from a set of BioParts before they are all assembled and the device tested. This approach has been proven to work for several biological systems, a good example of which were the timers in yeast demonstrated by Ellis et al. (2009). Using the characterisation data for two libraries of BioParts and a model based on those BioParts it was possible to generate timers from a network motif practically identical to the toggle switch of Gardner et al. (2000). Another demonstration of the ability to predict design output was demonstrated by Ceroni et al. (2010) who inserted LacI operator sequences into promoters which were then characterised and the results used to predict how similar devices would function. The ability to build feedback loops from components that have been separately tested has also been shown(Rosenfeld et al., 2007).

For some BioParts it has been shown that their activity is predictable enough that they can be designed to function to a set of specifications *in silico*. One examples of this is the biophysical modelling of the ribosome binding site (RBS) which led to the development of the RBS calculator by Salis (2011). For other BioParts similar models are currently in development such as the model of termination by transcriptional terminators (Chen et al., 2013). These models are now beginning to be combined and may begin to successfully predict the output from a designed module or collection of parts in the near future. The first example of such a combined model that predicts how BioParts will function as part of an expression module has been demonstrated to be both accurate and reliable (Carothers et al., 2011). New BioPart inter-operability and decoupling tools have been developed which will in future make this task even easier. In particular there are now tools to decouple mRNA based BioParts from sequences upstream of their position by a variety of methods (Lou et al., 2012, Qi et al., 2012). While many models have been tested and proven relatively accurate for RNA based BioParts production of models for other BioParts is lagging behind.

Libraries of BioParts have begun to be produced for some of these BioParts which lack detailed models. These include libraries of terminators(Chen et al., 2013, Cambray et al., 2013), orthogonal bacterial repressors(Stanton et al., 2013) and constitutive promoters(Mutalik et al., 2013). A new approach to developing BioParts libraries by insertion of repressor operator sites has been demonstrated to produce libraries of repressible promoters that may allow the prediction of

how the new BioPart will function in advance. This has been shown to be a viable strategy for the development of orthogonally repressible promoters in yeast using TAL effector proteins (Blount et al., 2012) and in *E. coli* using repressors from a range of bacterial strains(Stanton et al., 2013).

In addition to the rational design of these BioParts there have been significant efforts to re-engineer bacterial strains towards minimal strains more useful for a range of applications. Many of these have been based around knocking out chunks of the *E. coli* genome with examples being the removal of many of the sequences responsible for unwanted DNA recombination inside cells (Posfai et al., 2006, Umenhoffer et al., 2010) and another project which generated an *E. coli* strain demonstrating increased product yield of threonine (Mizoguchi et al., 2007). Much of this work has built upon earlier genome minimisation projects carried out on other organisms such as *Bacillus subtilis*(Westers et al., 2003) . Synthetic Biologists have also been working to synthetically build new minimal organisms with the Venter Institute eventually demonstrating the production of a synthetic genome and its successful propagation (Gibson et al., 2010). This synthetic genome was the result of years of work following the determination of the likely minimal genome of a *Mycoplasma* strain (Glass et al., 2006). Some labs have taken these approaches a step further and have begun to try to make protocells which are the absolutely minimal lipids or vesicles and components capable of maintenance and propagation(Sole et al., 2007).

2.4 Applications

It is the applications of Synthetic Biology which have really emphasised the potential of the field. The applications of Synthetic Biology are many and have varied significantly but a few of the more promising areas (in terms of research funding or industrial activity) stand out where the use of biological molecules or organisms has allowed a challenge to be attacked from a completely different angle to most other engineering disciplines.

Biosensors have become a key application area as biological systems continually have to monitor the surrounding environment and tend to be able to detect much lower concentrations of compounds than can easily be detected by other means (generally electrical or chemical). Good examples of where Synthetic Biology has been used to find solutions for problems in sensing is with heavy metal contamination and in particular arsenic where multiple designs for biosensors have been tested that can produce pH (de Mora et al., 2011) or luminescence (Sharma et al., 2013) based outputs and are being looked at for use in the field. Sensors for other molecules have also been

demonstrated including detection of zinc by an RNA aptamer (Rajendran and Ellington, 2008), detection of TNT by a protein receptor(Looger et al., 2003) or by a combination of DNA and antibody(Medintz et al., 2005). Biosensors have even been developed for cancer where RNA logic was used to distinguish tumour types and trigger apoptosis in cells that were identified as belonging to a particular class of cancer(Xie et al., 2011).

Another key application area for Synthetic Biology has been medical technologies and therapeutics. Synthetic Biology has had strong applications in this area as living organisms possess sensors for many bio-molecules which are medically useful, can recognise and manipulate biological motifs and can generate agents which can be aimed at medical targets. Living organisms have developed many ways of exploiting or dominating local populations which make them a good source of compounds and strategies that can be repurposed to fight infections in an age where antibiotic resistance is becoming concerning(Laxminarayan et al., 2013). Good examples of this come from efforts to disrupt bacterial populations in stable biofilms which help to protect them from attack. Two solutions have been developed to attack the biofilms; one that uses an engineered bacterio-phage to both attack potential pathogen cells in a biofilm and disperse the biofilm (Lu and Collins, 2007) and a second that uses the population sensing system of certain bacteria to target them for attack with anti-microbial compounds repurposed from *E. coli* (Gupta et al., 2013). Another good example of Synthetic Biology used for medical purposes was the proof of principle development of *E. coli* cells which have been targeted to invade cancer cells (Anderson et al., 2006). Developments have also been made with regards to implant technology for higher Eukaryotes. Synthetic Biology designed implants have been proven to work in a mammalian chassis (Wieland and Fussenegger, 2012) and shown to be capable of regulating diet in obese mice (Rossger et al., 2013). Perhaps the most well known and prestigious achievement of Synthetic Biology in the medical area has been the development of *E. coli* and yeast strains to allow the large scale production of artemisinic acid which is the precursor to the anti-malarial compound artemisinin (Ro et al., 2006).

A final area where Synthetic Biology has been shown to have significant applications is in the large scale production of biological molecules and compounds. This area of the field has received large amounts of investment as it has the potential to rapidly produce large amounts of currently expensive products and chemicals. Synthetic Biology has already found solutions to the production of many biological molecules such as lycopene (Rodriguez-Villalon et al., 2008) and short chain alkanes (Choi and Lee, 2013) as well as making significant advancements in the production of other biofuels (Fortman et al., 2008). In addition to these advances Synthetic Biology has provided a new toolbox to companies and labs looking to work in this area. This toolbox includes techniques such as

Multiplex Automated Genome Engineering (MAGE) (Wang et al., 2009), reduction of genome size (Trinh et al., 2008) and the development of scaffold structures to increase the efficiency of enzyme pathways (Dueber et al., 2009).

Synthetic Biology has pushed the boundaries of the engineering of biology. Synthetic Biologists have found new ways to solve problems and have begun to find ways to break through the complexity and noise of living systems to allow the design of robust system for a diverse range of applications. There are however still foundational problems to be tackled including the lack of detailed BioPart characterisation data and the natural variation of biological systems (Kwok, 2010). Some of those traditional engineering approaches and concepts will need to be applied in order to solve these issues and achieve the ability to design the living systems that are desired. One area that has particularly deserved attention is the application of automation to Synthetic Biology and in particular for the purpose of allowing the rapid characterisation of more BioParts in a more standardised format.

3 Specific Problem

3.1 Characterisation

Bioparts carry out the simplest biological functions suitable for engineering biological systems. In order to combine BioParts together it is important to understand they function on a quantitative basis. The process to determine these BioParts quantitatively is known as characterisation and is of vital importance to the bottom up engineering of biology.

It has been found that characterising BioParts is frequently a difficult and slow process. Unlike most other engineering disciplines it has been difficult to observe the output or activity of a BioPart directly. Many BioParts are active as part of a protein or RNA expression system in a living system and so required the use of some kind of biological reporter system to observe their behaviour. Reporters such as enzymes responsible for a colour change, for example LacZ (Zhang and Bremer, 1995), production of light (Brasier and Ron, 1992) or the changing of light from one wavelength to another by a fluorescent protein such as modified GFPs (Andersen et al., 1998) have all been demonstrated for identifying biological expression behaviours. However for these outputs to be properly utilised care has to be taken to ensure that output accurately reflects or can be used to calculate the desired process, for example by monitoring GFP fluorescence levels over time to convert them into parameters suitable for the modelling of test systems (Ronen et al., 2002).

Different BioParts often require different approaches in order to determine their characteristics that would allow them to be integrated into a design. For a few of the BioParts such as promoters, ribosome binding sites (RBSs) and other expression related BioParts it has been possible to characterise them all by the use of GFP. A type of BioPart known as transcriptional terminators has also been characterised with GFP by the clever use of an additional fluorescent reporter (Chen et al., 2013) and RNAse sites(Cambray et al., 2013). Other reporters which have been demonstrated for use as characterisation tools include a fusion of GFP and the LacZ alpha fragment named Gemini (Martin et al., 2009) and RNA aptamers that bind chemicals to make fluorophores such as the Spinach aptamer (Pothoulakis et al., 2013).

Characterisation *in vivo* has tended to be a slow process requiring large amounts of preparation and assay time. An attempt to reduce the amount of time characterisation requires has been the development of many new *in vitro* characterisation systems. How much these *in vitro* systems predict how BioParts will behave *in vivo* is debatable however it has been shown that at

least for simple systems they can show comparable results (Chappell et al., 2013). The Tx-Tl reaction mix system was developed with the explicit purpose of allowing a large number of BioParts to be tested in a first pass, concentrating resources for *in vivo* characterisation onto a smaller number of designs more likely to succeed (Sun et al., 2013). Transcription translation systems have also been demonstrated that characterise promoters from a mixed population based on transcript abundance (Patwardhan et al., 2009, Kosuri et al., 2013). In the longer term it has been suggested that microfluidics will be used to further enhance the rate at which new BioParts and designs can be tested (Gulati et al., 2009). There are limitations to *in vitro* characterisations however as the systems tend to have short life-spans and protein supplementation may be required. As a result it is difficult to get accurate expression of regulatory and other proteins when multiple component systems are tested, for example a feed forward loop or inverter.

Many of the most prominent examples of characterisation have been carried out on promoters *in vivo* using methods that vary primarily on the type of data they produce. Promoters exhibit both simple and complicated behaviours and they are one of the key control elements required for control of biological systems. A large portion of the field of Synthetic Biology has been focused on the engineering, design and characterisation of systems in *E. coli* and on these characterisations.

3.1.1 Data and Metrology

The methodology to characterise BioParts *in vivo* has been slowly evolving over time. A large number of these methods were based upon monitoring production of fluorescence for a test construct and optical density over time. The results of these characterisations were the normalised fluorescence level or the change in fluorescence over time for a set of BioParts. This output data format was simple and could be produced by most labs with access to a microplate plate reader however data comparison was difficult due to the arbitrary scales. This methodology was common to many characterisations and was sufficient to prove conclusions in the past for a variety of engineered systems for example Lee et al. (2011). For the long term future of an engineering discipline such arbitrary units were likely to prove to be not up to requirements.

A couple of methods to improve the metrology in Synthetic of Biology have since been developed to tackle this problem and generate a data format more suitable for comparing data between labs. Chronologically first was the characterisation of the F2620 Biobrick by Canton et al.

(2008). The design behind the Canton characterisation was to abstract the input and output of an inducible protein system to produce a black boxed module that took a signal of acyl homoserine lactone (AHL) and converted it to transcriptional activity from a promoter which could be coupled to other modules. Canton took the F2620 Biobrick (which contains the pLux promoter and also encodes the *LuxR* gene under the control of the constitutive in *E. coli* pTet promoter) and added a GFP expression cassette to allow the output of the promoter to be tested. The promoter was then characterised using much the same method as used previously with fluorescence and optical density (OD) monitored to allow calculation of results. Canton took this method further however and attempted to convert the results of this characterisation into data with an absolute set of units. The demonstrated results were initially the direct output in units of molecules of GFP per cell and these were further converted into specific BioPart characterisation results in units of Polymerases per Second or PoPS. The first of these conversions was achieved by doing calibration experiments with purified GFP to convert the arbitrary fluorescence scale into molecules of GFP and carrying out plating experiments with solutions of cells at various optical densities to determine the colony forming units (an approximation of the population level) per OD unit. This first conversion was highly appropriate as converted the signals observed into units of the source of the signals i.e. representing the results as they were observed. The second conversion to PoPS required the application of several levels of parameters based on other experiments. This measure of the promoter was ideal from an engineering perspective but from a biological perspective may be particularly prone to error (as it requires calculation based on mRNA levels which may not be consistent across time) and may be difficult to reproduce. The methods and results achieved by Canton are of a very high standard but have been observed to take significant time and effort for others to implement.

A contrasting approach was that taken by a group of Synthetic Biologists in the USA and Europe lead by Jason Kelly which resulted in the development of Relative Promoter Units (Kelly et al., 2009). Relative Promoter Units were the output of a promoter relative to the output of promoter J23101 (a promoter BioBrick from the Parts Registry) with a similar 5' UTR on the associated transcript. It has been suggested that calculating results in RPU may reduce the amount of error observed as the reference and tested promoter are held in a similar genetic and chassis context thus removing some of their impact from their results (Kelly et al., 2009). The RPU results are simple for any lab to calculate as the only requirement is an extra control in the experiment. However the results are again relative and not in absolute units so they may be difficult to utilise in design programmes. Unlike GFP per cell per second, RPU has begun to achieve a decent level of uptake by the field however with many recent characterisations using RPU or a derivative as the output units (Davis et al., 2011, Zucca et al., 2012).

Additionally it has been shown that results can be expressed in RPU for data produced on both a plate reader and a flow cytometer and increasingly flow cytometry is being utilised in characterisations. Unlike a plate reader where samples are read as a population flow cytometry scans individual cells for their size and fluorescent properties. This has allowed distributions inside the population to be identified, for example *E. coli* cells carrying slightly different copy numbers of vector (Beal, 2012). As a result flow cytometry has been increasingly used for characterisation. Recently it has been used in a large study by to characterise a series of promoter libraries and the output of numerous promoter – ribosome binding site combinations (Mutalik et al., 2013), as well as numerous other projects including AND gate characterisation(Anderson et al., 2007), and promoter library characterisation in yeast (Ellis et al., 2009) among others (Moon et al., 2011, Temme et al., 2012b).

Metrology has also an issue for flow cytometry with many of these flow cytometry characterisations reported using arbitrary fluorescence or normalised fluorescence values. Some labs have begun to define flow cytometry output with an attempted to adopt an approach similar to RPU - Relative Expression Units (Temme et al., 2012b) and most notably the use of standard calibration beads as part of the TASBE tool chain(Beal, 2012).

3.1.2 Chassis and Characterisation conditions

Biological systems are well known for being inherently noisy (Pilpel, 2011). Characterisation carried *out in vivo* uses a living chassis that is inherently influenced by its environment. There has become an increasing awareness that large numbers of biological processes inside the cells are highly dependent upon growth rate which is itself influenced by the conditions of the characterisation. It has been suggested that changes to growth rate will influence the amount of proteins and macromolecules(Klumpp et al., 2009), plasmids copy numbers(Klumpp, 2011) and even the activities of RNA polymerases inside the cell(Klumpp and Hwa, 2008). Work has begun to find ways to predict these factors (Scott and Hwa, 2011) but they should be considered when designing experiments.

A standard characterisation strain has not been selected for use across all of Synthetic Biology and labs have characterised with a selection of strains with often no real indication of why those strains where chosen for a particular project. Some of the *E. coli* strains used more commonly for characterisation have included the almost wild-type strain MG1655 (Canton et al., 2008, Moser et

al., 2013) and the parent strain of the Keio collection BW25113 (Mutalik et al., 2013, Cambray et al., 2013). However many characterisations have been carried out using cloning strains such as DH5α (Kelly et al., 2009, Ceroni et al., 2010), BL21 (Chappell et al., 2013), Top10 (Mutalik et al., 2012, Qi et al., 2012) and DH10B (Temme et al., 2012a, Salis et al., 2009, Moser et al., 2012). The reasons for strain selection have rarely been reported though may be for protein expression related reasons. The effect of using a particular chassis on the function of a circuit has been studied by Cardinale et al. (2013) with many of the chassis previously mentioned. Many of the strains tested by Cardinale et al. (2013) performed slightly differently when carrying the same devices. The variations owing to the strain have the potential to be a barrier to the sharing of characterisations across the field and with the development of increasingly minimal *E. coli* strains such as MDS42 (Posfai et al., 2006) and MGF-01 (Mizoguchi et al., 2007) there may soon be cause for establishing a standard set of strains which can be shared.

Fortunately media selection is an area where characterisation efforts have tended to stick to similar choices and increasing standardisation may be occurring. Many projects have been carried in traditional *E. coli* cloning, growth or protein production media such as LB (Lee et al., 2011) but there has been uptake of more defined media. The switch to defined media is a particularly welcome development as all the components are at set concentrations making reproduction easy. For the last 5 years BioPart characterisations have been increasingly carried out in some form of M9 frequently supplemented with casamino acids and thiamine. M9 media supplemented this way has probably been used because of its use in standard characterisations (Canton et al., 2008, Kelly et al., 2009).Recently one of the most defined media formulations described, EZ rich defined MOPS media based on the original Neidhardt recipe(Neidhardt et al., 1974) , has begun to be used by labs for characterisation(Cambray et al., 2013, Mutalik et al., 2013). Switching to this media has been advantageous to the field as it is a completely defined media and so batches should be practically identical which will hopefully decrease the variability in experiments between labs.

Characterisation has frequently been carried out in microplates and this brings a couple of challenges for the characterisation conditions. Microplates have been considered to have poor aeration kinetics and therefore it is possible that cells grown in them may become limited by availability of oxygen (Doig et al., 2005). If oxygen is no longer available to cells it has been shown that they will begin to utilise anaerobic metabolic pathways (Choe and Reznikoff, 1991) which would radically alter the internal metabolism of the cell (Salmon et al., 2003) and ultimately the interactions with any genetic circuit inside. As many characterisation have been carried out using GFP this could be particularly concerning as GFP requires oxygen for formation (Heim et al., 1994).

At present only a single *in vivo* sensor for oxygen has been demonstrated to function in *E. coli* and this was based on FRET between an oxygen sensitive flurophore and an oxygen insensitive flurophore (Potzkei et al., 2012). Previous experiments monitoring gene expression suggest proteins which may make viable biosensors for oxygen or anaerobic metabolism, for example the fumarate and nitrate reductase (FNR) which is one of the key transcription factors responsible for the transition to anaerobic metabolism and is itself directly repressed by oxygen (Unden et al., 2002).

It is also well acknowledged that microplates well close to the edge of microplates often exhibit different results from those in the centre and this has been termed the 'edge effects'. This effect has been shown to be at least partly caused by the outer wells of a microplate heating more rapidly when the plate is moved between different temperature environments (Oliver et al., 1981, Burt et al., 1979). While this was originally observed during ELISA immunoassays (Wreghitt et al., 1982) it has also been seen to affect the growth of cells (Fukazawa et al., 1995, Oliver et al., 1989). In some experiments it has been possible to overcome these edge effects by pre-incubating at room temperature (Lundholt et al., 2003), filling plate edge wells with media or water as a buffer (Oliver et al., 1989) or possibly by simply covering the plate (Burt et al., 1979). It has also been demonstrated that for ELISA at least there may be more statistically significant variation between wells randomly than as a result of the edge effect (Shekarchi et al., 1984).

The field of Synthetic Biology has been attempting to head in the direction of industrialisation and the first experiments to test Synthetic Biology designs (specifically logic gates) in conditions similar to industrial scale-up conditions was recently carried out and demonstrated that for some designs scale up was possible (Moser et al., 2012).

3.1.3 Genetic and host context

Beyond the environmental context (media and conditions) encountered by the chassis it has also been observed that the genetic context of a BioPart can have a large effect on the results of a characterisation. Most BioParts have been characterised while hosted on a plasmid vector and the components of this physical DNA may have had influence the results obtained for the tested BioParts. As yet there are no clear standardised vectors specifically for characterisation with a range of different expression or assembly vectors being utilised instead. These include vectors designed for expression and/or assembly via the use of either the BglBrick standard (Anderson et al., 2010, Lee et al., 2011) or BioBrick standard (Canton et al., 2008, Kelly et al., 2009). Non standard plasmids that

have been used for characterisation tend to contain an important BioPart (Qi et al., 2012), have been regularly used for cloning in the lab or have been designed for the characterisation (Mutalik et al., 2013). Recently a standardised vector format (pSEVA) was been developed (Silva-Rocha et al., 2013) and switching to use this plasmid family may be beneficial in the future. Using a common plasmid would keep sequences and local context the same, reducing variation in characterisation experiments.

For many purposes however it has been necessary to build new plasmids for characterisation in order to remove unknown interactions. Many of the parts of a plasmid have been shown to alter the results obtained from a BioPart characterisation and one of the more clearly understood plasmid parts in terms of interaction with characterisation is the origin of replication. While these origins are considered to be well understood, plasmid copy number has been shown to vary with growth rate (Klumpp, 2011) and at least one origin has been shown to be affected by read through from strong promoter systems (Stueber and Bujard, 1982). In addition to this it has been observed that a BioPart can burden the system in which it is held and this can potentially reduce the chassis growth rate (Canton et al., 2008) and alter the characterisation (Dekel and Alon, 2005). Cells are highly balanced living machines and it has been seen before that systems within these organisms can easily become overloaded (Cookson et al., 2011). This has been handled previously by using a lower copy number replication origin such as p15a, SC101 or ColE1. Controlling copy number itself has been shown to be feasible using a combination of certain origins and a dial plasmid strain(Kittleson et al., 2011) but for most characterisations the low copy origins have be used.

As the function of many BioParts is tied to their physical shape and structure (often dictated by sequence) the presence of nearby sequences can alter their properties. This is particularly true of promoters and RNA BioParts. Promoters have been shown to rely upon sigma factor and regulator protein binding which can both be influenced by nearby sequences such as poly-A tracts upstream of the sigma binding sites(Ellinger et al., 1994) and initiation blocking sequences which have been found immediately after the transcription start site of *E. coli* sigma 70 promoters (Chan and Gross, 2001). It has been shown that some BioParts (particularly promoters) can be protected from these sequences by addition of extra flanking DNA to form a physical barrier (Davis et al., 2011).

RNA BioParts have been shown to be even more dependent upon context as the transcript they are part of has a global characteristic - its stability. Addition of RNA hairpins have been shown to increase the output from expression constructs(Carrier et al., 1998), as has the introduction of transcriptional terminators to a sequence(Aiba et al., 1991). This is most likely due to their interaction with the RNAse genes for *E. coli*, slowing their easy cleavage and degradation (Grunberg-

Manago, 1999). The ribosome binding site also tends to be close to the start of the transcript and it has been shown that they can be heavily influenced by surrounding structures (Kudla et al., 2009, Salis et al., 2009). It has been suggested that this can cause issues when they are used with certain promoters which contain nucleotides after the transcription start site which become incorporated into the mRNA sequence (Kelly et al., 2009). This may be particularly problematic as DNA binding sites for some proteins have been shown to be more effective with palindromic sequences (Park et al., 1993) and may therefore be more likely to form structures in the RNA which may prevent ribosome access to the RBS. Fortunately new context decoupling devices have been designed that will allow the targeted cleavage of the RNA to remove unwanted pieces of sequence and structure (Qi et al., 2012, Lou et al., 2012).

3.2 Standards and documentation

Accurate documentation is vital for an engineering discipline, particularly given the variation in procedures carried out across the field. As has been documented above characterisation assays for the similar BioParts in different labs are often carried out with different conditions, different reporters and in a different chassis. Any one of those changes could have had a significant impact on characterisation results and unless all the experimental variables were recorded it would be impossible to account for their affects on the results. At present large amounts of BioPart characterisations have been published (Mutalik et al., 2013, Cambray et al., 2013), requiring a decent reporting level for the experimental steps carried out. While most data is still reported this way, attempts to define ways of reporting and exchanging data have also been worked on.

To better document Biopart characterisations attention first turned to how this problem had been tackled in other engineering disciplines, a decision that led to production of the first datasheet for a Synthetic Biology device(Canton et al., 2008). The F2620 datasheet was been a seminal moment in the field, setting down a gold standard both in terms of the methods used for characterisation (see 3.1.1) and for the information that was useful to know about a characterisation when choosing a BioPart. The Canton datasheet documented many of the characterisation conditions under which the data was obtained, including the cell strain, a brief description of the media used and some details regarding the plasmid the BioPart was carried on. This information covered a large number of the variation areas previously discussed and would have been a major help to anyone attempting to debug the promoter activity if an unexpected behaviour was observed

in the lab. While there was an indication that the BioPart was qualitatively tested in a separate set of conditions and behaved the same there was unfortunately no indication of which properties were tested and to what level of detail. The datasheet itself was backed up by electronic files on the parts registry (Parts Registry, n.d) which gave access to the full sequence information and raw data files.

An alternative datasheet was produced to document the characterisation of a series of BglBrick vectors (Lee et al., 2011). These datasheets focussed far more on the experimental results side of the datasheet, possibly because they were designed to describe the properties of the entire plasmid or because they were aimed at a different target audience. In those datasheets a more limited set of information about the characterisation was included with only the plasmid map and components indicated on the sheet (and in the plasmids name) along with the media used for a particular experiment. As a datasheet for a BioPart characterisation it was lacking in useful information but to help set up fermentations to produce proteins (their likely aim) it would be a very useful datasheet.

While no other datasheet have been produced so far there have been interesting suggestions for what should be included in a datasheet with specific attention placed on what kind of BioPart was being characterised as for certain BioParts different protocols would be required (Arkin, 2008). Some of the additional features identified by Arkin (2008) were also observed by other groups within the field of Synthetic Biology to be potentially useful information leading steadily to the development of various tools. This began with the Parts Knowledge Base (Galdzicki et al., 2011) and later resulted in a request to the Synthetic Biology community to release full plasmids sequences for all constructs used in testing (Peccoud et al., 2011). This simple request was important for the field as at the time published results often did not include full sequences which may have been vital for obtaining the correct results.

To tackle many of these problems a large collaborative effort was undertaken to develop a standard way of annotating DNA sequences to ensure that they could be transferred lab to lab without loss or change of information. The ultimate objective was to standardise a common information core that could be used by the next generation of tools that would be developed by the field. This effort finally resulted in the completion of the Synthetic Biology Open Language (SBOL) (Quinn et al., 2013). SBOL has rapidly been taken up as the language of choice for DNA annotation with many tools designed to work with it from day 1 including TinkerCell (Chandran et al., 2009), j5 (Hillson et al., 2012) and iBioSim (Myers et al., 2009).

3.3 Promoter BioParts and Promoter Engineering

Promoters have been one of the highest priority BioParts to characterise. They are one of the key BioParts in most control systems as they are capable of responding to a signal, usually in concert with a protein regulator. The simplest promoters however are constitutive, i.e., permanently turned on in the presence of their controlling sigma factor. While the initiation of transcription from a promoter is a very complicated process it is based on DNA sequence motifs which have been shown to offer significant potential for engineering. See Hook-Barnard and Hinton (2007) for a good review of how promoters work.

Many libraries of promoters have been created and these have generally been libraries of constitutive promoters. One of the sets of promoters most commonly used by Synthetic Biologists (particularly students in the iGEM competition) is the Anderson promoter collection. This set of promoters is particularly used as one of the Anderson collection promoters was established as a reference promoter as part of the RPU characterisation standard (Kelly et al., 2009). Additionally the promoters in this library have frequently been used to test the accuracy of characterisations. Other libraries of *E. coli* sigma 70 constitutive promoters have been generated by Alper et al. (2005) and Mutalik et al. (2013) which possess a wide range of activities.

Regulated promoters have proved to be much more complex and interesting to work with. Many of these promoters have been taken from nature such as the *lac* and *araBAD* promoters which have been used in various forms for decades. A few *E. coli* native promoter and regulator systems such as *rhaB* promoter – RhaS regulator system have already been developed into expression vectors (Wegerer et al., 2008) and other previously discovered systems may make ideal candidates for characterisation – for example the *xylF* promoter – XylR regulator system (Song and Park, 1997). Some of these *E. coli* native systems have shown interesting interactions – for example rhamnose induction by RhaR and RhaS regulator proteins (Egan and Schleif, 1993) - which could lead to unintended but potentially useful behaviours. New approaches to developing regulated promoters have been developed in the form of promoter engineering.

It has been shown that it is possible to define modules in the DNA sequence of a promoter and exchange these modules to engineer new promoters. This method was employed to develop at least one of the BioFAB constitutive libraries(Mutalik et al., 2013) but has also been used to engineer multiple input promoters useful for logic functions that respond to an activator and a repressor (Cox et al., 2007). An alternative approach to develop repressible promoters by the insertion of operator

sequences into promoter libraries has also been demonstrated recently with the insertion of TetR and TAL operators into a constitutive yeast promoter (Blount et al., 2012). The large volumes of sequencing data for various prokaryotic organisms have enabled this approach to generate new orthogonally repressed promoters. Orthogonally repressed promoters for *E. coli* have been generated using either repressor genes (Stanton et al., 2013) or anti sigma factors (Rhodius et al., 2013) from other organisms.

3.4 Automation in Biology

The use of automation has lead to significant improvements in throughput of repetitive processes and its application to Synthetic Biology would be beneficial. While liquid handling automation has been identified as potentially useful technology for Synthetic Biology (Kahl and Endy, 2013) the application of automation to most laboratory processes is relatively sparse and has often been focused on software to make genetic designs or assembly strategies more efficient. For this purpose that software for optimising BioBrick assembly strategies (Densmore et al., 2010) and the j5 software was developed (Hillson et al., 2012). Following this it has been demonstrated that it was possible to apply automation to the physical assembly of DNA (Leguia et al., 2011). Within the community there has been a push to standardise robot running programmes and this has lead to the development of a new robotic platform programming language known as Par-Par (Linshiz et al., 2013).

So far only a few examples of the application of automation have been published. An assay for β-galactosidase that is appropriate for use in prokaryotes, yeast and even mammals has been demonstrated on an automated platform (Vidal-Aroca et al., 2006) but the most relevant use of automation to date has come from the Alon lab (Kaplan et al., 2008). The Alon lab has used an automated platform to carry out simple GFP expression experiments using assays similar to some of those used previously for characterisation. This is particularly noteworthy because this method was originally demonstrated by Kaplan et al. (2008) in a 96 well plate format and proved successful enough that a second generation of these experiments were carried out by Zaslaver et al. (2009) suggesting that the development of protocols to automate characterisation may be a productive endeavour.

3.5 The approach at CSynBI: SynBIS

The EPSRC and BBSRC funded Centre for Synthetic Biology and Innovation at Imperial College has been developing technologies to aid the Synthetic Biology community. One ongoing project has been the development of the Synthetic Biology Information System (SynBIS) (Kitney and Freemont, 2012) and related tools and techniques. One of the key components of SynBIS is a database designed to hold and document BioPart characterisation experiments. The aim of the database was to develop a large repository of BioPart characterisations using known, validated methods which yield high quality data. Key to this was the standardisation of the key parts of these procedures; the characterisation, the data import and analysis and the data storage and display.

3.6 Project Aims

1) To test the current methodologies and standards to determine useful techniques which can be used to generate and share BioPart characterisation data.

2) Develop an automated methodology for characterising BioParts in a high-throughput manner without sacrificing data quality or relevance.

3) Characterise an initial batch of BioParts to be used in the development of SynBIS.

4 General Methods

4.1 Bacterial strains and media

E. coli strain Top10 (see table 4.1 for strain genotypes and source) was used for all cloning to reduce the risk of mutation and recombination. For characterisation *E. coli* Top10, DH10B, MG1655 and MDS42 were used initially to identify suitable strains. All standardised characterisation (automated protocol v1 and beyond) was carried out using MG1655 from the Busby lab[2] as this was chosen to be a common strain for the ST-Flow working group.

Strain	Genotype	Source
Top10	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697*gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen
DH10B	F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (araleu)7697 galU galK λ- rpsL nupG /pMON14272 / pMON7124	Invitrogen
MG1655 (ATCC)	F- lambda- ilvG- rfb-50 rph-1	ATCC® 700926™
MDS42	MG1655 multiple-deletion strain(Posfai et al., 2006)	Scarab Genomics
MG1655 (Busby)	F- lambda- ilvG- rfb-50 rph-1	Busby lab[2]

Table 4.1 Strains of *E. coli* used in this work.

Antibiotics were used at the following concentrations: ampicillin (100µg/ml), kanamycin (50µg/ml), chloramphenicol (35µg/ml) and streptomycin (50µg/ml) (All Sigma Aldrich). All overnight cultures were grown in Luria broth. Early characterisation assays were carried out in M9 supplemented media with the following formulation: M9 salts (Formedium), Casamino acids (0.2% W/V) (Becton Dickinson), 1mM thiamine hydrochloride (Sigma), 2mM magnesium sulphate, 0.1mM calcium chloride, 0.4% (W/V) glycerol or glucose and antibiotic. Supplemented M9 was chosen as it had been used in many previous characterisations (Kelly et al., 2009, Canton et al., 2008).

Later characterisation assays were carried out in MOPS EZ Rich Medium (Teknova) with 0.4% or 0.5% (W/V) carbon source (usually glucose) and antibiotic. EZ rich MOPS media was chosen for characterisation to improve standardisation as this was the only available fully defined media available at the time.

[2] http://www.birmingham.ac.uk/staff/profiles/biosciences/busby-steve.aspx

4.2 Cloning and DNA manipulation

4.2.1 Plasmids and DNA preparations

Plasmids minipreps were carried out according to manufacturer's protocol (Qiagen or EZNA/Omega biotech). PCR purifications, gel purifications and midi-preps were carried out according to the manufacturer's protocol (Qiagen). Most of the plasmids used were taken from the Parts Registry (n.d) although a few were provided by other members of the CSynBI and these are listed in table 4.2.

Transformation was carried out with chemically competent cells prepared using the CCMB method developed by Hanahan et al. (1991). Transformation was carried out using 50µl of cells and 5µl of ligation/CPEC mix or 1µl of miniprep DNA. Cell aliquots were allowed to thaw on ice before 50µl volumes were transferred to sterile 1.5ml eppendorf tubes where they were kept for 10 minutes prior to addition of DNA. Following addition cells were left for 30 minutes before heat shock at 42°C for 45 seconds and then returned to ice for 2 minutes. 200µl of pre-warmed LB was then added and cells rested for an hour before 100µl or all of the cell suspension was plated.

Plasmid/DNA	Contents	Source
pSB3K3	KanR modules, p15a replication origins	Parts registry
pCT	RFP expression module	Tom Ellis
Top10 Genome	FNR expression cassette	Lysed from cells (Invitrogen)
pAC1-ndh	ndh - GFP expression casette	Ben Stell/Kirsten Jensen
pAC1-narK	narK - GFP expression cassette	Ben Stell/Kirsten Jensen
pAC1-dmsA	dmsA - GFP expression casette	Ben Stell/Kirsten Jensen
I13504	B0034 based GFP expression module	Parts registry
I13507	mRFP gene	Parts registry
K316004	mRFP expression module pieces	Parts registry
MG1655 Genome	XylR gene, RhaS gene	Lysed from cells (ATCC)
I741018	XylF promoter	Parts registry

Table 4.2 Plasmids which were the source of DNA used in Cloning

4.2.2 General Cloning and PCR

All restriction enzymes, T4 ligase and T4 PNK were purchased from New England Biolabs (NEB). Restriction enzymes were used according to the manufacturer's protocols. For restriction cloning ligations were carried out using equimolar amounts of DNA in 1X ligation buffer. If a single restriction enzyme was used or enzymes which left complimentary overhangs plasmid dephosphorylation was carried out using RAPID alkaline phosphatise (Roche) according to the manufacturer's protocols.

For round the plasmid PCR and re-ligation cloning T4 PNK was first used to phosphorylate DpnI digested PCR fragments using a mix of 7µl DNA, 1µl T4 ligase buffer and 1µl T4 PNK. The mix incubated at 37°C for 30 minutes before denaturation at 65°C for 20 minutes. Following this the mix was split into two tubes (4.5µl in each) and T4 ligase added to one tube and water other (to act as a background test).

For oligo based cloning oligos were originally diluted to 10µM (each oligo) in a 20ul reaction with water 1X T4 ligase buffer. These reactions were heated to 95°C in a water bath before the bath was switched off and allowed to cool overnight. The following day oligos were phosphorylated using T4 PNK which was added into the annealed oligo mix as this already contained the 1X T4 ligase buffer and the reaction carried out with the same conditions as for PCR product phosphorylation. Following this phosphorylated annealed oligos were ligated into a cut plasmid by T4 ligase.

All PCR reactions (excluding colony PCR) were carried out using Phusion polymerase (NEB, Thermo or Finizymes) using buffer HF, 2% DMSO, 0.2mM each dNTP and 0.25µM each primer. The annealing temperatures used would generally be the lower Tm of the annealing region of the two primers and never above 68°C. All PCR protocols contain an initial 30 second 98°C denaturation, a final extension of 72°C for 2-3 minutes and cycles of 98°C denaturation for 10 seconds, annealing for 30 seconds and extension at 72°C for 15 seconds per kb DNA plus 15 seconds. All PCR reactions were carried out using G-Storm thermocylcers. If PCR fragments were to be used for cloning they were immediately digested with DpnI in the PCR reaction mix prior to purification.

Two main types of cloning PCR were carried out depending on what DNA manipulation was being carried out; generation of new plasmid modules (using CPEC or In-Fusion) or transfer of existing modules (using restriction enzyme cloning). Module generation PCR was carried out using a 30 cycle PCR programme - 5 cycles at the annealing region Tm followed by 25 cycles with annealing at 68°C - with plasmid DNA (5ng) or genomic DNA (2µl of a bacterial colony boiled in 100µl of ddH_2O)

as the template. The new module would be flanked by homology sequences to allow CPEC or In-Fusion cloning into a plasmid. Module transfer was carried out with a 10 cycles PCR programme - 5 cycles at the annealing region Tm followed by 5 cycles with annealing at 68°C - on plasmid DNA (~40ng). The module was then cut with the relevant restriction enzyme and transferred into the new plasmid.

Colony PCR was carried out using OneTaq DNA polymerase (NEB) using the manufacturer's buffers and same PCR conditions and protocol as for Phusion PCRs (with extension time increased to 60 seconds per kb plus 60 seconds). OneTaq was chosen because it uses the same buffer as Phusion (and so should work in the same conditions). Sequencing was carried out by Eurofins MWG or Source Bioscience. Prior to sequencing DNA was amplified by a templiphi kit (GE Healthcare) to improve sequence read quality and length.

CPEC reactions were carried out with an initial denaturation (98°C for 30 seconds) followed by 15 cycles of denaturation at 98°C (10 seconds), annealing at 62°C (30 seconds) and extension at 72°C for 15 seconds per kb plus 15 seconds and a final extension of 2 minutes. CPEC reactions were carried out using the Phusion PCR mix minus primers and an equimolar amount of each plasmid piece.

All the primers used are detailed in full in table 4.3 (Minimal plasmids primers), 4.4 (SV plasmids primers), 4.5 (Anderson Oligos) or 4.6 (Synthetic promoter primers).

4.2.2.1 Cloning of Minimal Vectors

The 3 module minimal vector was first designed by Eichenberger et al. (2009). The 3 module plasmid backbone was assembled by In-Fusion (Clontech) according to the manufacturer's protocol following PCR amplification of the 3 modules. The modules were generated by PCR amplification using primers p15a F 100bp and p15a R 100bp with plasmid pSB3K3 (module 1 – p15a), primers Kan F 100bp and Kan R 100bp with plasmid pSB3K3 (module 2 - KanR) and primers MCS F 100bp and MCS R 100bp with plasmid I13504 (module 3 – MCS/GFP output).

Promoters (J23101, P_{tac} and P_{lacUV5}) were introduced into the 3 module backbone plasmid by oligos generating plasmids pAC4, pAC5 and pAC7 respectively. The promoters were ordered as unphosporylated oligos (see table 4.2) and annealed prior to ligation into the dephosphorylated plasmid using T4 ligase.

Primer	Sequence	Module and (plasmid)
Kan F 100bp	catttaaatcggtttaaacggtaatgctctgccagtgttac	KanR-minimal (Minimal 3M)
Kan R 100bp	tagctagccggtttaaacgcttgccacggaacggtctg	KanR-minimal (Minimal 3M)
p15a F 100bp	cgacgagacgatttaaatggggtatgagtcagcaacaccttc	p15a-minimal (Minimal 3M)
p15a R 100bp	cgtttaaaccgatttaaatgcccttaacgtgagttttcgttc	p15a-minimal (Minimal 3M)
MCS F 100bp	gcgtttaaaccggctagctacattaacctataaaaataggcgtatcacg	GFP expression (Minimal 3M)
MCS R 100bp	ccatttaaatcgtctcgtcgcgaccgagcgcagcg	GFP expression (Minimal 3M)
J23101 oligo F	AATTCGCGGCCGCTTCTAGAGtttacagctagctcagtcctaggtattatgctagcTA	J23101 promoter
J23101 oligo R	CTAGTAGCTAgctagcataatacctaggactgagctagctgtaaaCTCTAGAAGCGGCCGCG	J23101 promoter
Ptac F	AATTCGCGGCCGCTTCTAGAGttgacaattaatcatccggctcgtataatgtgtggaattgtgagTA	pTac promoter
Ptac R	GCGCCGGCGAAGATCTCaactgttaattagtaggccgagcatattacacaccttaacactcATGATC	pTac promoter
PlacUV5 F	AATTCGCGGCCGCTTCTAGAGccccaggctttacactttatgcttccggctcgtataatgtgtggaattgtgagTA	pLacUV5 promoter
PlacUV5 R	GCGCCGGCGAAGATCTCggggtccgaaatgtgaaatacgaaggccgagcatattacacaccttaacactcATGATC	pLacUV5 promoter
FNR 4 F	GATTCATAAGATGGAGCGCGacattaaacaatttgtgccagcttgttc	FNR expression (FNR plasmid)
FNR 1r R	CCATTTAAATCGTCTCGTCGtcaggcaacgttacgcgtatg	FNR expression (FNR plasmid)
MCS 4r R	cgcgctccatcttatgaatccgaccgagcgcagcg	GFP expression (FNR plasmid)
ChaMCS Seq R	Gtagtgacaagtgttggc	Promoter sequencing
KanF Seq	Tgtaagcagacagttttattg	FNR Casette Sequencing

Table 4.3 Primers used to generate sequences or assemble plasmid pieces into the minimal or FNR plasmids.

Promoter sequences in oligos are the lower case letters. All primers except FNR 4 F, FNR 1r R, MCS 4r R and ChaMCS seq R were designed by either Eichenberger et al (unpublished) or T. Ellis.

4.2.2.2 Cloning of FNR plasmids

A new minimal plasmid was generated by addition of a fourth module containing an *fnr* expression cassette following the *GFP* gene. The *fnr* cassette was made by amplifying the section of the *E. coli* MG1655 genome containing the *fnr* gene, its promoter and the sequence downstream likely to contain the terminator with primers FNR 4 F and FNR 1r R and In-Fusion cloning it into the pAC1 plasmid amplified with primers p15a F 100bp and MCS 4r R. The ndh, nark and dmsA promoters had previously been cloned into 3 module plasmids by Ben Stell and Kirsten Jensen as part of an earlier project. To transfer the promoters into the 4 module plasmid, restriction cloning using the EcoRI and BsaI restriction enzymes was used.

4.2.2.3 Cloning of SVa plasmids

A more appropriate standardised vector backbone for characterisation was designed and the initial version named SV for Standard Vector. The 'a' suffix was added later when modifications to the plasmid for better operation were required (the overall topology remaining approximately the same each time). For the characterisation of inducible promoters 4 modules were required; origin, resistance, output and regulatory modules. A low copy origin was needed and p15a was already well established for characterisation. Similarly Kanamycin resistance was chosen for both its use in characterisation before but also its bactericidal nature. The GFP output module was based on the module from the minimal vector and retained the BioBrick prefix for promoter cloning. The regulator module was designed to allow production of regulatory proteins (see 6.1.2 for details) but contains RFP for constitutive experiments (to help cloning and to potentially act as a second reporter). The topology was chosen to reduce the likelihood of interaction between the modules.

All SVa PCRs were module generation (30 cycle PCRs). The initial SVa backbone plasmid was generated by a CPEC reaction containing all 4 modules, further SVa plasmids carrying regulator genes were generated by using CPEC to swap the RFP gene for the appropriate regulator. The primers used in the various steps can be found in table 4.4.

Plasmid SVa was generated by CPEC from 4 separate modules. The GFP module was generated from the I13504 BioBrick by PCR with primers Exp Cas Fwd and Exp Cas Rev. The KanR and p15a resistance modules were generated from plasmid pSB3K3 by PCR using primers SVKan Fwd and SVKan Rev and SV15a Fwd and SV15a Rev respectively. The RFP module was generated by a 2 step

process. Primers RFP RBCas Fwd and RFP RBCas Rev were used to amplify the RFP gene found in BioBrick I13521 and primers C3 RBCas Fwd and C3 RBCas Rev were used to amplify the plasmid backbone (including promoter, RBS and terminator) from BioBrick K316004 and the resulting linear fragments joined by CPEC. This produced the RFP expressing regulator module which amplified from this plasmid using the SR Fwd and SR Rev primer pair.

SVa-X was generated by CPEC following PCR amplification of the *xylR* gene from the *E.coli* MG1655 genome with primers XylR g Fwd and XylRw g Rev and the SVa backbone with primers C3 RBCas Fwd and XylR v Rev. SVa-S was generated by CPEC following PCR amplification of the *rhaS* gene from MG1655 genomic DNA with primers RhaS g Fwd and RhaSw g Rev and the SVa backbone with primers C3 RBCas Fwd and RhaS v Rev.

SVa-X-18 (carrying promoter *xylF*) was generated by cutting SXVa with EcoRI and XbaI and ligating in the promoter containing fragment released when BioBrick plasmid I174018 was cut with EcoRI and SpeI. To generate SVa-S-rhaB a fragment carrying the sequence for the *rhaB* promoter was generated by PCR on the MG1655 genome with primers rhaB Fwd and rhaB Rev The fragment was cut with XbaI and SpeI and ligated into plasmid SVa-S previously cut with XbaI and dephosphorylated. Transformanted *E. coli* were spread plated on 0.5% rhamnose plates and screened for fluorescence. The CAP binding sites were removed from both *rhaB* and *xylF* promoters by round the plasmid PCR using the Prefix Rev primer with either the rhaB-nc Fwd or xylF-nc Fwd primer, with the resulting linear fragment phosphorylated prior to blunt end ligation. The sequences in the *xylF* promoter downstream of the transcription start site were removed by the same method using primers SCAR-RBS Fwd and xylF-ntss Rev. These steps generated plasmids SVa-S-rB and SVa-X-xF.

Primer	**Sequence**	**Template**	**Module**
SVa Primers			
RFP RBCas Fwd	***ACGGCAGGTATTCGGCTCCA***tactagagaaagaggagaaatactagatggc	I13521	StdReg
RFP RBCas Rev	***AGAGGTTCAGCCCGTAGTCC***ttattaagcaccggtggagtgac	I13521	StdReg
C3 RBCas Fwd	***GGACTACGGGCTGAACCTCT***tcacactggctcaccttcgg	K316004	StdReg
C3 RBCas Rev	***TGGAGCCGAATACCTGCCGT***gctagcataatacctaggactgagctagc	K316004	StdReg
SR Fwd	***CTCCAGCGGCGTAAAGGTCT***tttacagctagctcagtcctaggtattatg	StdRFP	StdReg
SR Rev	***GGTAACTCCCGCTGTAGACG***aaataataaaaaagccggattaataatctggc	StdRFP	StdReg
SV15a Fwd	***GGTAGCGTTCAGACTCCTCG***ACGCATCTTCCCGACAACG	pSB3K3	Ori-15

SV15a Rev	***ATCGGAACTCGGCTGCGTCA***gagataggtgcctcactgattaagc	pSB3K3	Ori-15
SVKan Fwd	***CGTCTACAGCGGGAGTTACC***CTGATCCTTCAACTCAGCAAAAGTTCG	pSB3K3	KanR
SVKan Rev	***TGACGCAGCCGAGTTCCGAT***CTCGAGTCCCGTCAAGTCAGC	pSB3K3	KanR
Exp Cas Fwd	***AGACCTTTACGCCGCTGGAG***gaattcgcggccgcttctaga	pAC4	Output
Exp Cas Rev	***CGAGGAGTCTGAACGCTACC***tataaacgcagaaaggcccacc	pAC4	Output
SVb Primers			
SV Ins Fwd	***GCCCTAGGTCTATGAGTGGTTGCTGGATAAC***gaattcgcggccgcttctagag	SVa	-
SV Ins Rev	***AGCAGATAGGGACGACGTGGTGTTAGCTGTG***ctccagcggcgtaaaggtct	SVa	-
SVc Primers			
Swa-B14-R Rev	***CGCGATTTAAAT***aaataataaaaaagccggattaataatctggc	SVb	-
SpeI-1r-R Fwd	***CTGCAGCGGCCGCTACTAGT***ctccagcggcgtaaaggtct	SVb	-
SpeI-Ins-G Rev	***CTGCAGCGGCCGCTACTAGT***cacagctaacaccacgtcgtcc	SVb	-
AscI-B15-G Fwd	***CGCGGGCGCGCC***tataaacgcagaaaggcccacc	SVb	-
AscI-6-O Rev	***CGCGGGCGCGCC***ggtagcgttcagactcctcg	SVb	-
FseI-5r-O Fwd	***CGCGGGCCGGCC***atcggaactcggctgcgtca	SVb	-
FseI-Kan-A Rev	***CGCGGGCCGGCC***ctcgagtcccgtcaagtcagc	SVb	-
SwaI-4r-A Fwd	***CGCGATTTAAAT***cgtctacagcgggagttacc	SVb	-
SV Reg primers			
C3 RBCas Fwd	***GGACTACGGGCTGAACCTCT***tcacactggctcaccttcgg	SVa	Regulator
RhaS v Rev	***TTGGTGAATCTTTTTTACCGGGGTA***tggagccgaatacctgccgt	SVa	Regulator
RhaS g Fwd	***TACCCCGGTAAAAAAGATTCACCAA***atgaccgtattacatagtgtggatttttttcc	MG1655 genome	Regulator
RhaSw g Rev	***AGAGGTTCAGCCCGTAGTCC*****tta**ttattgcagaaagccatcccgtc	MG1655 genome	Regulator
XylR v Rev	***AGGTGGGGATATTGGATGAGTTACG***tggagccgaatacctgccgt	SVa	Regulator
XylR g Fwd	***CGTAACTCATCCAATATCCCCACCT***atgtttactaaacgtcaccgcatcac	MG1655 genome	Regulator
XylRw g Rev	***AGAGGTTCAGCCCGTAGTCC*****ttat**tacaacatgacctcgctatttacatcg	MG1655 genome	Regulator
RhaS RBS primers			
RS_RBS25II_Fwd	***CCCTCTGCTTTGGCCC***tggagccgaatacctgcc	SVc-S	Regulator
RS_RBS25II_Rev	***GGGGCTACT***atgaccgtattacatagtgtggatttt	SVc-S	Regulator

RS_RBS50I_Fwd	***ATCCGGCCGCGTC***tggagccgaatacctgcc	SVc-S	Regulator
RS_RBS50I_Rev	***CAAGGGGTTCTT***atgaccgtattacatagtgtggatttt	SVc-S	Regulator
RS_RBS50II_Fwd	***GAGATAGCTTCCT***tggagccgaatacctgcc	SVc-S	Regulator
RS_RBS50II_Rev	***GGAGGCAATCAG***atgaccgtattacatagtgtggatttt	SVc-S	Regulator
p.xylF primers			
Prefix-Rev	CTCTAGAAGCGGCCGCGAATTCC	SVa-18, SVa-rhaB	Promoter
pxylF-nc Fwd	gaaagataaaaatctgtaattgttttcccctgtttagttg	SVa-18	Promoter
SCAR-RBS Fwd	TACTAGAGAAAGAGGAGAAATACTAGATGCGTAAAGGAG	SVa-18nc	Promoter
XylF-ntss Rev	taaccaatttttagcaactaaacaggggaaaac	SVa-18nc	Promoter
p.rhaB primers			
rhaB Fwd	***CTTCTAGAG***caggaatgcggtgagcatcacatc	MG1655 genome	Promoter
rhaB Rev	***CTGCAGCGGCCGCTACTAGTA***tacgaccagtctaaaaagcgcctgaattc	MG1655 genome	Promoter
rhaB-nc Fwd	atctttccctggttgccaatggcc	SVa-rhaB	Promoter
Sequencing			
NS1 F	AGACCTTTACGCCGC	-	Output
NS6 R	CGAGGAGTCTGAACG	-	Output
NS4 F	GTAACTCCCGCTGTAG	-	StdReg
NS1 R	TCCAGCGGCGTAAAG	-	StdReg
NS5 F	GACGCAGCCGAGTTC	-	Resistance
NS4 R	CGTCTACAGCGGGAG	-	Resistance
NS6 F	GGTAGCGTTCAGACTC	-	Origin
NS5 R	ATCGGAACTCGGCTG	-	Origin

Table 4.4 Primers used to generate sequences or assemble pieces into the SV series of plasmids.

Key: Reg is short for regulator, StdReg is short for Standard Regulator, Ori-15 is short for p15a origin, KanR is short Kanamycin resistance cassette. A schematic for the plasmid is shown in figure 6.1.

4.2.2.4 Cloning of SVb plasmid

SVb (Standard Vector b) was designed to remove a possible design mistake in the original SVa plasmid prior to characterisation with the plasmid. The output and regulator module promoters were close together and it was considered that large proteins binding at one promoter may affect transcription at the other. The upstream promoter insulator sequence from Davis et al. (2011) was inserted into SVa immediately upstream of the BioBrick prefix to add a distance barrier between the two promoters. Plasmid SVb was assembled from SVa by round the plasmid PCR using primers SVIns F and SVIns R followed by phosphorylation of PCR product and ligation with T4 ligase.

4.2.2.5 Cloning of SVc plasmids

A final modification to the SV series of plasmids was the addition of restriction site between the 4 modules to make it easier to clone and allow a degree of compatibility with pSEVA (Silva-Rocha et al., 2013). The SVc was the finished plasmid for all automated characterisation.

The main backbone plasmid (SVc) was generated in a two step process from SVb. First SVb was PCR amplified with primer pairs AscI-B15-G Fwd and SwaI-B14-R Rev and SwaI-4r-A Fwd and AscI-6-O Rev separately before restriction cloning via the AscI and SwaI sites. The resulting product was again amplified in two separate PCR reactions with primer pairs SpeI-1r-R Fwd and FseI-Kan-A Rev and FseI-5r-O Fwd and SpeI-Ins-G Rev prior to restriction cloning with SpeI and FseI. The resulting plasmid was sequenced in full using the sequencing primers in tables 4.3 and 4.4. The Anderson promoter variants (SVc-100, SVc-101 etc. to SVc-119) were generated using oligo annealing and restriction cloning. Oligos for the Anderson promoters were either already existing in the lab or ordered so that when annealed they would form fragments as if the promoters had been cut from BioBrick vectors. i.e. cut EcoRI and SpeI – see primer table 4.4 for details. The oligos were annealed, phosphorylated and ligated as detailed in section 4.2.2.

Regulator plasmids SVc-S, and SVc-X were generated by amplifying the respective SVa plasmid with primers SpeI-1r-R Fwd and SwaI-B14-R Rev and restriction cloning the resulting fragments with SwaI and SpeI into plasmid SVc. All the regulator plasmids were midi-prepped and fully sequenced to ensure the backbone was identical to the SVc backbone plasmid.

Reference plasmids SVc-X-101 and SVc-S-101 were generated by GFP module transfer from plasmid SVc-101. The GFP module of SVc-101 was amplified by PCR using primers AscI-B15-G Fwd

and SpeI-Ins-G Rev (10 cycles only – see general methods 4.2.2) and then restriction cloned into the regulator plasmids using EcoRI and BsaI or AscI.

The minimal *rhaB* promoter was transferred from SVa-S-rB by restriction cloning using EcoRI and BsaI, generating plasmid SSVc-rB. The ribosome binding site attached to *rhaS* in all the SVc-S plasmids was switched for new RBSs by round the plasmid PCR, generating plasmids SVc-S25II, SVc-S50I and SVc-S50II (with no promoter, promoter J23101 and promoter *rhaB* variants for each vector).

The minimal *xylF* promoter was transferred from SVa-X-xF by restriction cloning with EcoRI and BsaI into SVc-X, generating SVc-X-xF.

Primer	**Sequence**
100-O-F	AATTCGCGGCCGCTTCTAGAGttgacggctagctcagtcctaggtacagtgctagcTA
100-O-R	CTAGTAGCTAgctagcactgtacctaggactgagctagccgtcaaCTCTAGAAGCGGCCGCG
102-O-F	AATTCGCGGCCGCTTCTAGAGttgacagctagctcagtcctaggtactgtgctagcTA
102-O-R	CTAGTAGCTAgctagcacagtacctaggactgagctagctgtcaaCTCTAGAAGCGGCCGCG
103-O-F	AATTCGCGGCCGCTTCTAGAGctgatagctagctcagtcctagggattatgctagcTA
103-O-R	CTAGTAGCTAgctagcataatccctaggactgagctagctatcagCTCTAGAAGCGGCCGCG
104-O-F	AATTCGCGGCCGCTTCTAGAGttgacagctagctcagtcctaggtattgtgctagcTA
104-O-R	CTAGTAGCTAgcacaatacctaggactgagctagctgtcaaCTCTAGAAGCGGCCGCG
105-O-F	AATTCGCGGCCGCTTCTAGAGtttacggctagctcagtcctaggtactatgctagcTA
105-O-R	CTAGTAGCTAgcatagtacctaggactgagctagccgtaaACTCTAGAAGCGGCCGCG
106-O-F	AATTCGCGGCCGCTTCTAGAGtttacggctagctcagtcctaggtatagtgctagcTA
106-O-R	CTAGTAGCTAgctagcactatacctaggactgagctagccgtaaaCTCTAGAAGCGGCCGCG
107-O-F	AATTCGCGGCCGCTTCTAGAGtttacggctagctcagccctaggtattatgctagcTA
107-O-R	CTAGTAGCTAgcataatacctagggctgagctagccgtaaaCTCTAGAAGCGGCCGCG
108-O-F	AATTCGCGGCCGCTTCTAGAGctgacagctagctcagtcctaggtataatgctagcTA
108-O-R	CTAGTAGCTAgcattatacctaggactgagctagctgtcagCTCTAGAAGCGGCCGCG
109-O-F	AATTCGCGGCCGCTTCTAGAGtttacagctagctcagtcctagggactgtgctagcTA
109-O-R	CTAGTAGCTAgcacagtccctaggactgagctagctgtaaaCTCTAGAAGCGGCCGCG
110-O-F	AATTCGCGGCCGCTTCTAGAGtttacggctagctcagtcctaggtacaatgctagcTA
110-O-R	CTAGTAGCTAgcattgtacctaggactgagctagccgtaaaCTCTAGAAGCGGCCGCG
111-O-F	AATTCGCGGCCGCTTCTAGAGttgacggctagctcagtcctaggtatagtgctagcTA
111-O-R	CTAGTAGCTAgcactatacctaggactgagctagccgtcaaCTCTAGAAGCGGCCGCG

112-O-F	AATTCGCGGCCGCTTCTAGAGctgatagctagctcagtcctagggattatgctagcTA
112-O-R	CTAGTAGCTAgcataatccctaggactgagctagctatcagCTCTAGAAGCGGCCGCG
113-O-F	AATTCGCGGCCGCTTCTAGAGctgatggctagctcagtcctagggattatgctagcTA
113-O-R	CTAGTAGCTAgctagcataatccctaggactgagctagccatcagCTCTAGAAGCGGCCGCG
114-O-F	AATTCGCGGCCGCTTCTAGAGtttatggctagctcagtcctaggtacaatgctagcTA
114-O-R	CTAGTAGCTAgctagcattgtacctaggactgagctagccataaaCTCTAGAAGCGGCCGCG
115-O-F	AATTCGCGGCCGCTTCTAGAGtttatagctagctcagcccttggtacaatgctagcTA
115-O-R	CTAGTAGCTAgcattgtaccaagggctgagctagctataaaCTCTAGAAGCGGCCGCG
116-O-F	AATTCGCGGCCGCTTCTAGAGttgacagctagctcagtcctagggactatgctagcTA
116-O-R	CTAGTAGCTAgcatagtccctaggactgagctagctgtcaaCTCTAGAAGCGGCCGCG
117-O-F	AATTCGCGGCCGCTTCTAGAGttgacagctagctcagtcctagggattgtgctagcTA
117-O-R	CTAGTAGCTAgcacaatccctaggactgagctagctgtcaaCTCTAGAAGCGGCCGCG
118-O-F	AATTCGCGGCCGCTTCTAGAGttgacggctagctcagtcctaggtattgtgctagcTA
118-O-R	CTAGTAGCTAgcacaatacctaggactgagctagccgtcaaCTCTAGAAGCGGCCGCG
119-O-F	AATTCGCGGCCGCTTCTAGAGttgacagctagctcagtcctaggtataatgctagcTA
119-O-R	CTAGTAGCTAgcattatacctaggactgagctagctgtcaaCTCTAGAAGCGGCCGCG

Table 4.5 Oligos annealed to form Anderson promoters.

The lower case sequence is the promoter itself and when annealed the upper cases sequences form pre-cut EcoRI and SpeI restriction sites.

4.2.2.6 Cloning of Synthetic Promoters

To allow generation of the synthetic xylose inducible promoter library, certain members of the Anderson collection were ligated into SVc-X cut with EcorI and BsaI by the same method used to make the Anderson variant plasmids. The xylose library was generated from these plasmids by round the plasmid PCR using the respective forward primers (i.e. Xyl UP 1 106 for SVc-X-106) and one of the two reverse primers (Xyl UP 1 Rev or Xyl UP 2 Rev) prior to phosphorylation and ligation of the linear product.

Primer	Sequence	Promoter
Xyl UP 1 Rev	CGATTACGATTTTTGGTTCTCTAGAAGCGGCCGCGAATTCC	All X1 and X2
Xyl UP 2 Rev	CTATTGAGATAATTCACACTCTAGAAGCGGCCGCGAATTCC	All X3
Xyl UP 1 106 Fwd	AAAGATAAAAATCTGTAAtttacggctagctcagtcctaggtatag	106X1, 106X2
Xyl UP 1 113 Fwd	AAAGATAAAAATCTGTAActgatggctagctcagtcctaggg	112X1, 112X2
Xyl UP 1 114 Fwd	AAAGATAAAAATCTGTAAtttatggctagctcagtcctaggtacaatg	113X1, 113X2
Xyl UP 1 114a Fwd	AAAGATAAAAATCTGTAAtttatggctagctcagtccttggtac	114X1, 114X2
Xyl UP 1 114b Fwd	AAAGATAAAAATCTGTAAtttatggctagctcagtcctaggtataatg	114aX1, 114aX2
Xyl UP 1 112 Fwd	AAAGATAAAAATCTGTAActgatagctagctcagtcctagggattatg	114bX1, 114bX2
Xyl UP 1 116_117 Fwd	AAAGATAAAAATCTGTAAttgacagctagctcagtcctaggg	116X1, 116X2, 117X1, 117X2
Xyl UP 2 106 Fwd	CAGTGTGAAATAACATAAtttacggctagctcagtcctaggtatag	106X3
Xyl UP 2 113 Fwd	CAGTGTGAAATAACATAActgatggctagctcagtcctaggg	112X3
Xyl UP 2 114 Fwd	CAGTGTGAAATAACATAAtttatggctagctcagtcctaggtacaatg	113X3
Xyl UP 2 114a Fwd	CAGTGTGAAATAACATAAtttatggctagctcagtccttggtac	114X3
Xyl UP 2 114b Fwd	CAGTGTGAAATAACATAAtttatggctagctcagtcctaggtataatg	114aX3
Xyl UP 2 112 Fwd	CAGTGTGAAATAACATAActgatagctagctcagtcctagggattatg	114bX3
Xyl UP 2 116_117 Fwd	CAGTGTGAAATAACATAAttgacagctagctcagtcctaggg	116X3, 117X3

Table 4.6 Primers used to generate synthetic xylose inducible promoters.

The initial promoter or BioBrick prefix binding sequence is in lower case and the new operator sequence in upper case.

4.3 Manual characterisation experiments

4.3.1 Ideal growth condition experiments

An initial experiment was carried out to identify what the ideal conditions were for growing *E. coli* in microplates and how this related to flask growth used historically and industrially during scale up. The growth rate experiments were carried out with Top10 *E. coli* carrying plasmids pAC1 and pAC4 (flask) and pACf, pACf-101, pACf-narK, pACf-dmsA and pACf-ndh (microplate). Overnight cultures (2ml) were set up in LB with kanamycin and the following morning diluted 1 in 100 into fresh LB media to approximate OD of 0.1. Diluted samples were then transferred to microplates (100µl and 200µl per well), 15ml BD falcon tubes or 250ml Ehrlenmyer flasks (20ml or 100ml in flask). The samples were then grown at 37°C in either a Thermo MaxQ 6000 incubator (0 or 200rpm) or a Labnet Vortemp 56 microplate incubator (500rpm) and optical density measurements taken approximately every hour. Microplates were covered with ringed lids (Costar) to prevent evaporation. Optical density of cultures at 600nm was determined using one of two pieces of equipment; the tube and flask cultures were measured using a Thermo Spectronic BioMate5 spectrophotometer whereas microplate samples were measured in the BMG Labtech Omega plate reader (with path length correction included). Fluorescence readings were taken by a BMG labtech Omega plate reader using the 485nm and 510nm excitation and emission filters with a gain of 1400.

4.3.2 Bit to Atom to Bit Characterisation (in collaboration with the BioFAB)

The Bit to Atom to Bit experiment was devised collaboratively between the BioFAB and members of the CSynBI to test methods for characterising and sharing the data produced by characterisation. The idea was take some of the BioFABs characterisation and construct data which was available digitally on line (i.e. in electronic bits) and build the physical DNA required to repeat the characterisation at the CSynBI (in atoms). The results of the characterisation should be reported back to the BioFAB via an electronic means (in bits, completing the bit to atom to bit loop).

DH10B *E.coli* was used for the characterisation in the BioFAB bit to atom to bit experiment. Cultures of E. coli carrying the pAC1, pAC4, pAC5 and pAC7 plasmids were grown overnight at 225rpm and 37°C in 3ml of LB in 15ml tubes with kanamycin (3 tubes per sample from separate colonies). The following morning cultures were diluted 1 in 60 into fresh pre-warmed LB or supplemented M9 media with 0.4% (w/v) glycerol and outgrown at 37°C and 225rpm in 15ml tubes.

After 90 minutes cultures were diluted directly into a black Greiner microplate with the dilution dependent upon the media. The outer wells at the edge of the microplate were filled with 200µl of water and only the central wells were used for growing samples in order to remove any edge effect. The dilutions were as follows; 30µl of M9 sample into 90µl fresh M9 and 20µl of LB sample into 100µl fresh LB. The samples in the plate were then grown inside the BMG-Labtech Omega plate reader at 37°C with 60 seconds of shaking (500rpm) immediately prior to taking a measurement. The plate was measured every hour for GFP (485,510; gain 1400), RFP (585/610; gain 2600) and absorbance (600nm). After 3 hours and 5 hours samples were taken (5µl per well) and transferred into a Costar 96-well microplate pre-filled with 195µl per well of water.

The Costar plates were kept on ice before being taken to a modified Becton-Dickinson FACScan flow cytometer with Cytek automated microplate sampler (AMS) adaptor. Samples were run on high flow rate until 20000 events within a gate corresponding to *E. coli* were observed or 30 seconds had elapsed. Flow cytometry settings used were all taken on the log scale; FSC sensor E01, SSC voltage350, FL1 voltage 700, threshold 52 (SSC).

4.3.3 Testing *E. coli* growth in rich MOPS media in microplates

As MOPS was a more suitable media for characterisation it was essential to test how cells would grow in the new media. Top10 *E.coli* was used for characterisation trials in MOPS EZ media containing 0.4% (w/v) glucose or glycerol. *E. coli* colonies containing plasmid pAC1 or pAC4 were grown overnight at 37°C and 225rpm in 3ml of LB with kanamycin. The following morning cultures were diluted 1 in 100 into fresh MOPS (containing either glucose or glycerol) or LB media (1ml in 15ml tube) and outgrown at 37°C and 225rpm for 90 minutes. After outgrowth cultures were again diluted into fresh MOPS media or LB media to 0.1 OD units and transferred to a Greiner microplate. The samples were grown in a Modified Mikura plate incubator at 37°C and 600rpm shaking for 5 hours. The microplate was covered with a ringed microplate lid while incubated. Every 30 minutes the plate was transferred to a Biotek Synergy HT plate reader and assayed for both GFP fluorescence (filters: 485±20 and 528±20, sensitivity: 79) and absorbance (600nm).

4.3.4 Identification of promoter induction range

Prior to full characterisation the range of induction was determined in MG1655 *E. coli* and supplemented M9 media with 0.4% glucose. Cultures of samples to be tested were set up in LB containing kanamycin and grown overnight and transferred to ice the following morning. In the afternoon the samples were diluted into M9 media in a costar clear microplate and outgrown at 30°C for 90 minutes with a microplate lid to prevent evaporation. Following 90 minutes outgrowth samples were transferred to a black 96 well Greiner microplate and immediately induced with a wide range of inducer concentrations. The microplate was then sealed with a breathe-easy seal (Sigma) and the microplate scanned on a repeating programme of fluorescence and absorbance measurements every 10 minutes overnight in the Synergy HT microplate reader (GFP filters: 485±20 and 528±20, sensitivity: 79). The assay was repeated until inducer concentrations were obtained which either covered the entire dose-response curve of the inducer or until the inducer began to significantly affect sample growth rate.

4.4 Relative Promoter Unit data analysis

4.4.1 Plate reader data

All characterisation experiments were run with 3 controls; media wells, a negative control – generally plasmid containing no GFP or an unexpressed GFP (usually pAC1 or SVc) and a reference construct – a GFP encoding transcript identical to the transcript of the tested BioPart under control of the J23101 reference promoter (Kelly et al., 2009). The three controls were required to enable accurate data analysis and calculate results in RPU. While generally absorbance data was collected some protocols called for collection of optical density data and this was handled the same way.

Background absorbance from the media was calculated by averaging the absorbance value for the media wells at each time point. This value was then subtracted from all sample wells where the sample was grown in that media to give the background corrected absorbance. The resulting background corrected absorbance values were plotted as the raw absorbance trace.

The auto-fluorescence was calculated to remove this from fluorescence data. It was observed that the relationship between absorbance and auto-fluorescence was complicated and so an equation fitting approach also used by Davis et al. (2011) was employed. For all time points the

background corrected absorbance and raw fluorescence results for the negative controls were averaged and the resulting values plotted against each other. A linear or second order quadratic was then fitted to the data with the equation of the fit indicating the auto-fluorescence of cells and media at any given absorbance. The background corrected absorbance for each sample and time point was then used to calculate each sample's auto-fluorescence which was then subtracted from the respective raw fluorescence, generating the background corrected fluorescence.

The growth rate and doubling time were initially calculated at the 3 hour time point using the following equations (after conversion of absorbance to Optical Density):

$$Growth\ Rate\ (GR) = \ln(cOD_i / cOD_{i-1})$$

$$Doubling\ time\ = (t_i - t_{i-1}) \times \frac{\log(2)}{GR_i}$$

Where cOD was the corrected Optical Density for the sample and t was the measurement time. This growth rate is also used as part of the flow cytometry analysis (see section 3.5.2). For automated characterisation experiments the Growth Rates software by Hall et al. (2014) was used as this proved to be more effective and reliable method of calculating growth rate.

The synthesis rate of fluorescence production was then calculated using the following equation from Kelly et al. (2009):

$$Synthesis\ rate\ (SS_{Sample}) = \frac{cFl_i - cFl_{i-1}}{(cAbs_i + cAbs_{i-1})/2} / (t_i - t_{i-1})$$

Where cFl was the corrected fluorescence for the samples

The promoter strength in Relative Promoter Units (RPU) was calculated from the synthesis rates using the following equation from Kelly et al. (2009):

$$Sample\ output\ (RPU) = \frac{SS_{sample}}{SS_{J23101}}$$

Where SS_{J23101} was the synthesis rate at that time point for the J23101 reference construct on the same day the sample was characterised. The promoter output in RPU was then calculated by averaging the calculated sample outputs. The error was calculated as a single standard deviation of those sample outputs.

Where curve fitting was required this was carried out using the fit function in matlab.

4.4.2 Flow cytometry data

Flow cytometry data analysis was carried out using either Cyflogic or Flowjo software. Raw flow cytometry files were gated in side scatter and forward scatter dimensions to remove non-cellular signal. The exact region gated depended upon the species of *E. coli* acting as a chassis for characterisation and for automated characterisations the time point at which data was collected. For DH10B *E. coli* gating was between 5.6 and 626.5 forward scatter and 6.9 and 181 side scatter. For data from MG1655 *E. coli* samples the gate was set for each set of experiment individually. Following gating the geometric mean and coefficient of variation for the FL1 (GFP) distribution for each sample were determined.

The background fluorescence was removed by subtraction of the average geometric mean of the negative control cells. The output of each sample in RPU was then derived by using the following equation taken from Kelly et al. (2009):

$$Sample\ output\ (RPU) = \frac{Synthesis\ rate_{Sample}}{Synthesis\ rate_{J23101}} = \frac{Geometric\ mean_{Sample}}{Geometric\ mean_{J23101}} \times \frac{Growth\ rate_{Sample}}{Growth\ rate_{J23101}}$$

Where each promoter output is calculated for each sample using the average reference (J23101) output for the respective day. The growth rate was only factored into the equation where there was a clear difference in growth rate. When this was the case the growth rate for each sample was the value calculated from its absorbance trace from the plate reader data. The resulting RPU values were averaged to obtain the promoter output in RPU characteristic and the error in the promoter output was a single standard deviation of the calculated RPU values.

5 Automated platform set up and characterisation methods

5.1 Theonyx platform

The platforms on which all automated experiments were designed, tested, debugged and run were Aviso-GMBH Theonyx platforms (based on Sias Xanthus hardware). The CSynBI had 2 Theonyx platforms and the majority of this work was carried out on the platform that was dedicated for characterisation and has modifications to aid this (see section 5.2). Both platforms were equipped with multiple pipetting blocks, tip racks and a Synergy HT plate reader (Biotek) as standard and were fully capable of handling microplates. The characterisation platform was initially set-up with pipetting needles.

The second Theonyx platform was a dedicated assembly platform and so was pre-calibrated and taught. This platform was used for the semi-automated characterisation (see 5.6.1) while the characterisation platform was being calibrated and set-up. The second platform came equipped with Tecan tip heads and was outfitted for disposable tips use. The only items used on this platform were the basic pipetting slots, tip drop station, tip racks, reagent troughs and the Synergy HT reader.

5.2 Platform modifications

Additional modules were added to the characterisation platform to make it more suitable for cell growth and assays. The pipetting needles were replaced with Tecan pipetting heads (Aviso-GMBH) to allow disposable tip use and thus prevent cross-contamination. To complement this a tip drop station was supplied by Dr. R. Weinzerl. A BioShake 3000 (Quantifoil Instruments) heated plate shaker was added to the platform to act as a heated pipetting position. A tube holding position was made from a polystyrene tube rack and an extra unnecessary raised pipetting position. Finally a robot useable incubator was made by modifying a Mikura Ventura 2000 plate incubator. The incubator lid was cut into two halves and the hinge removed with blocks to allow the gripper arm to manipulate the lid. A frame for 4 microplates was fitted inside of the incubator to hold plates in place during shaking. The incubator was clamped to the platform underneath to prevent movement relative to the rest of the platform.

5.3 Platform set up and GUI programming

5.3.1 General information and software

The Theonyx platforms were programmed using the Robo Manager GUI programmed by Aviso-GMBH which contains an in built database. Calibration was carried out using the Sias X-Util and scaling calculator firmware software (see 5.3.7).

5.3.2 Deck organisation

To enable accurate disposable tip usage each pump was calibrated by updating firmware scaling and movement values using the X-Util software. For details of this calibration see 5.3.7. The deck and modules carried on it were arranged so pipetting related blocks and tasks are carried out near the tip drop point at the far left of the platform and incubator handling at the far right. The deck was arranged in 6 bands left to right. Slots were labelled from the back of the platform to the front A to E (excluding the incubator). See figure 5.1 for a map of the deck. Slots B2 and C2 were used for pipetting in addition to slot D3 which is the heated slot. Slots D2 and E2 were used to place output plates for removal from the platform or microplate storage spots. Slots A2 and A3 were used for stacking microplates (up to 4 each) where many plates are required. The tube rack was located at slot B3 and tip racks in the 6 slots located B4 to D5. Plate lids for the incubator were kept in the 6th band.

5.3.1 Position teaching

Co-ordinates were taught for all tubes (wells are classed the same as tubes), tips and plates by visually checking for accuracy of positioning. Movement co-ordinates were also taught visually for plates and plate lids. For pipette tip use teaching 200µl Starlab tips were used, the GUI automatically calculated the positions for the other tip types using the reference positions. Pipette teaching was verified using a programme to ‘waste’ water from the taught trough/plate/tube. Tip rack positions were taught for each tip type separately and were initially taught by eye then verified (particularly depth) by running a programme that uses all the tips in a given rack (often the ‘waste’ programme mentioned above).

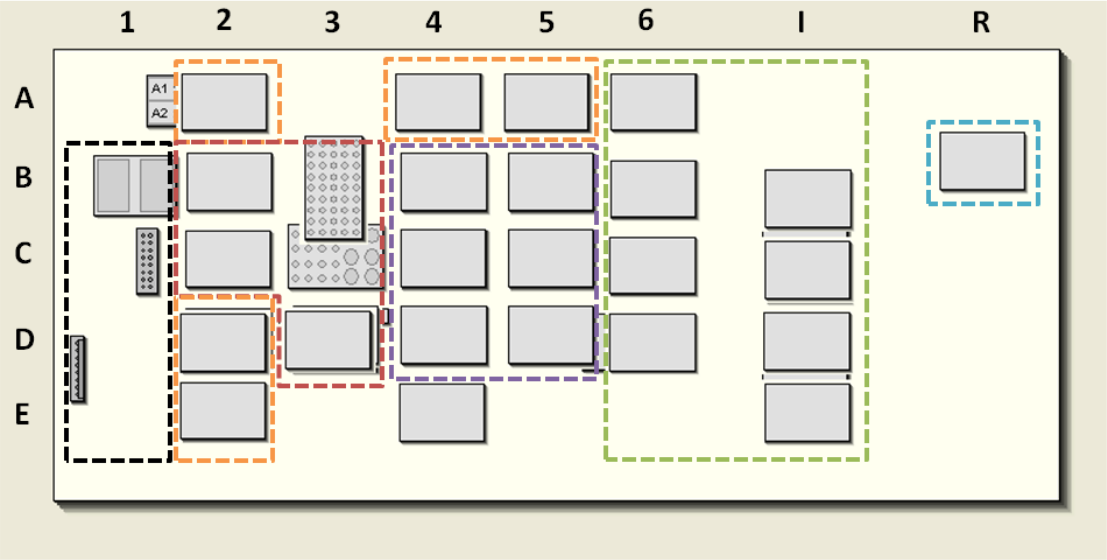


Figure 5.1 Deck layout of the Automated characterisation platform.

Items that are commonly used together were kept in similar areas to reduce movement times with pipetting carried out at the left hand of the platform and incubation and measurement on the right hand end. Boxed areas correspond to positions with the same or related functions as follows: Cyan box – Plate reader, Green box – Incubator and lid storage slots, Orange – plate storage (back of platform) and plate output positions (front of platform), Purple box – Tip racks, Red Box – Pipetting positions and 2 tube slots, Black box – Reagent (media) troughs, waste and wash position and tip drop station (in that order to prevent contamination).

5.3.2 Pipetting and mixing steps

Regular flushing between pipetting steps was required to ensure accurate pipetting. For most pipetting steps a system air gap and small amount of sample was used as a 'waste' volume to ensure accurate pipetting. The volumes of the air gap and waste were governed by the volume transferred. Generally air gaps were: 10µl for 20µl tips, 20-50µl for 200µl tips and 50-200µl for 1ml tips. The waste volume used was determined by the volume to be aspirated (generally this was 5µl for volumes 20µl or lower, 20µl for volumes 20µl to 200µl and 50µl for volumes above 200µl). Following pipetting the air gap and waste were dispensed at the wash station. The pipetting speed was set to pipette the volume required in one second unless the volume was under 10µl or over 450µl (the pump speed minimum and maximum respectively). The pipetting function of the platform was used to mix samples by cycles of aspirate and dispense.

5.3.3 Absorbance based dilution steps

The database built into Robo Manager allowed the use of absorbance based dilutions. The plate reader scanned an entire microplate for absorbance at 600nm before blanking with a pre-determined media absorbance value. The results were automatically uploaded into the database via a text file intermediate with a standardised format. The platform was capable of carrying out two different kinds of dilutions; one step dilution (addition of more media) and two step dilution (transfer to fresh media). Both dilutions were used in protocols and carried out with a 200µl tip, 100µl of air gap, pipetting speed of 100µl per second and a waste volume of 20µl.

5.3.4 Automated use of the Synergy HT plate reader

The plate reader was operated by the Robo Manager software running the external KC4 call software (KC4call.exe) and protocol files set up in advance using the plate reader's KC4 software (Biotek). Each command sent to the plate reader was sent in the form of 3 parameters; protocol (PRO), plate file (PLT) and export file (EXP). Mathematical procedures such as blanking were implemented in the protocol file prior to export of results from KC4. The Robo Manager database imported data from the text files as a list of samples with a specific header. As the fluorescence bulb required warming to take reliable measurements a separate protocol was run prior to transferring a microplate to the plate reader for a fluorescence measurement.

5.3.5 Platform pipetting calibration

Calibration was carried out for aspiration and dispense of liquid separately for each pipetting pump. This was achieved by measuring the actual volume taken up or ejected with a given tip and updating the motor scaling and move values inside the platform firmware. Calibration was carried out for each pump until the results were as good as or better than those specified by the manufacturer. Calibration was carried out for the maximum volume for each tip generally used i.e. 1000µl for 1000µl tips, 200µl for 200µl tips etc. with the appropriate tip firmly pushed onto the tip head. All calibrations were carried out with pipette speed and waste volume similar to those that would be used in regular use.

Calibration of the aspirate and dispense settings for each pump were carried out in two different ways depending upon volume. For the 1000µl dispense and the 200µl and 1000µl aspirate calibrations water and a precision scale were used. The balance was set up to allow access from pipetting heads and a beaker of water filled and measured on the scale. The beaker was then lifted to allow the tip to enter the water and the pump would aspirate or dispense liquid before the beaker was returned to the scale and the change measured. The weight difference was converted into microlitres (assuming 1µl = 1mg) and after 5 repeats the average result was used to update the pump firmware.

Calibration for lower volumes of dispense and aspirate were carried out using the Biotek QC check dye diluted to yield an absorbance of approximately 2.5 arbitrary absorbance units in 200µl of liquid. For the 200µl dispense 250µl of dye was aspirated per tip from a reagent trough and 200µl dispensed into a microplate. The absorbance was then measured to allow accurate calculation of dispensed volume. For aspirate and dispense 20µl the appropriate volume of dye was transferred within a microplate and the absorbance values before and after compared to determine the volume picked up or transferred. Following a round of 5 repeats the pump calibrations would be updated and then re-tested until the pump performed better than manufacturer specification.

5.4 Calibration of plate reader and flow cytometry measurements

5.4.1 Plate Reader fluorescence calibrations

The synergy plate reader has a PMT for detecting fluorescence that was set as part of the protocol to various sensitivities. A set of sensitivities which cover a range sensible for GFPmut-3b fluorescence inside cells were calibrated and monitored for variation over time by the use of a CLS96M (Spectro LLC) testing plate. The standard sensitivities calibrated for use and absolute unit determination were 50, 60, 70, 75, 80 and 85 (top optic). The results from the plate reader were arbitrary fluorescence units and absorbance which were not suitable for inter lab comparisons. To make this possible it was necessary to develop a calibration for the arbitrary results into absolute unit results which would be the same in other labs. He units chosen were molecules of GFP per cell and equivalent units of sodium fluorescein.

5.4.1.1 GFPmut-3b Protein purification

Plasmid pProEX-htb-GFPmut-3b which carried a gene encoding his tagged GFPmut-3b protein was transformed into BL21 *E. coli* for protein expression. A 5ml overnight culture of LB and ampicillin was used to inoculate a 500ml flask of LB with ampicillin. This flask was grown until the OD of culture reached 0.5 at which point IPTG was added to a concentration of 0.5mM to induce GFP-expression. After 4 hours the culture was centrifuged at 3000 × g for 15 minutes in a Thermo Sorval RC6 Plus centrifuge. Cells were washed with PBS and centrifuged again and then kept at -80°C ready for sonication. Prior to sonication cells were resuspended in 20ml of PBS with protease inhibitors and then sonicated on a 15 minute programme of 30 second on then 30 seconds off. The GFPmut-3b protein was then purified using an ÄKTA purification system by Ciaran Mckeown. Fractions obtained during purification were assessed for purity by PAGE gel and the two fractions (9 and 10) which appeared to contain little or no contaminating protein were kept for testing. Purity for these fraction appeared to be >90% on a brilliant blue stained gel.

5.4.1.2 GFP Protein Quantification

The fractions taken for testing were assayed for protein concentration by multiple methods. Initially fractions were tested by Bradford assay (Pierce) following the manufacturer's protocol. A series of dilutions of fraction 9 were also evaluated by absorbance at 280nm in a Greiner UV-star half area microplate to get a second measure of protein concentration. Both methods quantified the GFP per were not considered accurate enough for use in calibration. Based on this it was decided to quantify GFP protein using a Modified Lowry Kit (Pierce) as this was more accurate at lower protein concentrations. Samples and dilutions of both fraction 9 and 10 were assayed using the kit according to the manufacturer's protocol (microplate method).

5.4.1.3 Quantification of fluorescent signal from GFP in cells and lysate

As the exact environment inside cells is unknown and may impact upon GFP fluorescence it was necessary to quantify fluorescence in lysate and convert this to the fluorescent signal observed in the same cells used to make lysate. Lysate was made chemically using B-PER II lysis buffer (Pierce) according to a modified version of the manufacturer's protocol. Efficiency of lysis was determined by replica plating of lysed and unlysed *E. coli* cultures.

First the fluorescence from known amounts of GFPmut-3b in cell lysate was measured. The cell lysate used for this was generated from MG1655 *E. coli* not expressing GFP. Cell lysate was generated from MG1655 cells carrying plasmid SVc (which does not express GFP) which were grown in 5ml LB media cultures with antibiotic to an optical density of approximately 1 at 600nm. Cells were then spun down at 2683 × g for 10 minutes and resuspended in PBS. The cells in suspension were spun again with the same settings and resuspended in approximately 5ml of PBS. Following this the optical density was measured and the cells diluted to exactly 0.5 optical density units. Cells were lysed by adding an equal volume of B-PER II lysis buffer (Pierce).

Dilutions of purified GFPmut-3b in PBS were carefully pipetted to cover a range similar to those observed in experiments (based on purified GFP fluorescence in PBS). 40µl of these GFPmut-3b dilutions were pipetted into wells of a Greiner black walled 96-well microplate and then 160µl of the lysate with care taken to not generate bubbles in the wells. The microplate was then scanned on the standard fluorescence sensitivities (see above). The results were background corrected and a

curve fitted to determine the relationship between volume added (and thus molecules of GFP present) and fluorescence signal.

To complete the quantification and allow result calibration it was necessary to determine the change in fluorescence signal caused by cell lysis. To determine this MG1655 *E. coli* carrying either SVc, SVc-101, SVc-106, SVc-110 and SVc-111 or SVc, SVc-100, SVc-101, SVc—102, SVc-106, SVc-114 and SVc-114b were grown overnight in LB before dilution (1 in 50) into 3ml of 0.4% glucose MOPS medium. Samples were grown for around 4 hours before being placed on ice and spun at 2683 × g for 10 minutes. Following centrifugation samples were washed with PBS and centrifuged again using the same settings. Following this step the samples were resuspended in 1ml of fresh 0.4% glucose MOPS medium. Sample optical density was measure at 600nm and the samples were then diluted to an optical density of 0.5 units. 100µl samples were transferred into duplicate wells onto a Greiner black walled 96-well microplate. For each pair one sample was diluted with 100µl of MOPS media while the other sample was lysed by addition of B-PER II buffer. The plate was then scanned for GFP at the standard sensitivities (see above) and a curve fitted to the results following removal of background signal.

The two curves were then combined to determine the relationship between the number of molecules of GFP inside unlysed MG1655 cells in MOPS medium and fluorescence as observed by the plate reader at standard sensitivities.

5.4.1.4 Sodium fluorescein calibration

Sodium Fluorescein was diluted in DPBS (~ pH7.35) to generate stock solutions at concentrations of 5µg/ml, 4µg/ml, 2µg/ml and 1µg/ml. 10 fold serial dilutions were carried out to generate a range of concentrations from 5µg/ml down to 0.1ng/ml. Triplicate repeats for each solution were transferred to a Greiner black microplate (100µl per well) along with DPBS as a background control and scanned for GFP fluorescence with the range of fluorescence sensitivities. Following background correction a curve was fit to determine the relationship between observed fluorescence and Fluorescein concentration.

5.4.2 Flow cytometer day to day calibration

Initially standardised flow settings were used day to day to attempt to keep results consistent. Starting with the *rhaB* rhamnose inducible promoter dataset RCP-30-5A flow calibration particles from Spherotech were used for day to day flow cytometry calibration. Before a dataset was run a tube containing 500µl of sterile water and 3 drops of calibration particles were run on the flow cytometer. The instrument voltages were altered until the maximum channels (over 20 seconds) for the particles peaks matched as closely as possible with the standard channel set: 2, 15, 38, 113 273, 798, 2072 and 4216 for peaks 1 to 8 respectively. The settings to be used for flow cytometry were then updated prior to running samples.

5.4.3 Absorbance to cell number calibration

Plate readers generate arbitrary absorbance results. To finish result calibration into absolute units it was necessary to convert these arbitrary results into an absolute equivalent with the most logical being the number of cells in the sample.

Cell numbers were counted using Accucount cell counting beads (Spherotech). MG1655 *E. coli* carrying plasmid SVc-101 were grown overnight in 1ml of LB media containing Kanamycin. The following morning the *E. coli* were diluted down to 0.01 absorbance units (100µl, 600nm) in 0.4% glucose MOPS media and transferred to a Greiner black microplate (100µl per well in the first 11 columns with the final column of wells filled with 100µl of media). The *E. coli* were grown for the next 8 hours by the automated platform as if a characterisation experiment were being run with measurements taken every 15 minutes. Samples were taken every 45 minutes and diluted until there were 10-100 fold more cells than 25µl of counting beads so to give accurate results. The diluted cells were mixed with 25µl of cell counting beads into a final volume of 200µl and then run on the flow cytometer through the AMS with the following settings; FSC – E00, SSC – 300 (Threshold 40), FL1 - 640, FL2A - 670, FL2W - 600.

The resulting data was gated in the FL1 and SSC channels to exclude events which were not cells or beads. Following this the two populations were identified by the differences in their FL1 (GFP) and FL2A (RFP) signals and gated separately.

The number of cells in the sampled wells was then calculated using the following equation:

$$Number\ of\ cells\ in\ sampled\ wells = \frac{Cell\ events}{Bead\ events} \times Number\ of\ beads$$

The number of cells in the sampled wells was then multiplied by the dilution factor to yield the number of cells responsible for the absorbance values measured in the Greiner plate.

5.4.4 Applying the calibrations to generate absolute unit results

Data analysis was carried out to generate synthesis rates using the same steps as in methods 4.4.1 with the inclusion of the arbitrary to absolute unit conversion following the background signal removal step. Calibration of absorbance was performed using the following equation:

$$Number\ of\ cells = cABS\ \times\ 5.51E08\ +\ 5.66E07$$

Where cABS is the background corrected absorbance.

The fluorescence calibration equation was applied after the auto-fluorescence was removed by background subtraction but before any further analysis was performed. The equations for these calibrations were:

$$Molecules\ of\ GFP = cFl\ \times GsF$$

$$Equivalent\ fluorescein\ concentration = cFl\ \times fG + fI$$

Where cFl is the background corrected fluorescence signal, GsF is the GFP scaling factor for the used sensitivity setting , fG is the gradient of the fluorescein calibration curve and fI the intercept of the fluorescein curve (where necessary).

Following these calibrations synthesis rates were calculated in absolute units.

5.5 Aerobic growth verification

5.5.1 Aerobic growth verification

5.5.1.1 Demonstrating fnr biosensor activity

The initial tests did not clearly prove that the *fnr* based biosensors were indicating the presence of aerobic or anaerobic metabolism or reduced oxygen levels inside cells. A clearer test where oxygen availability would be restricted was required. An experiment was set up to reduce the surface area of a culture was devised as this would also limit oxygen availability. The *fnr* carrying minimal plasmids pACf, pACf-ndh, pACf-narK and pACf-dmsA were transformed into Top10 *E. coli.* Cultures were set off for each sample (4 each) and grown overnight at 37°C and 700rpm in a microplate (200µl per well). The following morning cultures were diluted down to 0.01 absorbance units (200µl volume) in LB containing 1% (W/V) glycerol in wells in a Greiner black 96 well microplate. The samples were grown for 6 hours at 37°C and 700rpm in the robot incubator as part of an automated protocol that transferred the microplate to the reader for absorbance and fluorescence measurements using sensitivity 79 every 30 minutes. After approximately 135minutes the incubator shaking was turned off for 60minutes after which the shaking was resumed until the end of the protocol. The data were analysed to obtain synthesis rate values for ndh, narK and dmsA. The ratio of synthesis rates for narK/ndh and dmsA/ndh were calculated and used to demonstrate the biosensor activity.

5.5.1.2 Standard procedure verification

As oxygen limitation or anaerobic metabolism could have a large effect on results it was necessary to verify that the standard automated characterisation procedure would not cause these conditions to arise. The *fnr* carrying minimal plasmids pACf-ndh and pACf-narK were transformed into MG1655 *E. coli* and assayed as part of an automated inducible characterisation run. Samples were run among the controls as part of the *rhaB* promoter characterisation experiment (see 5.7.3). The biosensors were set up in exactly the same way as other samples in the experiment. Results were analysed as in the previous biosensor experiment (5.5.1.1).

5.6 Automated characterisation testing and optimisation

5.6.1 Semi automated characterisation workflow

This protocol was the only one designed, scripted and run on the second of the CSynBI's two Theonyx platforms as it had been pre-calibrated. Overnight cultures containing minimal plasmids pAC1, pAC4, pAC5, and pAC7, were set up in triplicate in 1 ml of LB media with kanamycin and grown in 15ml tubes overnight at 37°C and 225rpm shaking. The platform took the overnight cultures and diluted them 1 in 60 into fresh LB and M9 media in a Costar 96-well microplate and then outgrown for 90 minutes. After 90 minutes each M9 sample was diluted 30µl into 90µl of fresh media and each LB sample 20µl into 100µl of fresh LB directly into the black Greiner assay plate. The outer wells were filled with water to prevent any edge effect. Media wells were added to the plate and the assay started.

The plate was transferred to the platform's Synergy HT plate reader where the plate was immediately scanned for absorbance at 600nm and GFP fluorescence (sensitivity 79). The plate was incubated in the plate reader at 37°C. Every hour the plate was shaken at 900rpm for one minute and the plate then scanned for fluorescence and absorbance. After the 3 hour measurement the plate was transferred to a pipetting slot and a 5µl sample from each well was transferred to a microplate pre-loaded with 195µl per well sterile water. The assay plate was returned to the plate reader and the assay resumed until the 5 hour time point when a second set of samples were transferred into another fresh microplate preloaded with 195µl per well sterile water. Both of the output plates were kept on ice before being run on the flow cytometer using the AMS as soon as possible after preparation. The flow cytometry settings were the same as had been used for the manual version of this protocol (4.4.1). The results were analysed the same way as before.

5.6.2 Optimisation of automated workflow

5.6.2.1 Testing suitability of absorbance based dilution

All automated work from this point on was designed, scripted, debugged and tested on the characterisation platform using the Synergy HT plate reader. Volume based dilutions had proved to yield inconsistent cell populations at the start of experiments and an alternative strategy using absorbance based dilution was tested. The outer wells were used for samples for the first time. The first 3 runs were carried out at 37°C but later runs were carried out at 30°C to allow the assay to

collect more data before samples entered stationary phase. Samples to be tested along with negative and reference controls were set up in triplicate for overnight growth in 1ml of LB with kanamycin in a 15ml tube. Samples were grown in the tubes at 30 or 37°C overnight at 225rpm. The next day samples were transferred into a Costar microplate (200µl per well) and supplied to the platform. The platform diluted the samples either 1 in 30 or 1 in 60 into fresh 0.4% glucose MOPS. These cultures in MOPS media were scanned for absorbance at 600nm and diluted into a fresh plate to a range of absorbance values (each column diluted to a different absorbance) in 100µl before a 90 minute outgrowth in the microplate incubator (30°C or 37°C at 600-700rpm). 5 minutes before the end of outgrowth a black Greiner (assay) plate was filled with 85µl of fresh media per well except for the final column which was filled with 100µl of media as a control. This assay plate was transferred to the heated slot to keep warm during the second dilution. The outgrowth plate was transferred to the plate reader and scanned for absorbance at 600nm and the samples diluted to a range of absorbance values again.

Following dilution 15µl of each sample was transferred to the assay plate and the assay was started with an initial read of fluorescence and absorbance (sensitivity 79 and 600nm respectively). The plate was assayed for 5 hours with the plate remaining in the incubator at 30 or 37°C and 600-700rpm between measurements in the plate reader which were carried out every 30 minutes. While in the incubator the plate was covered by a Costar plate lid to prevent evaporation and reduce any edge effect. Measurements were carried out with fluorescence before absorbance to minimise plate time in plate reader (see 5.3.4 for details).

5.6.2.2 Standard strain testing of MG1655 and MDS42

While preliminary testing had used Top10 *E. coli* this was considered an unsuitable candidate for characterisation as it was a lab strain with an interesting genotype. More suitable strains needed to be tested for use in the characterisation protocol. MG1655 and MDS42 were selected as they represented two types of *E. coli* likely to be used for characterisation; MG1655 as a 'wildtype' and therefore lacking in mutations *E. coli* and the MDS42 strain which while minimal had a relatively *E. coli* normal phenotype. MG1655 and MDS42 *E. coli* transformed with plasmids SVc, SVc-100, SVc-101, SVc-102, SVc-106 and SVc-113 and SVc-114 were grown overnight at 37°C and 700rpm in LB in a microplate with kanamycin (3 colonies for each picked and placed in separate wells). The following day samples were diluted 1 in 60 into fresh 0.4% glucose MOPS medium and then diluted to 0.045 absorbance units in 100µl. The samples were outgrown for 90 minutes before dilution to 0.08

absorbance units. 15µl of the diluted samples were then transferred to an assay plated set up with 85µl per well pre warmed 0.4% glucose MOPS media in the first 6 columns. Samples in MG1655 were arrayed in the first 3 columns and MDS42 samples in columns 4-6. Column 7 contained 100µl of media per well as a control. The assay was then carried out exactly as described at 30°C and 700rpm as was described in 5.6.2.1 with the exception of measurements being taken every 7.5 minutes between 120 minutes and 240 minutes.

5.6.2.3 Optimisation of dilution for characterisation in MG1655

MG1655 proved to be a good characterisation candidate but the growth remained inconsistent between samples in each run and the population at the beginning of the characterisation was clearly too low. The original dilutions had been chosen using Top10 *E. coli* so an experiment to find dilutions more optimal for MG1655 *E. coli* was designed. MG1655 *E. coli* transformed with SVc and SVc-101 were grown overnight in 200µl per well of LB with Kanamycin in a costar microplate at 30°Cand 700rpm with a breathable film sealing the plate. Samples were diluted, outgrown and assayed in 0.4% glucose MOPS medium at 30°C and 700rpm using the protocol used in the dilution testing (5.6.2.1) but with different dilution values. This time the dilution values were altered following each test run to 'hone in' on optimised values for dilutions which work reliably and do not stray into stationary phase.

Following the MG1655 optimisation the logarithmic fitting script developed with Catherine Ainsworth was used to fit 3 of the key growth parameters (growth rate, carrying capacity and log phase) to the absorbance data. Fitting of the growth model was carried out in R to the modified logistic growth model discussed by Zwietering et al. (1990). The fit model was as follows:

$$y(t) = \frac{A}{1 + exp\left(\frac{4\mu}{A}(\lambda - t) + 2\right)}$$

Where y is the cell population at time t, A is the carrying capacity, µ is the specific growth rate and λ is the lag time. The absorbance data from many different data runs (at least n=8 per data run) was used as an indicator of population size and the remaining parameters were then obtained by fitting to the curve. Resulting distributions were plotted to observe patterns in growth characteristics and used to help determine optimal dilution settings.

5.7 Standard automated characterisation

5.7.1 Standard automated characterisation of constitutive promoter BioParts

The dilution optimisation resulted in a protocol which produced reliable growth for characterisation and the standard workflow was built around this. The first standard automated protocol was designed to allow the characterisation of a library of constitutive promoters and was designed and executed on the characterisation Theonyx platform. The standard characterisation conditions of 30°C and 700rpm shaking with 0.4% glucose MOPS media and MG1655 Busby *E. coli* were used to characterise the Anderson constitutive promoter library available on the Parts Registry (n.d).

A diagram for the workflow can be seen in figure 6.16. The Anderson collection was sequence verified and all 20 members plus two additional members generated during cloning were transformed into MG1655 *E. coli* for characterisation. All samples were taken as single colonies from agar plates and transferred to 200µl per well LB media with kanamycin in a Costar microplate. The plate was kept overnight in the platform incubator at 30°C and 700rpm covered by a breathable film. Included among the samples were cells carrying the SVc plasmid as a negative control and the SVc-101 plasmid as a reference positive control. Prior to starting the automated programme output and dilution plates (for flow cytometry) were placed on the deck pre-filled with the following volumes of water per well: Output 1 - 190µl, Dilution - 195µl and Output 2 - 160µl. The heated block was preheated to 37°C and the incubator and plate reader were pre-heated to 30°C. MOPS media with 0.4% glucose was pre-heated to 37°C and poured into an autoclaved reagent trough and placed in the trough rack (the media was kept in an incubator at 37°C when not being used).

The characterisation began with the transfer of 282µl of MOPS media into wells on a 96-well clear Costar microplate – 1 for each well of sample/control. The sample plate was moved from the incubator to a pipetting slot and each sample mixed (20µl, 2 cycles of aspirate and dispense) prior to 18µl of sample being transferred to the media filled preparation plate. The sample plate was then removed and the preparation plate moved to the plate reader for the first dilution measurement. Following measurement the samples were diluted to 0.0117 absorbance units (100µl) via two-step dilution with fresh MOPS media into new wells on the same plate.

The preparation plate was then transferred to incubator slot 1 where it was covered with a plate lid for outgrowth (90 minutes). 5 minutes before the end of outgrowth the assay plate was transferred to a pipetting slot and media dispensed (85µl) into each well which was to hold sample

during the assay. A column of media (100µl) was also dispensed. Following this the assay plate was transferred back to the incubator and covered with a plate lid. The preparation plate was then transferred to the plate reader and a second absorbance measurement carried out. The resulting values were used to perform a one step dilution to 0.024 absorbance units (100µl) with fresh MOPS media for all samples and controls, with the assay plate then transferred to the heated slot and 15µl per sample transferred from the preparation plate to the assay plate to begin the characterisation procedure.

The assay plate was then transferred to the incubator for a brief shake prior to the initial reading. A fluorescence reading at the standard sensitivity setting of 70 was carried out first followed by an absorbance read at 600nm. The microplate was then returned to the incubator and covered with a plate lid once again. From this point on a regular incubate–reading cycle with 15 minutes between measurements was performed until the 180 minute mark.

After 180 minutes the first sample for flow cytometry was taken. Immediately following the measurement the assay plate was transferred to the heated slot and the first output plate to a pipetting slot. For each microplate well containing a sample, media or control 10µl was transferred to the first output plate. The assay plate was returned to the incubator and the 15 minute incubate-reading cycle resumed from the 195 minute measurement. The output plate was removed and carried on ice to the flow cytometer where it was immediately assayed using the AMS attachment.

After 360 minutes of assay the assay plate was again transferred to the heated position and the dilution for flow cytometry sampling. For each well containing a sample or control 5µl was transferred to the dilution plate to dilute the cells. 40µl of these diluted samples were then mixed and transferred to the second output plate. Finally the output plate was carried on ice to the flow cytometer and immediately assayed using the AMS. For both sets of flow cytometry measurements the settings used (all logarithmic) were as follows: FSC – E01, SSC – 375, FL1 – 700, FL2A – 800 and FL2W – 800. All samples were run through the AMS on low flow rate.

Data produced using the standard workflows were analysed using the same procedure as for manually characterised BioParts with the inclusion of the data calibrations where appropriate.

5.7.2 Standard characterisation of inducible promoter BioParts

As the output from an inducible promoter is related to the concentration of an inducer it was necessary to modify the protocol that worked very successfully for constitutive promoters. It was decided to add the induction step at 120 minutes at this stage the bacterial population had reached a good level for accurate measurement while allowing a large amount of time for data collection. The inducible promoters chosen for testing were *xylF* promoter and *rhaB* promoter which are both induced by sugars and initial testing (see 4.3.4) indicated high concentrations of inducer and thus a large volume of liquid relative to the sample would need to be transferred (e.g. 1.5M xylose in 25µl). To make this possible it was necessary to dilute the inducer in EZ rich MOPS without carbon source to prevent dilution of media components. To compensate for this the media used during this experiment contained 0.5% (W/V) glucose and the final volume after induction would be 125µl. The xylose induction concentrations (all in mM) were: 300, 250, 150, 100, 80, 60, 50, 40, 35, 30, 25, 22, 18, 15, 12, 10, 8, 6, 4, 3, 2, 1, 0.5 and 0mM. The rhamnose induction concentrations (all in mM) were: 300, 250, 150, 100, 75, 45, 30, 18, 12, 7.5, 4.5, 3, 1.8, 1.2, 0.75, 0.45, 0.3, 0.18, 0.12, 0.075, 0.045, 0.03, 0.018 and 0.

Plasmids SVc-X-xF, SVc-X-101 and SVc-X and plasmids SVc-RII-rB, SVc-RII-101 and SVc-RII were transformed into MG1655 for use in the characterisation of *xylF* promoter and *rhaB* promoter respectively. Single colonies were used to inoculate three 1ml LB and kanamycin cultures in eppendorf tubes before 200µl was transferred to 4 wells in a microplate for overnight growth. This was carried out to ensure all 4 wells would contain the same biological sample but would not require replication into extra wells the next morning. The control carrying plasmids would get 4 wells for overnight growth each with a single colony used to inoculate each well. These samples were grown in the microplate overnight in the incubator at 30°C and 700rpm covered by a breathable film.

The characterisation procedure was then carried out as for constitutive promoters (see previous protocol) except the wells from overnight were duplicated to make columns of 8 wells for each biological repeat of the tested BioPart and full columns for both controls, making a 5 column by 8 row grid. The samples were diluted, outgrown and then diluted again as before. The assay plate was prepared with 11 columns of wells containing 85µl of media for samples and the final column containing 100µl of media as an absorbance blank. Samples were transferred to the plate as follows; BioPart sample 1 to columns 1, 4 and 7, BioPart sample 2 to columns 2, 5 and 8, BioPart sample 3 to columns 3, 6 and 9, reference samples to column 10 and negative samples to column 11.

The constitutive programme was again followed until the 120 minute measurement after which all the wells were induced with pre-heated inducer. Columns 1-3 were induced with the highest inducer concentrations, columns 4-6 the middle concentrations and 7-9 the lowest concentrations. The control samples and media were induced with the 3 highest inducer concentrations, the 2 median concentrations and the 3 lowest concentrations. Following induction the plate was transferred back to the incubator until it was time for the 135 minute reading. From this point on the protocol used exactly matched the constitutive programme previously described.

For fluorescent measurements the sensitivity was set to 80. For the *xylF* promoter characterisation the flow cytometry settings matched those used for the constitutive promoter characterisation. For the *rhaB* promoter characterisation the flow cytometry beads were used to adjust the flow cytometer prior to measurement (see 5.4.2 for details). All data from this experiment was analysed using the same procedure as for manually generated data analysis.

5.7.3 Standard automated characterisation of carbon source context

Testing BioParts in media containing different carbon sourced required a few more adjustments to the standard framework. Unlike previous characterisations it was decided to characterise at 37°C and 700rpm shaking as it was considered that the change in carbon source could lead to slower growth. The protocol also required the use of two plates simultaneously which was achieved by effectively running the same programme twice with an offset of approximately 8 minutes. To keep the platform running to time the flow cytometry sampling steps were carried out by hand using a calibrated multi-channel pipette. Batches of MOPS media were created with 7 different carbon sources in addition to MOPS. Stock solutions of 20% (w/v) glycerol, xylose, arabinose, rhamnose, sucrose, mannose and maltose were filter sterilised and added to carbon free MOPS media to a final concentration of 0.4% (7 separate batches). The batches were split in three and immediately frozen to prevent contamination or precipitation.

The work flow was set up to assay 4 constitutive promoter BioParts (excluding controls) carried on plasmids SVc-106, SVc-110, SVc-111 and SVc-115 in MG1655 *E. coli* in media containing one of 8 carbon sources. Triplicate biological repeats for each along with SVc and SVc-101 controls were used to inoculate separate wells in 200µl of LB and kanamycin in a costar microplate. The samples were grown overnight at 37°C and 700rpm in the microplate under a breathable film. The following morning batches of media were thawed, pre-heated to 37°C and transferred to 8 clearly marked troughs for use in the characterisation.

The complexity of this assay and the preparation steps required a different plate layout which was retained in both stages of the protocol and is shown in figure 5.2. The preparation steps roughly followed the constitutive protocol but with 2 costar microplates each carrying 4 different batches of media. Each plate had 18 wells in each 3 column quarter filled with 188µl per well of media (one batch per quarter). 12µl of each sample was transferred 8 times – 4 times to each plate – to start 8 cultures per sample in all 8 batches of media (1 per batch).

The overnight sample plate was removed and the preparation plates were then measured for absorbance and the samples diluted to 0.004 absorbance units (100µl) into a fresh pair of costar microplates. The fresh plates were transferred to the incubator to outgrow for 90 minutes and the 2 assay plates were prepared with the same pattern of media (see figure 5.2) during this time. Following this the outgrowth plates in the incubator were scanned for absorbance and the samples diluted again into fresh plates to 0.016 absorbance units (100µl). 15µl of each sample was then transferred to the respective well on the assay plate.

While the second assay plate was being prepared the first plate was measured for fluorescence using sensitivity 70 and absorbance at 600nm. When the second assay plate was fully set up the plates were incubated in the on platform incubator (covered with a plate lid) for 5 hours with measurements for fluorescence and absorbance taken every 16 minutes. Samples were taken for flow cytometry using the same dilutions and plates as the constitutive and inducible characterisation with the exception of samples being taken at 128 minutes and 256 minutes (rather than 180 and 360) and microplates stored on ice. The samples were run through the flow cytometer using the AMS and the flow cytometry settings used for previous automated characterisations (see 5.7.1 and 5.7.2).

Data was analysed using the same analysis method as for other characterisation data. Growth rates were calculated with the software developed by Hall et al. (2014).

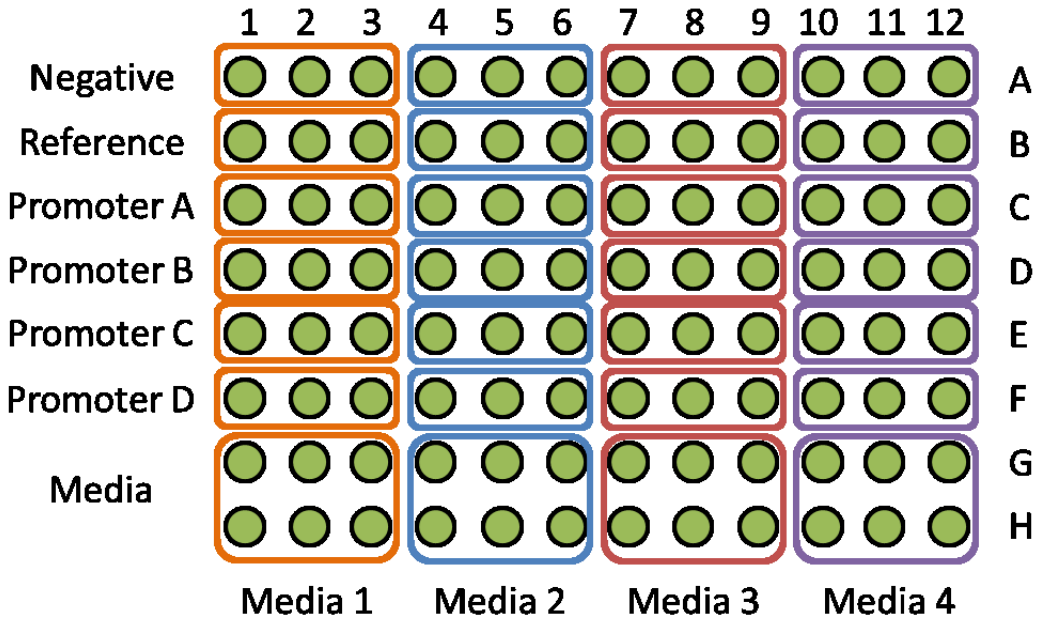


Figure 5.2 Media layout used for context characterisation experiment

For the carbon source context experiment 2 plates were run, each loaded with the same samples in MOPS media with one of 4 carbon sources (8 total). In this experiment promoter A was J23106, promoter B J23110, promoter C J23111 and promoter D J23115.

5.7.4 Automated characterisation of inducible promoter BioParts libraries

For the synthetic promoter libraries a characterisation protocol based around the standard framework for induction was required that could assay many samples at a time. As the same regulator would be used with the synthetic promoters it was considered that the inducer transfer curve would be the same shape with different amplitude (level of output). On this basis it was decided to design the protocol to test a reduced number of induction concentrations which would indicate roughly the maximum and basal output from the promoter. The induction concentrations of 300, 100 and 0mM were chosen (with stocks made in carbon source free MOPS media).

The library workflow was similar to the single inducible promoter BioPart workflow (5.7.3). As such large volumes of inducer were to be used and so 0.5% glucose MOPS was used as the growth medium for assay and incubation conditions of 30°C and 700rpm. Samples for 8 BioParts in MG1655 *E. coli* were set off in biological triplicate from single colonies for overnight growth in 200µl of LB and kanamycin in a costar microplate. These samples were arranged so that each sample would be represented once per column in the first 3 columns of the microplate. Samples carrying SVc-X and SVc-X-101 were included in the fourth column as controls (4 wells each as in 5.7.2). The plate was grown overnight covered with a breathable film in the incubator.

Following overnight growth the inducible BioPart protocol was followed through the dilutions, and outgrowth (with only duplication of the controls to use 8 wells each). Each column of sample replicates was sub-cultured (15µl into the 85µl of media) to 3 columns on the assay plate; the first column of sample replicates were transferred to columns 1, 4 and 7 of the assay plate, the second column of sample replicates were transferred to columns 2, 5 and 8 of the assay plate and the third column of the sample replicates were transferred to columns 3, 6 and 9 of the assay plate. The control samples were transferred (15µl per well) to columns 10 (reference controls) and column 11 (negative controls) of the assay plate. Column 12 contained 100µl per well media.

The characterisation was carried out as for the single BioPart inducible workflow (5.7.3) at 30°C and 700rpm with the plate reader fluorescence sensitivity set to 75. After 120 minutes induction was again carried out with samples in the first three columns induced to a final concentration of 300mM xylose, columns 4-6 induced to 100mM xylose and column 7-9 with 0mM xylose. The controls and media were induced with all three concentrations. The characterisation protocol following this point including flow sampling steps and settings matched the single BioPart inducible workflow (5.7.3). Data analysis was again carried out using the same procedure as for manually generated data analysis.

5.7.5 Automated characterisation of X3 synthetic promoter library

For the characterisation of the X3 xylose inducible library a modified version of the inducible BioPart library workflow was used. These modifications only are detailed here but the majority of this procedure was identical to the previous characterisation (5.7.4). The characterisation was carried out in 0.4% (W/V) glucose MOPS on the characterisation platform at 30°C and 700rpm using MG1655 *E. coli*. Controls used in this experiment were SVc-X and SVc-X-101. The only difference was a change in the volume of cells taken for flow cytometry analysis was increased to give larger flow cytometry sample sizes. After 180 minutes all wells were mixed for 2 cycles of 50µl of aspirate and dispense with an aspirate speed of 35µl/s and a dispense speed of 20µl/s (no offset) immediately before transfer of 35µl of sample to output plate 1. After 360 minutes 5µl from each well was transferred as normal to the dilution plate containing 195µl water then 100µl of each diluted sample was transferred to the output plate which contained 100µl per well of water.

5.7.6 Automated characterisation of the X114aX synthetic promoters

While it had been considered that the shape of the transfer function of the synthetic promoters would match the one observed for the *xylF* promoter as they shared the same regulator protein this had not been proven. To test this it was decided to characterise in detail the transfer function of three of the synthetic promoters based on a single constitutive promoter (5.7.2). The 114a promoter was chosen because of the high on:off ratio and maximum signal for 2 of the three promoters. To characterise these promoters the single inducible promoter BioPart characterisation procedure was repeated with promoters X114aX1, X114aX2 and X114aX3 each replacing one of the biological repeats of the *xylF* promoter. The only modifications to the protocol were the increased flow sampling settings used during characterisation of the X3 synthetic promoter library (5.7.5). The inducer concentrations used were the same as in the *xylF* promoter characterisation. The characterisation was carried out in 0.4% (W/V) glucose MOPS on the characterisation platform at 30°C and 700rpm using MG1655 *E. coli*. Controls used in this experiment were SVc-X and SVc-X-101.

6 Results

6.1 Standard plasmids for characterisation

6.1.1 3 and 4 module minimal plasmids

A vector was required to host the BioParts for characterisation. While the pSB series of plasmids produced by the parts registry were widely distributed and used they were very inflexible outside of the multiple cloning site. As such if a pSB vector was chosen the components of the backbone would be difficult to replace without moving to a new plasmid. A vector where various modules could be swapped in and out and new modules added was desirable to allow modifications in future to add new functional modules or test different types of BioPart.

A new vector topology had been designed by Eichenberger et al. (2009) at Imperial to remove potentially unnecessary parts of plasmids and thus effectively minimise them. This plasmid system was chosen as it was designed to allow generation of plasmids from separate modules by using recombination and homology based DNA assembly techniques such as CPEC, In-Fusion and Gibson cloning. As a result the modules in the plasmid could be quickly and easily swapped. The minimal plasmids consisted of 3 modules separated by unique homology regions (labelled 1-3 in figure 6.1 A). The homology regions separated a replication origin, a resistance cassette and a multiple cloning site. The kanamycin resistance cassette was chosen as kanamycin is a particularly selective bactericidal antibiotic which is broken down slowly and so ideal for use. The p15a origin was chosen as it is a low to medium copy origin and should impart little or no burden on the chassis while avoiding the difficulties associated with working with very low copy origins such as pSC101. For the characterisation module the BioBrick MCS containing the I13504 BioBrick was chosen. This BioBrick contained the B0034 high output RBS, the gene encoding the GFPmut-3b protein and the B0015 terminator. This was chosen for its similarity to constructs used as part of the work of Kelly et al. (2009) but with a stronger RBS which would increase the level of signal for detection.

Promoters can be ligated into the BioBrick prefix retained as part of BioBrick I13504. Given the small size of constitutive promoters (generally less than 40bp) this was considered easiest to achieve by annealing oligos and ligating them into the vector. This method was used to generate a reference construct for RPU output characterisation in accordance with Kelly et al.(2009) by ligation of the J23101 promoter in front of the aforementioned BioBrick to generate plasmid pAC4. For the BioFAB bit to atom to bit experiment (see 6.2.1) the *tac* and *lacUV5* promoters from the BioFAB pilot programme were chosen and synthesised as oligos which were annealed together and ligated into the pAC1 plasmid generating the testing constructs (pAC5-7).

A 4 module plasmid was later designed to allow testing of sensors for aerobic metabolism. For these sensors to work it was necessary to over express the FNR protein which was responsible for the regulation of these promoter biosensors. A new homology sequence was designed which would allow a cassette the *fnr* gene to be inserted after the GFP sensor module (figure 6.1 B). To avoid having to design an *fnr* expression module which may or may not be functional the genomic locus containing the *fnr* gene plus the sequence upstream likely to contain the promoter was amplified by PCR and inserted into the pAC1 plasmid by CPEC. This plasmid was designated pAC1f. The FNR protein repressed *ndh* promoter and FNR protein activated *dmsA* and *narK* promoters were cut from their holding plasmids and ligated into pACf to make the Biosensor plasmids pACf-ndh, pACf-dmsA and pACf-narK respectively. A list of the minimal plasmids used and the key DNA sequences are included in tables 6.1 and 6.2.

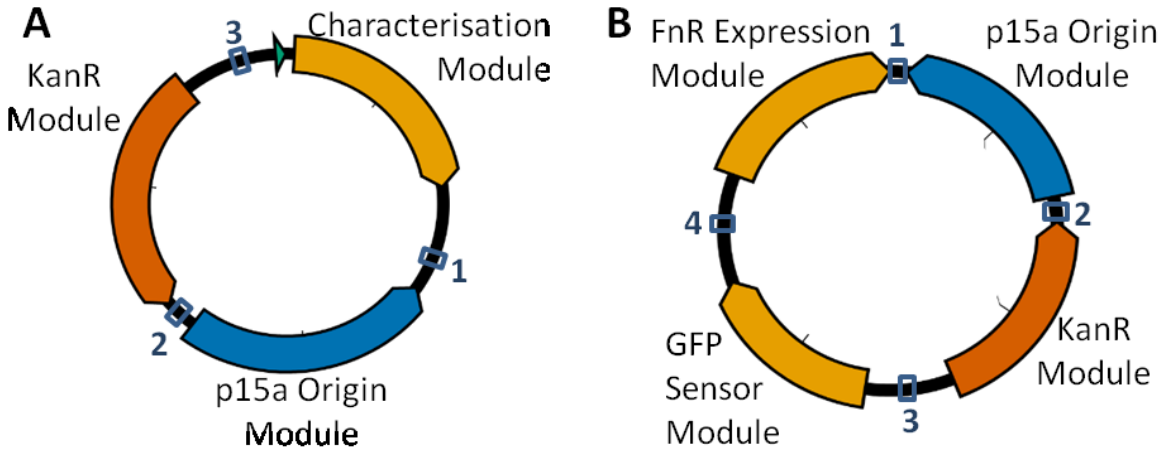


Figure 6.1 Topology and features of the minimal plasmids used in early experiments.

The minimal plasmids were used for early trial characterisation experiments and to develop FnR based biosensors. Both plasmids were put together with a combination of CPEC and/or InFusion using the homologous sequences (numbered boxes). The 3 module plasmid displayed in A was designed to allow testing of a promoter cloned into the characterisation module using the Biobricks cloning site inside. B shows an extended 4 module minimal plasmid designed to over-express the Fnr transcriptional regulator in addition to containing a module with GFP under expression of an Fnr regulated promoter (positively or negatively). These plasmids were put together to produce paired sensors which could attempt to form an *in vivo* aerobic/anaerobic metabolism sensor.

Plasmid	Modules	Promoter
pAC1	p15a origin (minimal), KanR (minimal), I13504	N/A
pAC4	p15a origin (minimal), KanR (minimal), I13504	J23101
pAC5	p15a origin (minimal), KanR (minimal), I13504	pTac
pAC7	p15a origin (minimal), KanR (minimal), I13504	pLacUV5
pACf	p15a origin (minimal), KanR (minimal), I13504, FnR	N/A
pACf-ndh	p15a origin (minimal), KanR (minimal), I13504, FnR	Ndh
pACf-narK	p15a origin (minimal), KanR (minimal), I13504, FnR	narK
pACf-dmsA	p15a origin (minimal), KanR (minimal), I13504, FnR	dmsA

Table 6.1 Minimal plasmids used in this research.

Module/Element	**Sequence**
Minimal KanR (Reverse orientation)	**GTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAA TGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATG AAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCG ACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAA TCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGT TCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTG ATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGA ATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTC TAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTAC GGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCA TCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCC CATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAA ATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCAT AACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTT GTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTGTTGAATAAATCGAACTTTTGCTG AGTTGAAGGATCAGATCACGCATCTTCCCGACAACGCAGACCGTTCCGTGGCAA** KanR protein translated sequence: **MSHIQRETSCSRPRLNSNMDADLYGYKWARDNVGQSGATIYRLYGKPDAPELFLKHGKGSVANDVT DEMVRLNWLTEFMPLPTIKHFIRTPDDAWLLTTAIPGKTAFQVLEEYPDSGENIVDALAVFLRRLHSIP VCNCPFNSDRVFRLAQAQSRMNNGLVDASDFDDERNGWPVEQVWKEMHKLLPFSPDSVVTHGDF SLDNLIFDEGKLIGCIDVGRVGIADRYQDLAILWNCLGEFSPSLQKRLFQKYGIDNPDMNKLQFHLMLD EFF***
P15a (Minimal)*	**GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGCTAGCGGAGTGTATACTGGCTT ACTATGTTGGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCA CCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGC TCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATG CCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCC CCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAA**

	AGATACCAGGCGTTTCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGT GTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTT CGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAAC TATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTG ATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGT GACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAA ACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCA AGAAGATCATCTTATTAAGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGG
GFP output in Biobrick MCS (no promoter)*	**CATTAACCTATAAAAATAGGCGTATCACGAGGCAGAATTTCAGATAAAAAAAATCCTTAGC** **BioBrick Prefix** **RBS B0034** **TTTCGCTAAGGATGATTTCTGGAATTCGCGGCCGCTTCTAGAGAAAGAGGAGAAATACTAGATGC GTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATG GGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAA TTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGGTG TTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCG AAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGA AGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAG ATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCA GACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCG TTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACA ACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCC TTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAATAATACT AGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTG TTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTA** **BioBrick Suffix** **TATACTAGTAGCGGCCGCTGCAGTCCGGCAAAAAAACGGGCAAGGTGTCACCACCCTGCCCTTTTT CTTTAAAACCGAAAAGATTACTTCGCGTTATGCAGGCTTCCTCGCTCACTGACTCGCTGCGCTCGGT CG** GFPmut-3b protein translated sequence: MRKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFGYGV QCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI LGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLST QSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK**
fnr cassette	**ACATTAAACAATTTGTGCCAGCTTGTTCACACTTTTATGTAAAGTTACCCTTAACAACTTAAGGGTT TTCAAATAGATAGACATATATTTACATCTAATATCGGAATTCTCTGCTGTTAAGGTTT** **-35** **-10** **+1** **GCTTAGACTTACTTGCTCCCTAAAAAGATGTTAAAATTGACAAATATCAATTACGGCTTGAGCAGA CCTATGATCCCGGAAAAGCGAATTATACGGCGCATTCAGTCTGGCGGTTGTGCTATCCATTGCCAG GATTGCAGCATCAGCCAGCTTTGCATCCCGTTCACACTCAACGAACATGAGCTTGATCAGCTTGAT AATATCATTGAGCGGAAGAAGCCTATTCAGAAAGGCCAGACGCTGTTTAAGGCTGGTGATGAACT TAAATCGCTTTATGCCATCCGCTCCGGTACGATTAAAAGTTATACCATCACTGAGCAAGGCGACGA GCAAATCACTGGTTTCCATTTAGCAGGCGACCTGGTGGGATTTGACGCCATCGGCAGCGGCCATC ACCCGAGCTTCGCGCAGGCGCTGGAAACCTCGATGGTATGTGAAATCCCGTTCGAAACGCTGGAC**

	GATTTGTCCGGTAAAATGCCGAATCTGCGTCAGCAGATGATGCGTCTGATGAGCGGTGAAATCAA AGGCGATCAGGACATGATCCTGCTGTTGTCGAAGAAAAATGCCGAGGAACGTCTGGCTGCATTCA TCTACAACCTGTCCCGTCGTTTTGCCCAACGCGGCTTCTCCCCTCGTGAATTCCGCCTGACGATGAC TCGTGGCGATATCGGTAACTATCTGGGCCTGACGGTAGAAACCATCAGCCGTCTGCTGGGTCGCTT CCAGAAAAGCGGCATGCTGGCAGTCAAAGGTAAATACATCACCATCGAAAATAACGATGCGCTG GCCCAGCTTGCTGGTCATACGCGTAACGTTGCCT FNR protein translated sequence: **MIPEKRIIRRIQSGGCAIHCQDCSISQLCIPFTLNEHELDQLDNIIERKKPIQKGQTLFKAGDELKSLYAIRS GTIKSYTITEQGDEQITGFHLAGDLVGFDAIGSGHHPSFAQALETSMVCEIPFETLDDLSGKMPNLRQQ MMRLMSGEIKGDQDMILLLSKKNAEERLAAFIYNLSRRFAQRGFSPREFRLTMTRGDIGNYLGLTVETI SRLLGRFQKSGMLAVKGKYITIENNDALAQLAGHTRNVA***
J23101 promoter	**-35 -10** **TTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGC**
tac promoter	**-35 -10 +1** **TTGACAATTAATCATCCGGCTCGTATAATGTGTGGAATTGTGAG**
lacUV5 Promoter	**-35 -10 +1** **CCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAG**
ndh promoter	**FNR operator** **CTCTGTTTTTTGATCTCACCCGGTAAAGTCGCCTATCTTTTCAGCAACAAAAC** **FNR operator -35 -10 +1** **TTGATTAACATCAATTTTTGGTATGACCAATGCACCATTCATGTTATTCTCAATAGCGA**
narK promoter	**FNR operator** **ATTTTTATGAAGTCACTGTACTCACTATGGGTAATGATAAATATCAATGATAGATAAAGTTA** **FNR operator -35 -10 +1** **TCTTATCGTTTGATTTACATCAAATTGCCTTTAGCTACAGACACTAAGGTGGCAGACA**
dmsA promoter	**TACCCAATTTTTCTGAATCTAAAAAGCGCCTGCGGGGCTTTTTTTGTCTCCCT** **FNR operator -35 -10** **TTGATACCGAACAATAATTACTCCTCACTTACACGTAATACTACTTTC**
Homology Sequence 1*	**GACGAGACGATTTAAATGG**
Homology Sequence 2*	**CATTTAAATCGGTTTAAACG**
Homology Sequence 3*	**GCGTTTAAACCGGCTAGCTA**
Homology Sequence 4	**GATTCATAAGATGGAGCGCG**

Table 6.2 Full sequences of minimal plasmid modules

*denotes sequences designed/first used in this format by Eichenberger et al. (2009). For each module protein coding regions are displayed in **red** letters and the amino acid sequence is shown following the DNA sequence. Other annotations (where known) include the -35 box, -10 boxes and transcription start site (+1) of promoters (underlined), operator sequences in **green**, ribosome binding sites in **orange** and BioBrick prefix and suffix in **blue**. The transcription start site for the *dmsA* and J23101 promoters is the A base immediately following the promoter sequence.

6.1.2 SV Series plasmids

The modular minimal plasmids unfortunately proved to be unstable and prone to read through which was a concern for effective characterisation (Casini, 2010). Additionally a new module that would not suffer from read through was required to allow characterisation of inducible promoters in combination with a regulatory protein. On this basis it was decided that a new plasmid should be designed to eliminate these problems.

The new plasmid was designed with 6 homology regions; 4 to join together the 4 modules required for the plasmid and a further two to facilitate the swapping of genes within the regulatory module. The homology sequences were designed by requesting 20bp neutral sequences from a beta version of the R2O software(Casini et al., 2014) with settings altered to prevent inclusion of ribosome binding sites and common restriction sites. One half of the plasmid was designed to again carry the full KanR resistance cassette and the p15a origin orientated to face each other. The other half of the plasmid was designed to include the GFP reporter BioPart characterisation (testing) module and the regulator module orientated away from each other (figure 6.2 A). This design was chosen so as to reduce the risk of interaction between all the full size modules as they contain full transcription terminators and the orientation should reduce the effect of any residual read through.

The testing module was again based around GFP-mut3b and the B0034 RBS however the suffix that previously followed this BioBrick was removed resulting in the topology documented in SBOLv in figure 6.2 B. The regulator module contained the J23101 promoter flanked by neutral sequences for homology based cloning followed by a ribosome binding site and open reading frame combination (B0034 and mRFP1 (E1010) in the neutral plasmid) before a third neutral sequence and the B0014 terminator – see figure 6.2 B for a diagram in SBOLv annotation. This arrangement allowed the insertion of new regulator genes and ribosome binding sites by homology based assembly methods resulting in the production of semi-standardised regulator expression cassettes. When the RFP gene was swapped for a regulator a new RBS was designed using the RBS calculator(Salis, 2011) with a strength of 25000 units taking into account the sequence of neutral sequence 2 and the new regulator.

The basic plasmid was designated SVa. As the promoters for the testing module and regulator module were so close together the insulator sequence produce used by Davis et al.(2011) was inserted by round the plasmid PCR upstream of the BioBrick prefix to reduce the risk of undesirable interactions between these modules. The new plasmid design was designated SVb. A final upgrade was applied to the SVb plasmid to allow the swapping of modules using restriction cloning in order

to reduce the likelihood of mutations as a result of PCR steps in assembly and to achieve a degree of compliance with the pSEVA vector standard (Silva-Rocha et al., 2013). This was achieved by the introduction of restriction sites adjacent to the 4 module linking neutral sequences – see figure 6.2A. The finished base plasmid was designated SVc. A proposed naming system for the plasmids is explained in figure 6.3.

New ribosome binding sites were designed for the *rhaS* and *xylR* genes as part of the regulator cassettes. The ribosome binding sites and their respective genes were inserted into SVc using CPEC following PCR amplification of the genes from the genome resulting in plasmids SVc-S and SVc-X. A second version of SVc-S was designed with a new 25000 unit RBS design and designated SVc-SII. The *rhaB* promoter was cloned into SVc from the *E. coli* genome before removal of the cap binding sites and transfer into plasmid SVc-S to make SSVc-S-rB. The *xylF* promoter (from BioBrick I741018) was transferred into SVc before removal of CAP binding sites and finally transferred into SVc-X to make plasmid SVc-X-xF. Anderson library promoters were synthesised as oligos then annealed and phosphorylated prior to ligation into SVc. During this process 2 new promoters were generated and named 114a and 114b. The synthetic promoter libraries were generated by round the plasmid PCR and re-ligation with primers encoding operators for the XylR protein.

Key sequences for the SVc plasmids and promoters characterised in SVc plasmids (excluding synthetic xylose inducible promoters) are shown in tables 6.3 and 6.4. The full sequence for the family backbones SVc, SVc-X and SVc-S25II can be found in appendix A.

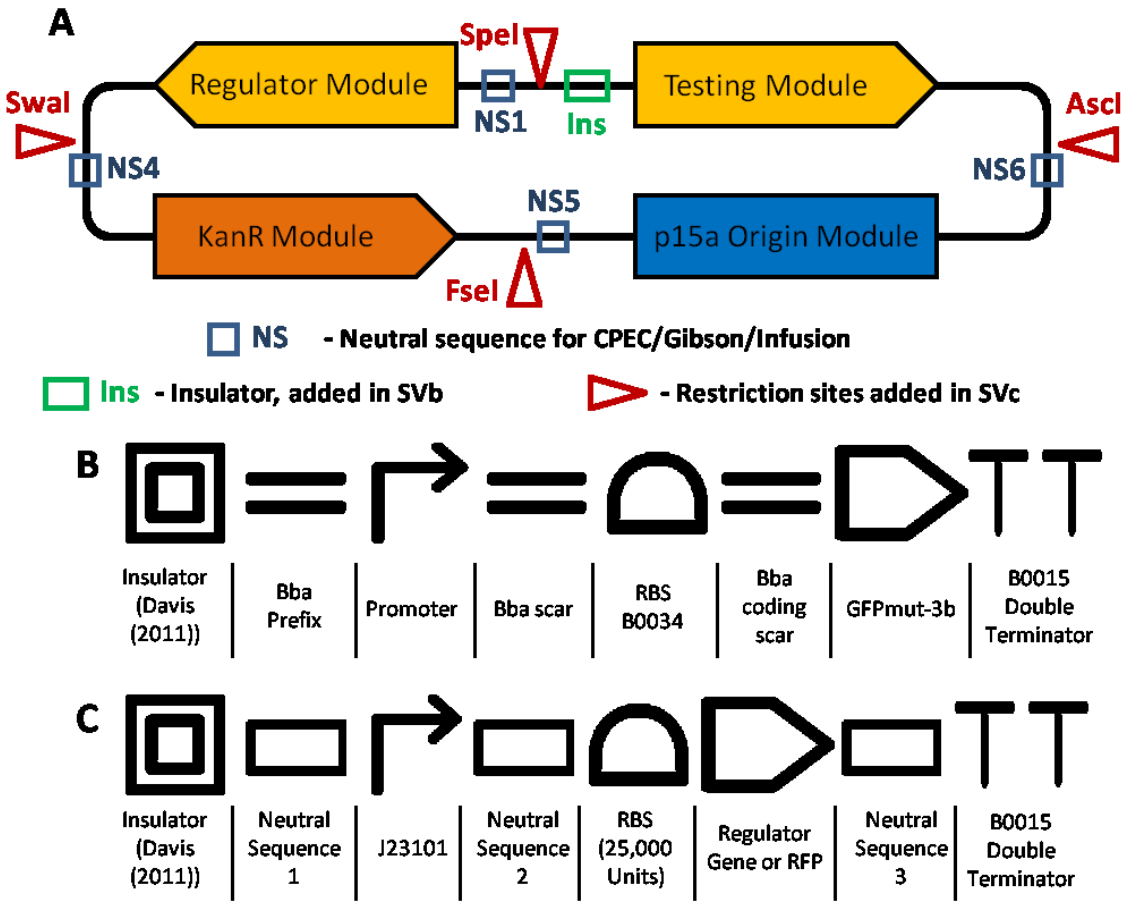
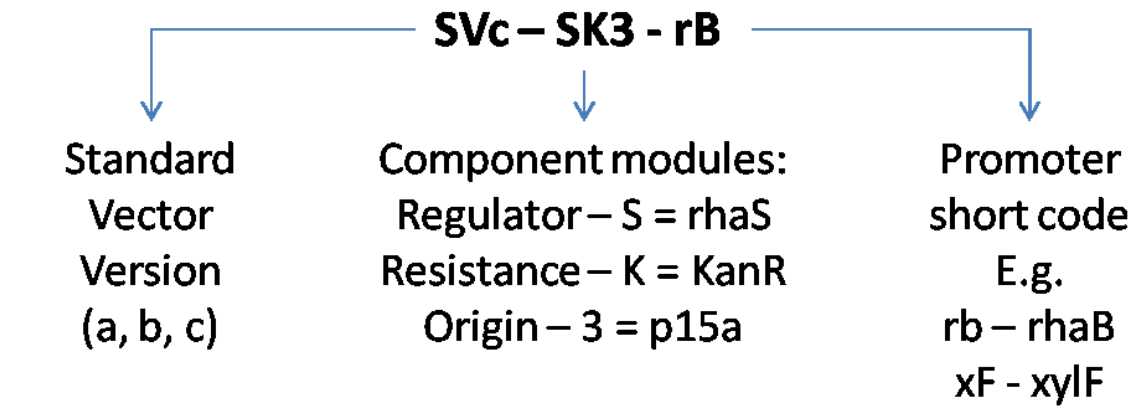


Figure 6.2 Topology and features of the SV series plasmids used for automated characterisation.

The SV series of plasmids were designed for efficient characterisation of individual promoters or promoters with over-expressed regulator. Diagram A is an overview of the topology of the SV series of plasmids, highlighting various features of the plasmid. The SV plasmids are made of 4 modules, the full KanR resistance cassette, the full p15a origin, a standardised RFP or regulator expression module and a GFP output module with the Biobrick BBa prefix retained for cloning. Also highlighted are 4 of the 6 neutral sequences which were used to put the plasmids together with CPEC. The green box indicates where an insulator sequence was added upstream of the promoter in SVb and the red arrows indicate the restriction sites added to make plasmid SVc. SVc plasmids were used in all automated characterisations). Diagrams B and C are detailed SBOL-v annotation maps of the expression/characterisation module (based on BBa_I13504 with a promoter insulator upstream of the prefix) and the regulator expression module where a regulator gene can be inserted (using neutral sequences 2 and 3) to be expressed by promoter J23101 (a unique RBS should be designed using the RBS calculator(Salis, 2011)) with the respective coding region and an output of 25,000 units.



Regulator	Code	Resistance	Code	Origin	Code
RhaS	S	KanR	K	F'	1
RhaS-25II	SII	AmpR	A	SC101	2
XylR	X	TetR	T	p15a	3
AraC	A	CamR	C	ColE1	4
LacI	La	SpecR	S	pUC	5
LuxR	Lu			R6K	6

Figure 6.3 Nomenclature for the SV series of plasmids.

The SV series of plasmids were designed to allow characterisation of many BioParts with many different regulators and possibly with different vector. While only the core SVc based KanR and p15a plasmids were built others were designed and a nomenclature devised that could cope with them. The name for a plasmid is split in three pieces separated by dashes; the plasmid design standard, the modules carried and the promoter carried for testing. The central part of the name was based on 3 characters to define the modules in the vector in the order of regulator, resistance and origin. An example and the letters to indicate other modules are shown in the table. Where the standard p15a origin and KanR resistance are used the name is often shortened to version-regulator-promoter e.g. SVc-X-xF.

Module/Element	Sequence
KanR Full	AAGATAAAAATATATCATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAA GGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTCCAGGCCGCGATTAAATTCCA ACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTG CGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCA AAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAA TTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCAC CACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGA AAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATT GTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGG TTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTG GAAGGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTC TCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGA GTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTT CTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAA ATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAAC ACTGGCAGAGCATTACGCTGACTTGACGGGACTCGAGGGCCGGCC KanR protein translated sequence: MSHIQRETSCSRPRLNSNMDADLYGYKWARDNVGQSGATIYRLYGKPDAPELFLKHGKGSV ANDVTDEMVRLNWLTEFMPLPTIKHFIRTPDDAWLLTTAIPGKTAFQVLEEYPDSGENIVDAL AVFLRRLHSIPVCNCPFNSDRVFRLAQAQSRMNNGLVDASDFDDERNGWPVEQVWKEMH KLLPFSPDSVVTHGDFSLDNLIFDEGKLIGCIDVGRVGIADRYQDLAILWNCLGEFSPSLQKRLF QKYGIDNPDMNKLQFHLMLDEFF*
p15a origin full	ACGCATCTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAATCACCAAC TGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTATGGCTGGATGATG GGGCGATTCGGGCCTGGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGCT AGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGTGCTTCA TGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATA TTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATG GCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGT GAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAA ATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT CCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCG CTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCT CCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAA CTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGG TAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGA CAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCA GAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTCTCAGAGCAAGAGATTAC GCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAAGGGGTCTGACGCTCAGTGGAA CGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGAT CCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTC

RFP/Blank regulator module.	**-35 J23101 Promoter -10 +1** **TTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGCACGGCAGGTATTCGGCTCCATACTAG** **RBS B0034** **AGAAAGAGGAGAAATACTAGATGGCTTCCTCCGAAGACGTTATCAAAGAGTTCATGCGTT** **TCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGT** **GAAGGTCGTCCGTACGAAGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCC** **GCTGCCGTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTT** **AAACACCCGGCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGG** **GAACGTGTTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGGACTCCTCCCTGC** **AAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGGTCC** **GGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTACCCGGAAG** **ACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACGGTGGTCACTAC** **GACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAGCTGCCGGGTGCTTAC** **AAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGAACAG** **TACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAATAAGGACTACGGGCTGAACCT** **CTCCACCGGTGCTTAATAAGGACTACGGGCTGAACCTCTCCACCGGTGCTTAATAAGGACT** **ACGGGCTGAACCTCTTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATATACTAG** **AGAGAGAATATAAAAAGCCAGATTATTAATCCGGCTTTTTTATTATTTATTTAAATCGCGA** **AAT** mRFP protein translated sequence: **MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS** **PQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKL** **RGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEIKMRLKLKDGGHYDAEVKTTYMA** **KKPVQLPGAYKTDIKLDITSHNEDYTIVEQYERAEGRHSTGA****
GFP output (no promoter)	**ACTAGTCACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTTGCT** **BioBrick Prefix RBS B0034** **GGATAACGAATTCGCGGCCGCTTCTAGAGAAAGAGGAGAAATACTAGATGCGTAAAGGA** **GAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGC** **ACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTA** **AATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGT** **TATGGTGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGA** **GTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACT** **ACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAA** **AAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATA** **ACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCA** **AAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATA** **CTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCC** **CTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCT** **GCTGGGATTACACATGGCATGGATGAACTATACAAATAATAATACTAGAGCCAGGCATCA** **AATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTG** **AACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGGC** **GCGCC**

	GFPmut-3b protein translated sequence: **MRKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT FGYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL KGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTP IGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK****
***xylF* gene and RBS**	**25000 unit RBS** **CGTAACTCATCCAATATCCCCACCTATGTTTACTAAACGTCACCGCATCACATTACTGTTCA ATGCCAATAAAGCCTATGACCGGCAGGTAGTAGAAGGCGTAGGGGAATATTTACAGGCG TCACAATCGGAATGGGATATTTTCATTGAAGAAGATTTCCGCGCCCGCATTGATAAAATCA AGGACTGGTTAGGAGATGGCGTCATTGCCGACTTCGACGACAAACAGATCGAGCAAGCG CTGGCTGATGTCGACGTCCCCATTGTTGGGGTTGGCGGCTCGTATCACCTTGCAGAAAGTT ACCCACCCGTTCATTACATTGCCACCGATAACTATGCGCTGGTTGAAAGCGCATTTTTGCAT TTAAAAGAGAAAGGCGTTAACCGCTTTGCTTTTTATGGTCTTCCGGAATCAAGCGGCAAA CGTTGGGCCACTGAGCGCGAATATGCATTTCGTCAGCTTGTCGCCGAAGAAAAGTATCGC GGAGTGGTTTATCAGGGGTTAGAAACCGCGCCAGAGAACTGGCAACACGCGCAAAATCG GCTGGCAGACTGGCTACAAACGCTACCACCGCAAACCGGGATTATTGCCGTTACTGACGC CCGAGCGCGGCATATTCTGCAAGTATGTGAACATCTACATATTCCCGTACCGGAAAAATTA TGCGTGATTGGCATCGATAACGAAGAACTGACCCGCTATCTGTCGCGTGTCGCCCTTTCTT CGGTCGCTCAGGGCGCGCGGCAAATGGGCTATCAGGCGGCAAAACTGTTGCATCGATTAT TAGATAAAGAAGAAATGCCGCTACAGCGAATTTTGGTCCCACCAGTTCGCGTCATTGAAC GGCGCTCAACAGATTATCGCTCGCTGACCGATCCCGCCGTTATTCAGGCCATGCATTACAT TCGTAATCACGCCTGTAAAGGGATTAAAGTGGATCAGGTACTGGATGCGGTCGGGATCTC GCGCTCCAATCTTGAGAAGCGTTTTAAAGAAGAGGTGGGTGAAACCATCCATGCCATGAT TCATGCCGAGAAGCTGGAGAAAGCGCGCAGTCTGCTGATTTCAACCACCTTGTCGATCAA TGAGATATCGCAAATGTGCGGTTATCCATCGCTGCAATATTTCTACTCTGTTTTTAAAAAA GCATATGACACGACGCCAAAAGAGTATCGCGATGTAAATAGCGAGGTCATGTTGtaataa** Translated XylF protein sequence: **MFTKRHRITLLFNANKAYDRQVVEGVGEYLQASQSEWDIFIEEDFRARIDKIKDWLGDGVIA DFDDKQIEQALADVDVPIVGVGGSYHLAESYPPVHYIATDNYALVESAFLHLKEKGVNRFAFY GLPESSGKRWATEREYAFRQLVAEEKYRGVVYQGLETAPENWQHAQNRLADWLQTLPPQT GIIAVTDARARHILQVCEHLHIPVPEKLCVIGIDNEELTRYLSRVALSSVAQGARQMGYQAAKL LHRLLDKEEMPLQRILVPPVRVIERRSTDYRSLTDPAVIQAMHYIRNHACKGIKVDQVLDAVG ISRSNLEKRFKEEVGETIHAMIHAEKLEKARSLLISTTLSINEISQMCGYPSLQYFYSVFKKAYDTT PKEYRDVNSEVML****
***rhaS* gene and original RBS**	**25000 unit RBS (later replaced by the 25II RBS)** **TACCCCGGTAAAAAAGATTCACCAAATGACCGTATTACATAGTGTGGATTTTTTTCCG TCTGGTAACGCGTCCGTGGCGATAGAACCCCGGCTCCCGCAGGCGGATTTTCCTGAACATC ATCATGATTTTCATGAAATTGTGATTGTCGAACATGGCACGGGTATTCATGTGTTTAATGG GCAGCCCTATACCATCACCGGTGGCACGGTCTGTTTCGTACGCGATCATGATCGGCATCTG TATGAACATACCGATAATCTGTGTCTGACCAATGTGCTGTATCGCTCGCCGGATCGATTTC AGTTTCTCGCCGGGCTGAATCAGTTGCTGCCACAAGAGCTGGATGGGCAGTATCCGTCTC ACTGGCGCGTTAACCACAGCGTATTGCAGCAGGTGCGACAGCTGGTTGCACAGATGGAA CAGCAGGAAGGGGAAAATGATTTACCCTCGACCGCCAGTCGCGAGATCTTGTTTATGCAA**

	TTACTGCTCTTGCTGCGTAAAAGCAGTTTGCAGGAGAACCTGGAAAACAGCGCATCACGT CTCAACTTGCTTCTGGCCTGGCTGGAGGACCATTTTGCCGATGAGGTGAATTGGGATGCC GTGGCGGATCAATTTTCTCTTTCACTGCGTACGCTACATCGGCAGCTTAAGCAGCAAACGG GACTGACGCCTCAGCGATACCTGAACCGCCTGCGACTGATGAAAGCCCGACATCTGCTAC GCCACAGCGAGGCCAGCGTTACTGACATCGCCTATCGCTGTGGATTCAGCGACAGTAACC ACTTTTCGACGCTTTTTCGCCGAGAGTTTAACTGGTCACCGCGTGATATTCGCCAGGGACG GGATGGCTTTCTGCAATAAtaaTACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGT CGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCAC ACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGGCGCGCC RhaS protein translated sequence: MTVLHSVDFFPSGNASVAIEPRLPQADFPEHHHDFHEIVIVEHGTGIHVFNGQPYTITGGTVC FVRDHDRHLYEHTDNLCLTNVLYRSPDRFQFLAGLNQLLPQELDGQYPSHWRVNHSVLQQV RQLVAQMEQQEGENDLPSTASREILFMQLLLLLRKSSLQENLENSASRLNLLLAWLEDHFADE VNWDAVADQFSLSLRTLHRQLKQQTGLTPQRYLNRLRLMKARHLLRHSEASVTDIAYRCGFS DSNHFSTLFRREFNWSPRDIRQGRDGFLQ**
RhaS RBS 25II	GGGCCAAAGCAGAGGGGGGGCTACT
NS1	AGACCTTTACGCCGCTGGAG
NS2	TGGAGCCGAATACCTGCCGT
NS3	AGAGGTTCAGCCCGTAGTCC
NS4	GGTAACTCCCGCTGTAGACG
NS5	TGACGCAGCCGAGTTCCGAT
NS6	GGTAGCGTTCAGACTCCTCG
X1 upstream sequence	XylR operator XylR operator AACCAAAAATCGTAATCGAAAGATAAAAATCTGTAA
X2 upstream sequence	XylR operator XylR operator TGTGAATTATCTCAATAGAAAGATAAAAATCTGTAA
X3 upstream sequence	XylR operator XylR operator TGTGAATTATCTCAATAGCAGTGTGAAATAACATAA
Promoter upstream insulator	CACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTTGCTGGATAA C

Table 6.3 Full sequences (excluding promoters) contained within the SVc plasmids used in this research

The SVc plasmid sequences are fully documented to ensure the exact constructs being used are known in the event of unexpected behaviour is observed. NS is short for Neutral Sequence. These sequences have been verified in all backbone plasmids by sequencing. For each module protein coding regions are displayed in **red** letters and the amino acid sequence is shown following the DNA sequence. Other annotations (where known) include the -35 box, -10 boxes and transcription start site (+1) of promoters (underlined), operator sequences in **green**, ribosome binding sites in **orange** and BioBrick prefix and suffix in **blue**. The full sequences for the SVc, SVc-X and SVc-S25II backbone plasmids can be found in Appendix A. The promoter insulation sequence was taken from Davis et al. (2011).

Promoter	Sequence
J23100	-35 -10 ttgacggctagctcagtcctaggtacagtgctagc
J23101	tttacagctagctcagtcctaggtattatgctagc
J23102	ttgacagctagctcagtcctaggtactgtgctagc
J23103	ctgatagctagctcagtcctagggattatgctagc
J23104	ttgacagctagctcagtcctaggtattgtgctagc
J23105	tttacggctagctcagtcctaggtactatgctagc
J23106	tttacggctagctcagtcctaggtatagtgctagc
J23107	tttacggctagctcagccctaggtattatgctagc
J23108	ctgacagctagctcagtcctaggtataatgctagc
J23109	tttacagctagctcagtcctagggactgtgctagc
J23110	tttacggctagctcagtcctaggtacaatgctagc
J23111	ttgacggctagctcagtcctaggtatagtgctagc
J23112	ctgatagctagctcagtcctagggattatgctagc
J23113	ctgatggctagctcagtcctagggattatgctagc
J23114	tttatggctagctcagtcctaggtacaatgctagc
J23115	tttatagctagctcagcccttggtacaatgctagc
J23116	ttgacagctagctcagtcctagggactatgctagc
J23117	ttgacagctagctcagtcctagggattgtgctagc
J23118	ttgacggctagctcagtcctaggtattgtgctagc
J23119	ttgacagctagctcagtcctaggtataatgctagc
J23114a	tttatggctagctcagtccttggtacaatgctagc
J23114b	tttatggctagctcagtcctaggtataatgctagc
Promoter xF (xylF without CAP sites)	XylR operator XylR operator GAAATAAACCAAAAATCGTAATCGAAAGATAAAAATCTGTAATTGTTTTCCCCTGT TTAGTTGCTAAAAATTGGTTA
Promoter rB (rhaB without CAP sites)	RhaS operator RhaS operator -35 ATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAAT -10 +1 TCAGGCGCTTTTTAGACTGGTCGTA

Table 6.4 The Anderson and non-synthetic inducible promoters characterised in this research

Please note, the sequences for the synthetic xylose inducible promoters are built from Anderson sequences and the xyl upstream sequences in table 6.3 but can otherwise be found in figure 6.29. The -35 box, -10 boxes and transcription start site (+1) of promoters are underlined (where known) and operator sequences are highlighted in **green**. The transcription start of the Anderson promoters (J23XXX) are the A base immediately following the end of the shown sequence.

6.2 Manual characterisation and testing ideal growth environment

6.2.1 BioFAB Bit to Atom to Bit (Manual Characterisation)

An initial experiment (termed the Bit to Atom to Bit experiment) was set up in collaboration with the BioFAB based in the USA to investigate the storage and transfer of BioPart characterisation data. The effort was termed Bit to Atom to Bit as this was the core of the objective; taking data stored in bits (i.e. data on a computer) from the BioFAB Data Access Client, reproducing and characterising the physical BioPart DNA (Atoms) at the CSynBI, then returning the resulting data to BioFAB electronically (i.e. in Bits). While methods had been developing for a few years to improve characterisation techniques the documentation of experiments had been neglected outside the work of Canton et al. (2008) where datasheet had been generated to transfer some of the information pertaining to a BioPart characterisation experiment. This experiment was taken as an opportunity to test the state of the art in the field at the time. At the time the most recent publications were those by Canton et al. (2008) and Kelly et al. (2009) so the state of the art was considered to be an assay for GFPmut-3b fluorescence in a plate reader at 37°C with a reference plasmid included for RPU calculation.

To this end 3 promoters (*tac, lacUV5* and the *T7A1* promoter) were chosen from the BioFAB pilot process as the candidates to be tested and cloned into pAC1. For the characterisation the commonly used DH10B strain of *E. coli* was picked as they have previously been used for characterisation experiments. The choice was taken to characterise in both LB and M9 media as LB is commonly used in many labs and M9 has been used in many of the existing characterisation publications. The characterisation was set up using outgrowth of overnight culture to reduce the background fluorescence that resulted from the accumulation of GFP overnight.

Samples for the promoters and pAC1 and pAC2 (carrying the J23101 reference promoter) were started from fresh transformation plates (3 colonies for n=3) and were grown overnight in LB media. Samples were dilution into LB or M9 supplemented media for a 90 minute outgrowth before a second dilution and transfer to the assay plate. The plate was then assayed in a BMG Omega plate reader with fluorescence and absorbance scanned every hour with shaking for around 10 minutes prior to measurement. This extended characterisation time frame was chosen to allow the acquisition of data points across the entire growth profile of *E. coli* as promoter output has been shown to vary during growth by Zaslaver et al. (2009). Samples were taken from the microplate twice during the characterisation at the times considered to be the most worthy of investigation (around mid-log and in late log phase). These samples were analysed by flow cytometry.

The decision was taken to use a datasheet as the medium for returning data to the BioFAB as this enabled the transfer of the data in addition to a large amount of detail regarding the characterisation itself. As no datasheet had yet been suggested for constitutive promoter BioParts a new design was developed for this purpose. To clearly display all the important information about the characterisation the datasheet was split across two pages; a first page displaying the data (figure 6.4) and a second page displaying the key metadata associated with the characterisation (figure 6.5). Data obtained from cells grown in LB and M9 media was analysed and displayed separately on the first page of the data sheet and each half split into plate reader and flow cytometry data as the methods give fundamentally different data. The flow cytometry results were reported in background corrected fluorescence and Relative Promoter Units PRU while plate reader data was reported in OD normalised fluorescence (also background corrected) and RPU as these were considered the most appropriate result formats and corresponded well with the BioFAB's data format. Also included on the data page were the raw trace files for both types of data to allow observation of growth and expression trends.

The RPU results for the *lacUV5* promoter (bold numbers in the left hand graphs in figure 6.4) were fairly consistent between both media though differed more significantly between data types. Additionally the RPU value obtained at mid log (3 hours) differed from that obtained at late log (5 hours) particularly in the plate reader results. The RPU results for *tac* promoter shown in figure 6.6A were very mixed particularly for the plate reader and the cause of this was not particularly clear. The results for the *T7A1* promoter were discarded as the flow cytometry results for this BioPart appeared to show two distinct populations (figure 6.6 B) which may have been caused by sample contamination or possibly by BioPart loss or mutation as had been observed by Canton et al. (2008).

The second page of the datasheet (figure 6.5) displayed all the metadata associated with the BioPart. A map and the sequence for the BioPart (in this datasheet the sequencing trace) was included as this may be useful for a BioPart user to have to hand to understand the specific DNA environment the BioPart was characterised in. The BioPart and vector sequences were included with the datasheet electronically. The gate used for flow cytometry was supplied as co-ordinates as these may be useful for someone re-analysing the data. Finally an extended 'other metadata' section (right hand side of figure 6.5) was included which contains all the intricate experimental details someone would need in order to carry out the characterisation again and accurately documents the specific context under which the characterisation was carried out.

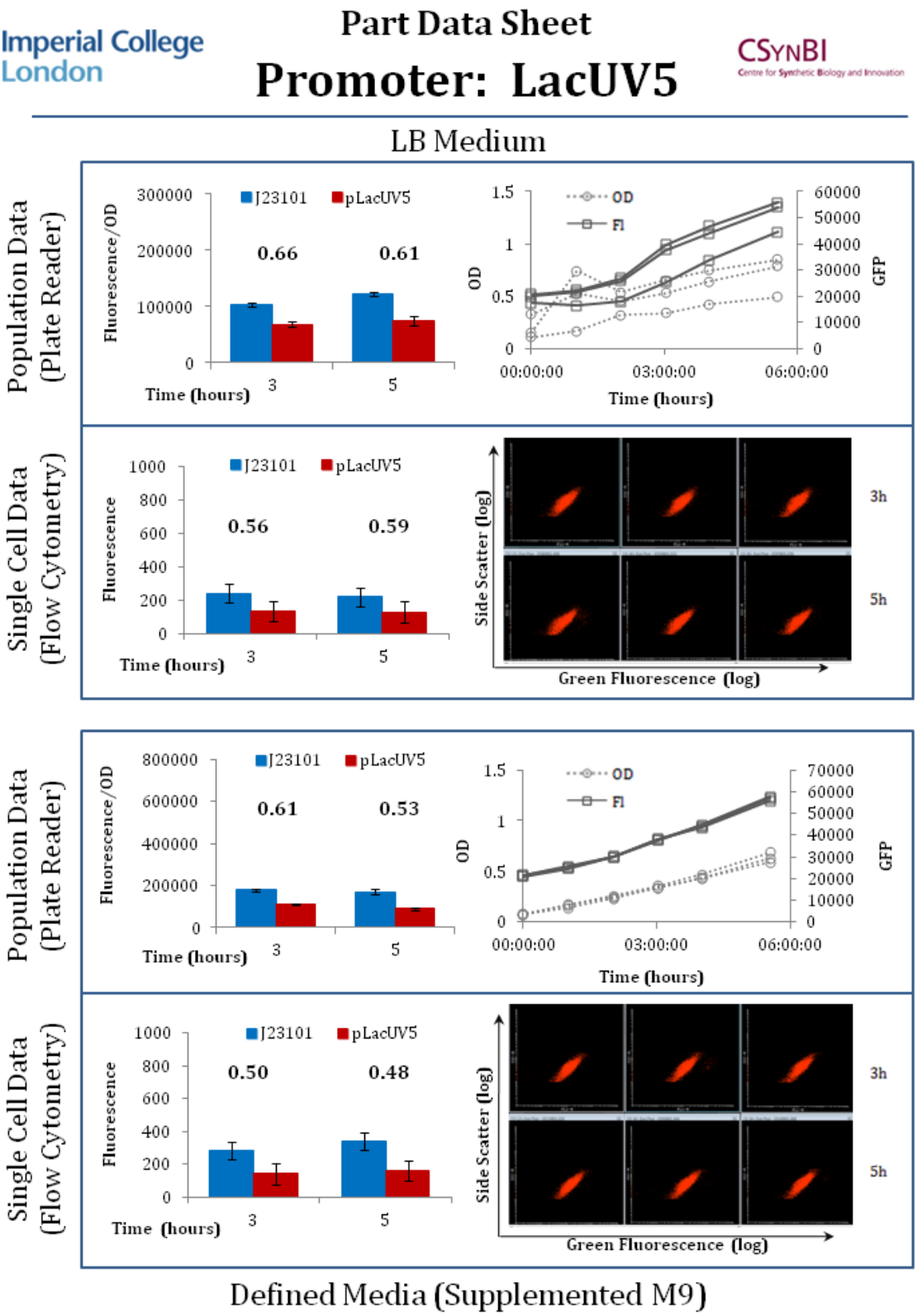


Figure 6.4 First page of a new datasheet design generated to document the characterisation of a constitutive promoter BioPart.

The first page of the datasheet was designed to display both the calculated characteristics (left hand side, bold numbers are promoter output in RPU) and raw data obtained by both a plate reader and flow cytometry. The results were split according to the media the cells were grown in during characterisation. Values for analysed results area means and the errors are one standard deviation. For all graphs, n=3.

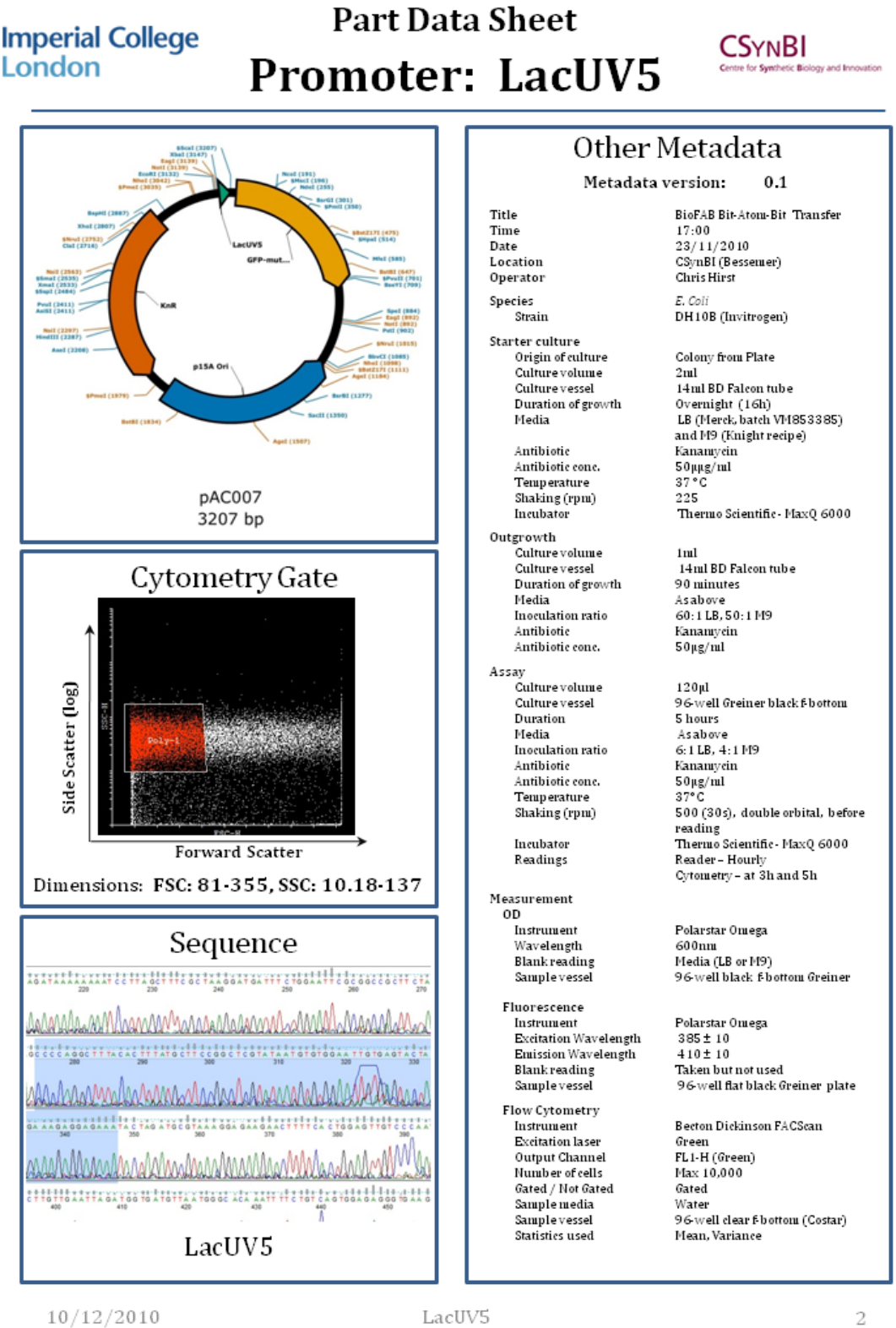


Figure 6.5 Second page of a new datasheet design generated to document the characterisation of a constitutive promoter BioPart.

The second page was focused on the metadata associated with a BioPart which should be instantly available with the characterisation data. The metadata is broken into 3 parts; the physical metadata (i.e. the DNA metadata – the plasmid topology and promoter sequencing results), the analysis metadata (flow cytometry gate) and the experimental metadata.

A

pTac	Plate Reader		Flow cytometer	
	3h	5h	3h	5h
LB Media	1.78	2.83	3.35	3.86
M9 Media	3.44	2.94	4.00	3.53

B

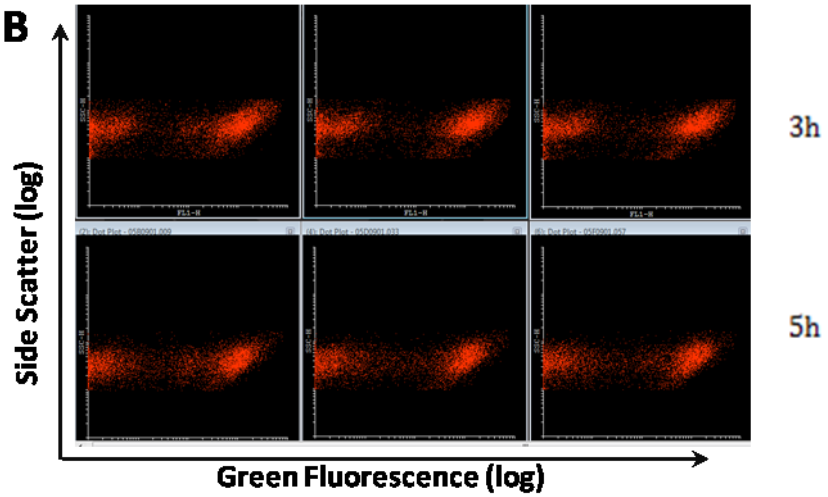


Figure 6.6 Results for *tac* and *T7A1* promoters as part of the BioFAB BAB experiment.

As part of the BioFAB BAB experiment the *tac* and *T7A1* promoters were also characterised. The RPU results for both data streams at 3 and 5 hours for *tac* are summarised in A. Flow cytometry dot plots for size and fluorescence for *tac* are displayed. The profiles suggested a mixed population of unknown origin so all *TA17* promoter results were considered unreliable. Values displayed are the mean RPU. For all results n=3.

6.2.2 Impact of volume and shaking on cell growth and metabolism

In preparation for optimising characterisation on an automated platform conditions under which characterisation was to be carried out were tested to identify conditions more suitable for use. There was particular concern for how much oxygen would be available to cells given the long time frame being considered and the acknowledged poor oxygen transfer parameters of 96 well microplates. The main factors which should be able to influence growth in a microplate are temperature, shaking and well volume. As 37°C is the optimal growth temperature for *E. coli* it was decided to vary the shaking and well volume primarily to monitor the effect on *E. coli* growth but also to attempt to test for aerobic or anaerobic metabolism inside those cells.

Growth experiments were set up to test the growth rate of *E. coli* MG1655 in LB media as this allowed faster growth than M9 media. Plasmids pAC1 and pAC2 from the BioFAB Bit to Atom to Bit experiment were chosen to use for most of the testing. In the microplate experiments samples containing biosensors for the *E. coli* fumarate and nitrate reducatase regulator protein (FNR) were included. The FNR protein has been shown to modulate the metabolism of *E. coli* cells in the absence of oxygen (Salmon et al., 2003) and plasmids carrying a copy of the *fnr* gene causing over-expression of the protein inside cells were developed for oxygen sensing. The plasmids also carried a promoter which controlled GFP expression that was either positively or negatively regulated by the FNR protein in the absence of oxygen (figure 6.8 A) (Choe and Reznikoff, 1991, Salmon et al., 2003). As GFPmut-3b requires oxygen for fluorophore formation these sensors were designed to be used as a pair based on comparison between an activated promoter and a repressed promoter in separate samples (as there should be similar amounts of oxygen in each). The use of a pair of sensors should also help to remove any impact from a change in growth rate that may occur as this should also affect each sensor equally.

Experiments were carried out in 3 identical microplates were set up from LB overnight cultures containing *fnr* biosensors, pAC1 and pAC2 plasmid carrying MG1655 in wells with a volume of 100 or 200µl. The 3 plates were grown at 37°C and 0rpm, 200rpm or 500rpm with measurements taken every hour for 5 hours. In parallel MG1655 cultures containing pAC1 were grown in flasks and measured over a similar time frame to obtain a baseline comparison to conditions approaching optimal growth (20ml/flask or 5ml/tube) or used during industrial scale up (100ml/flask).

None of the samples in the microplate showed a particular difference in growth rate compared to the other samples on that plate in the same media volume. When the optical density of the samples in 100µl of media were plotted and compared there was little change as the speed of

shaking was increased. Samples at 200 and 500rpm grew at almost precisely the same rate and had similar levels of error. Samples grown at 0rpm appeared to have a potentially higher rate of growth (figure 6.7A) but as the samples approached saturation the measurements become prone to error due to an unknown cause. The growth rate for samples in 200µl of media at 500rpm was the same as samples grown at 200 or 500rpm in 100µl but as is shown in figure 6.7B there was a distinct drop in growth rate as the rate of shaking was reduced. These growth results were compared to those obtained by growth in flask or tubes and as can be observed in figure 6.7C the samples in 200µl at 200rpm grew at the same rate as cells in a flask with 100ml of LB at 200rpm shaking implying that the conditions may be similar.

It was hoped that the ratio of the fluorescence per cell of samples carrying an FNR activated promoter versus the FNR repressed promoter would give a rough estimate of the current metabolic state of the cell in regards to anaerobic metabolism. The ratio of the promoters in each pair at time point zero were an appropriate baseline as the cells up to this point were grown in tubes rotated to allow proper aeration. At time zero (figure 6.8B) the ratio of the sensors was fairly even across all the volumes and conditions implying that any change would be the results of the conditions they have been introduced to. The results after two hours of growth in figure 6.8C appear to agree with the optical density results. In the 200µl samples there appeared to be a relationship between shaking speed and the promoter ratio for both the *dmsA*/*ndh* and *nark*/*ndh* promoter pairs with more 'anaerobic' output in the unshaken samples. For the 100µl per well samples there was no clear difference for shaken samples but again there was a much larger 'anaerobic' output when samples were not shaken. These results seemed logical as shaking would increase the surface area through which oxygen could diffuse into the media thus increasing oxygen availability. These results would suggest that conditions of at least 500rpm of shaking and 100µl per well of media should be used for characterisation.

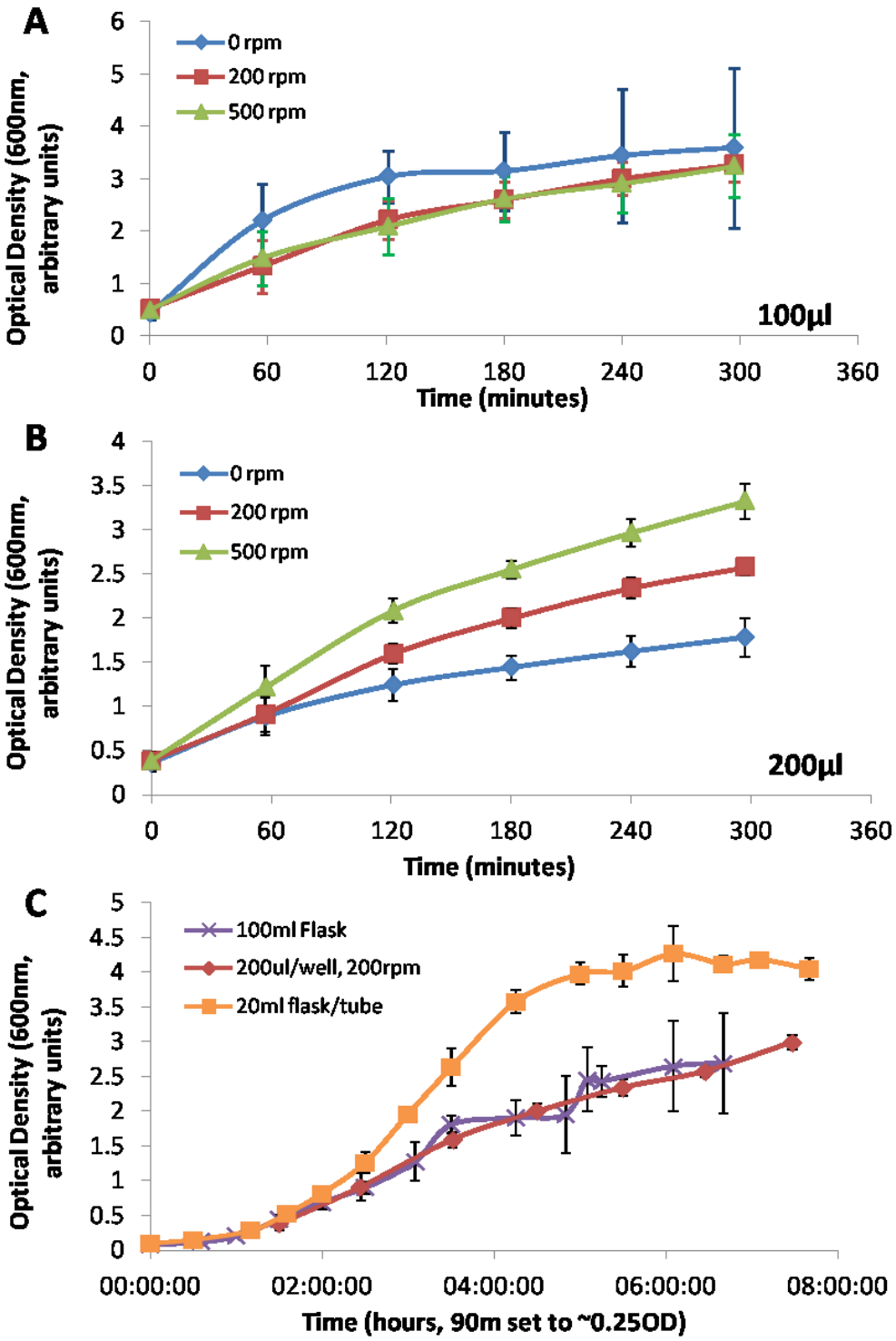


Figure 6.7 Comparison of *E. coli* MG1655 growth in various setting in microplate and in conditions likely to be encountered in laboratories and industry.

Samples of *E. coli* were grown at various shaking speeds likely to be encountered in a microplate reader to determine the growth obtained with 100 (A) or 200µl (B) of sample per well (n=10). For 100µl per well shaking has a limited effect. Growth in 200µl at 500rpm gives comparable results but lower shaking speeds with 200µl of media leads to a reduction in growth. The results were compared to samples grown in a flask (100ml) or under optimal observed conditions (20ml in a flask or 5ml in a tube) with n=3 and n=4 respectively. The results for 200µl/well at 200rpm were similar to 100ml in a flask indicating that growth better than this should be appropriate for characterisation. Values displayed are means and the error is one standard deviation.

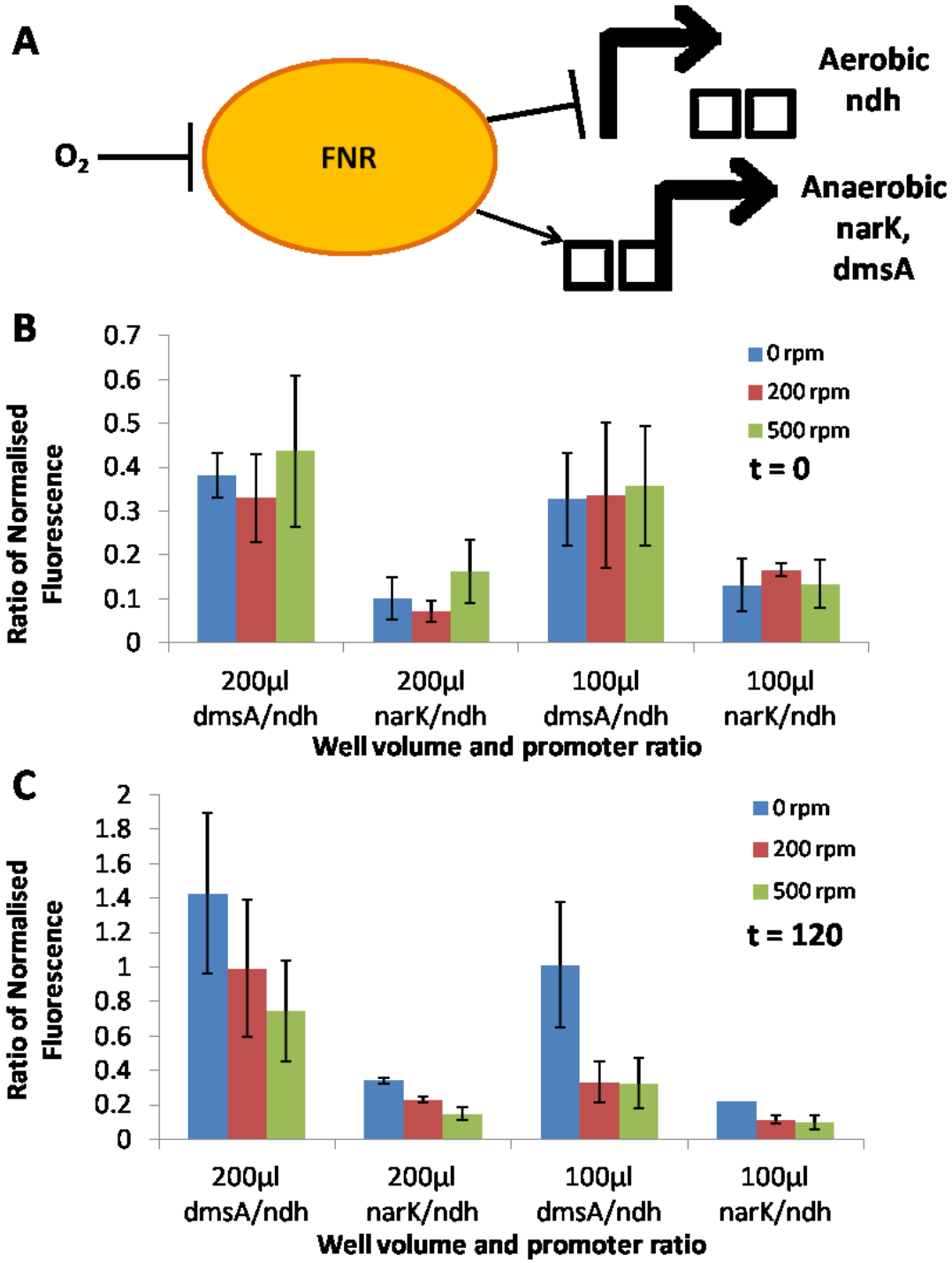


Figure 6.8 Identifying growth conditions for *E. coli* MG1655 in a microplate likely to allow aerobic or anaerobic growth using paired sensors.

2 pairs of promoters (*narK*/*ndh* and *dmsA*/*ndh*) where one is activated and the other repressed by the FNR protein (Unden et al., 2002) were used as *in vivo* sensors for aerobic or anaerobic metabolism (A). The mean ratio of fluorescence per optical density unit for these pairs was calculated for the start of the experiment (B) and after 2 hours(C). B indicates that there was no significant difference between the output ratios of the promoters for any of the samples immediately following transfer to microplates. After 2 hours (C) it was observed that the ratio values increased in all the unshaken samples. While the 100µl/well shaken sample ratio values appeared to remain roughly the same there was a notable increase in the ratio values for the 200µl/well samples. The error was one standard deviation. For all results, n=3.

6.2.3 Test of the ability of automated platform to reproduce results obtained by manual characterisation

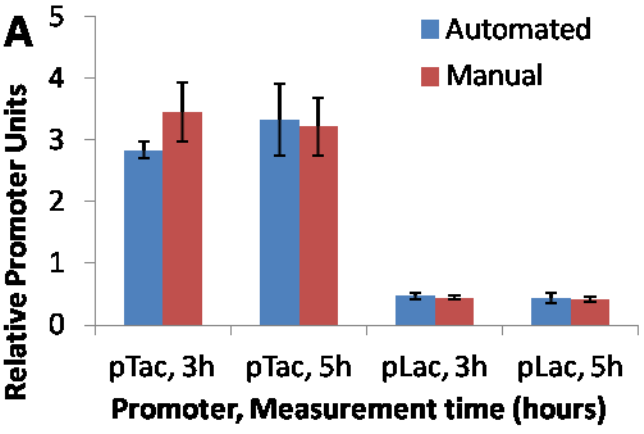
Automation has been used to increases the accuracy and throughput of processes to levels unachievable by humans. Characterisation was an ideal candidate for improvement by the addition of automation but how well automation equipment would be able to cope with the task was unknown. As liquid handling platforms work with microplates primarily it was important to find out if it could cope with the small volumes required as well as a human? An experiment was set up to test if it could cope with the assay and how well it performed in comparison to a human manually carrying out the protocol.

The earlier testing had indicated that a volume of 100µl and 500rpm shaking would be most appropriate it was necessary to upgrade the Theonyx liquid handling platform (detailed in methods 5.1 and briefly in results 6.3.1) to be suitable for purpose. Fortunately while the robotic platform to be used was being prepared, taught and calibrated (see methods 5.3) a second Theonyx platform was available and already partially calibrated in the CSynBI. While this platform was primarily being used for DNA preparation techniques it carried all the necessary features to carry out a trial characterisation. This platform was programmed to run an automation acceptable version of the manual characterisation protocol carried on the BioFAB BAB samples and so it was decided to make the platform would rerun the initial characterisation experiment to see how reproducible the results were and if it could improve the accuracy of results. One of the few protocol modifications was the inability of a plate lid to fit into the plate reader and as the platform could not handle plate films the plate had to be left exposed and so the outer wells were filled with water to reduce any risk of edge effect. Again the *tac*, *lacUV5* and *T7A1* promoters were characterised at 37°C in a plate reader (a Biotek Synergy HT) with shaking prior to measurement. Samples were again taken at 3 hours and 5 hours for flow cytometry analysis.

The automated platform carried out the manual characterisation well by efficiently outgrowing and assaying the samples well with minimal user input. The results for the BioFAB BAB promoters were consistent in M9 media for both *tac* and *lacUV5* as can be seen in figure 6.9A with 3 of the 4 RPU results being almost the same. The results for LB were less consistent. Evaporation proved to be higher than expected causing flow cytometry sampling to intermittently fail at the 5 hour sample point and unfortunately making the results unreliable due to wildly inconsistent n numbers.

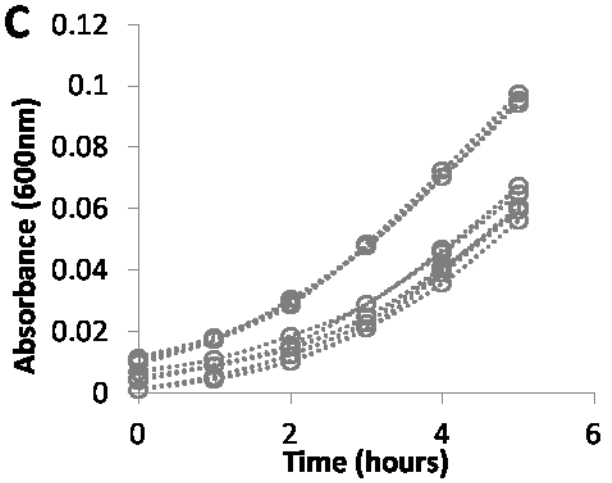
The accuracy of results for the characterisation was compared by calculating the average standard deviations of the RPUs at 3 hours and 5 hours for the *tac* and *lacUV5* promoters as percentage of the RPU value at that time (i.e. the coefficient of variation). These results from the plate reader data are shown in the table in figure 6.9B however some of the change in accuracy may be down to the change in measurement equipment. For LB media there was a significant drop in error upon automation although for M9 error there was a small increase.

In an attempt to determine how the platform and equipment change could both increase and reduce error the data was looked into in more detail and it was discovered that if the standard deviations were calculated for each day (3 RPUs) rather than across all 9 together the average error dropped dramatically for both M9 and LB. Upon closer inspection this was observed to be caused by the differences in the amount of cells and possibly growth rate at each time point on different days and the effect this has on fluorescence. This can be observed in the raw absorbance and fluorescence traces for the *lacUV5* promoter in figure 6.9C and D. For future experiments the much tighter control of *E. coli* growth was needed to improve the quality of characterisation and a way to use plate lids or seals was required.



B

	Average Error		
	Manual	Automated	Automated (days separated)
LB Media	19.7%	9.81%	7.31%
M9 Media	11.1%	13.5%	5.81%



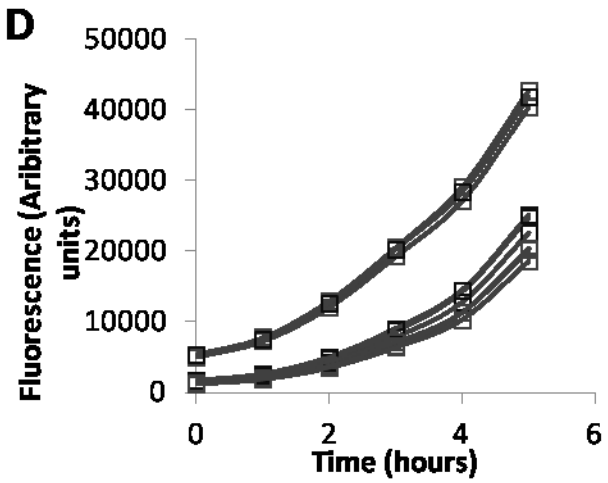


Figure 6.9 Comparison of results from manual and semi-automated characterisation.

The same samples that were tested manually before were re-characterised on an Aviso Theonyx platform to see how accurately it could carry out the procedure in comparison to a researcher. The platform reproduced similar results to those produced by manual characterisation in M9 media (A) with 3 of the 4 RPU scores obtained similar in both experiments (n=3 manual, n=9 automated). There was a significant reduction in error/deviation (calculated as the standard deviation as a percentage of the mean value) associated with measurements when carried out in LB media but apparently an increase in error when M9 media was used (B) (n=3 manual and automated separate days, n=9 for total automated). However if the error for individual runs was calculated instead this was significantly lower than for the manual characterisation. This difference in accuracy across days compared to within days may be due to small differences in starting optical densities and thus many samples lagging behind as is visible in the absorbance (C) and fluorescence (D) traces for the *lacUV5* M9 samples.

6.3 Developing Automated Characterisation

6.3.1 Theonyx platform set-up and pipette calibration results

The platform for characterisation was an Aviso Theonyx liquid and plate handling robot based on a Sias Xantus hardware chassis with an Aviso programmed GUI for easier scripting. The Theonyx platform arrived at the Centre for Synthetic Biology and Innovation in its default configuration with pipetting needles, a few pipetting positions, a tube rack and a Synergy-HT plate reader. To be used for characterisation it needed many upgrades to make it suitable for working with *E. coli* samples. For characterisation it would require a sample incubator, the tip needles swapping for disposable tip heads, tip racks and thorough teaching and calibration of the pipetting pumps to enable the accurate use of tips.

To improve cell growth a microplate incubator was attached to the platform to allow incubation of samples outside of the plate reader and under cover of a plate lid. Inspiration was taken from an incubator of the same model and with similar modification in the lab of Dr. R. Weinzerl. A Mikura Ventura 2000 was purchased and modified by Gary Jones to allow the platform to access plates and store them inside during incubation. This incubator was clamped underneath the deck to prevent movement relative to the platform. The pipetting needles were replaced with Tecan tip heads which use disposable tips to remove the risk of cross-contamination. With all the correct components in place the tips required calibration. Calibration was vital for tip usage as the differing tip dimensions affect the accuracy of pipetting by the pumps. Calibration was carried out for aspiration and dispensing of liquids separately. For large this was easily achieved using a precision scale however at smaller a new method using dyes was devised.

The Synergy HT plate readers use a quality control dye based on Yellow number 5 powder for routine testing. Dye solution was transferred between wells on microplates multiple times and the absorbance changes tracked to calculate the volumes aspirated and dispensed. The statistics for these transfers were calculated and used to update the firmware settings until the average volume and pipetting variation were within the manufacturer's specifications.

The calibration and addition of new modules prepared the platform for use in characterisation as it had the tools to subculture biological samples without risk of cross-contamination due to the use of pipetting needles. With the calibration data it was possible to calculate the error for each tip and this is shown in figure 6.10 and for all but one tip at one volume it was possible to reduce this error in pipetting volume to significantly less than the manufacturer's specifications. The platform

pipetting error was similar to the error of commercially available hand pipettes above 5µl. The variation values obtained for particularly low volumes (20µl and less) were probably over estimations given the small absorbance changes at these volumes. For the 20µl or less volumes the detection error from the plate reader alone would be 1%.

	Aspirate					Dispense		
Volume	**1000**	**200**	**20**	**5**	**1.5**	**1000**	**200**	**20**
Hardware Specification	**0.50%**	**1.00%**	**1-10%**	**10.00%**	**20.00%**	**0.50%**	**1.00%**	**1-10%**
Pipette 1	0.49%	0.29%	**1.68%**	4.96%	9.23%	0.14%	0.41%	**4.87%**
Pipette 2	0.46%	1.12%	**0.69%**	6.58%	8.18%	0.15%	0.31%	**4.40%**
Pipette 3	0.16%	0.16%	**0.52%**	2.63%	13.3%	0.29%	0.80%	**3.78%**
Pipette 4	0.07%	0.87%	**1.40%**	7.07%	11.2%	0.19%	0.46%	**4.24%**
Pipette 5	0.43%	0.66%	**1.20%**	3.68%	10.3%	0.37%	0.79%	**3.42%**
Pipette 6	0.33%	0.66%	**1.04%**	3.35%	11.1%	0.12%	0.50%	**3.57%**
Pipette 7	0.36%	0.67%	**2.09%**	2.52%	9.43%	0.24%	0.60%	**3.02%**
Pipette 8	0.39%	0.66%	**1.65%**	5.47%	9.37%	0.27%	0.30%	**2.97%**

Figure 6.10 Pipette calibration results for the characterisation platform in comparison to the manufacturer's specifications.

As part of re-tooling the platform to allow the use of disposable tips the pipette pumps needed calibrating to correctly use the tips. Water or a dye solution was repeatedly pipetted and the statistics calculated to allow updating of the equipment firmware. This resulted in observed pipetting variation under the manufacturer's specification and puts most in the range of hand pipettes. The error range for 20µl is put at 1-10% as there is no specific manufacturer specification. Only a single pump narrowly failed to have a lower variation than the specification and has not lead to any noticeable effects in experiments.

6.3.2 Testing absorbance based dilutions as a way to improve consistency

With the platform fully taught and calibrated the full characterisation methods needed to be developed by incorporating the advantages of using an automation system to improve the quality of results. The earlier testing had shown the need for improvements to the consistency of growth. For automated characterisation this was even more critical as it had been decided to transition from LB and M9 media which each have components which are not produced to a consistent specification to the fully defined *E. coli* optimised EZ MOPS media first developed by Neidhardt et al. (1974). Glucose was chosen as the carbon source as this would turn off catabolite repression systems in *E. coli*. To cope with all this, a new strategy for sample preparation was required.

An initial testing experiment with MOPS media (with 0.4% (W/V) glucose or glycerol) was prepared to compare growth in the robot incubator at 37°C with 600rpm shaking with growth in LB under the same conditions. Top10 *E. coli* carrying pAC1 or pAC4 were grown overnight in LB before being diluted to an absorbance of approximately 0.01 units (100µl) in appropriate media following outgrowth and grown for 7 hours with absorbance measurements taken every 3o minutes.

The *E. coli* grown in MOPS with either carbon source grew at approximately the same rate as in LB for the first couple of hours after which the rate of growth in LB diminished. Growth in glycerol MOPS was fairly consistent with growth slowing from around 4 hours as can be seen in figure 6.11A. The variation in absorbance of samples grown in MOPS media with glucose however began to increase rapidly at around 3 hours. This was caused by the growth of a few of the samples beginning to slow from around this point on while others continued to grow rapidly for a while longer. The significant drop in growth rate at around 4 hours indicated that there was only a short time frame for the collection of results at 37°C in MOPS media. While this had been acceptable for earlier testing for full characterisation as many data points as possible would be preferred. The inconsistent growth in glucose MOPS was also somewhat concerning so the decision was made to reduce the temperature of assay conditions for characterisation in order to cope with the rapid growth of *E. coli* in this media.

Diluting samples to a set absorbance to begin the assay had worked well and so the automated protocol was developed to allow the platform to carry out similar dilutions on samples for characterisation. A pair of absorbance based dilutions were introduced to the protocol with one either side of the outgrowth step to bring the cell populations into alignment. This modified protocol was used to test the new suggested characterisation conditions of 30°C and 600rpm with 0.4% glucose MOPS for effectiveness with Top10 *E. coli* carrying the pAC1 and pAC4 plasmids. A range of

dilution concentrations were used for each dilution step resulting in columns of samples in a microplate (6 columns each with an n of 8). Samples were then grown every 5 hours and measured every half an hour to determine how effective the set up had been.

The results for each day were positive with the majority of samples starting from a fairly consistent absorbance and growing at a very similar rate. When this was repeated across 3 days the results were much more mixed with only 1 of the 6 dilution concentrations used being consistent across all 3 days and within days (figure 6.11B). The dilution 'concentration' data (absorbance) was inspected and it was noted that the variation observed in samples following outgrowth was relatively high making the second dilution step essential. After 5 hours the absorbance reached was still relatively low which would need to be dealt with in future.

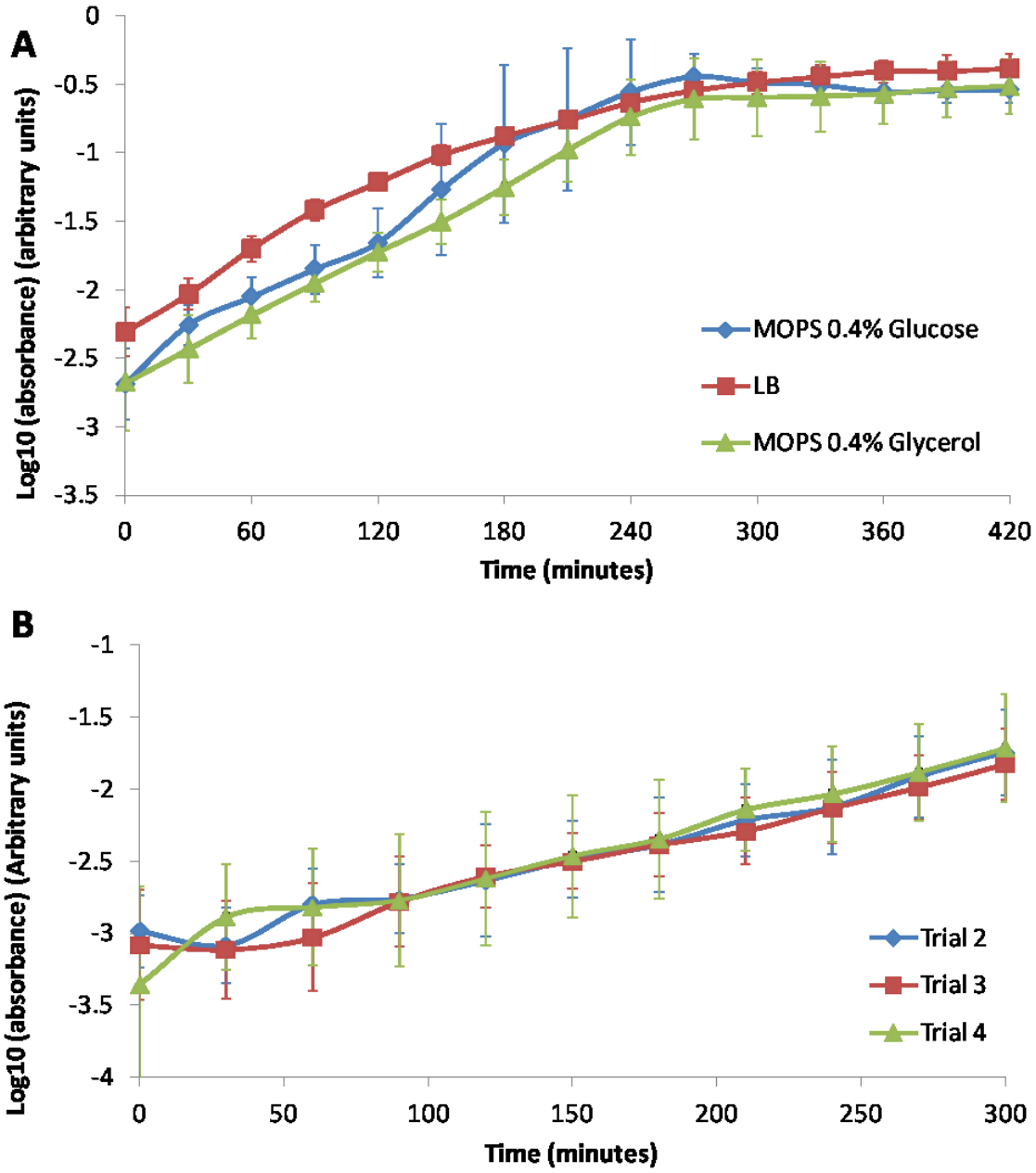


Figure 6.11 Improving consistency by more accurately diluting according to absorbance.

As MOPS is a rich media *E. coli* grew very fast and rapidly reach large population sizes. A) Top10 cells grown in MOPS with glucose or glycerol at 37°C initially grew at a similar rate to LB but rapidly approached stationary phase at which point variation increased (2 different plasmids, n=4). B) By using absorbance based dilution steps the automated set up protocol managed to more precisely replicate the set up for experiments resulting in consistent day to day cell growth in 0.4% glucose MOPS (2 different plasmids, n=8 per trial at 30°C). All errors are standard deviation.

6.3.3 Testing MG1655 and MDS42 for use as standard characterisation strains

With dilution based steps tested and shown to work it was considered a suitable time to test how well the automated platform could characterise BioParts in MOPS media. At this time involvement with the ST-Flow working group had begun which triggered changes for the protocol. The ST-Flow working group had reached a standardisation decision to work primarily from one media and two strains of *E. coli.* The media was agreed to be the glucose based EZ rich MOPS and the two stains for use were MG1655 from the Busby Lab (2013) and the multiple deletion strain MDS42 (Posfai et al., 2006). As both of these strains had been picked for future use for the CSynBI it was considered that this was an opportune time to make one of these the primary strain for testing on the characterisation platform and to optimise protocols around this strain. To identify which of the two strains would be most appropriate for automated characterisation the MOPS media characterisation test was carried out with both strains and the results compared.

The characterisation protocol had been slightly modified since its last use. The overnight cultures were still grown in LB but now at 30°C prior to absorbance based dilution. 6 plasmids carrying members of the Anderson constitutive promoter collection from the Parts Registry (n.d) along with a negative control were transformed into both strains for characterisation. A 90 minute outgrowth followed before the second dilution and transfer to an assay plate. The most consistent values from the Top10 *E. coli* test were used as the absorbance values to dilute to as they had allowed consistent growth although the final absorbance reached during assay were low. In an attempt to compensate for this the characterisation time was extended to 6 hours rather than the 5 used previously. The entire characterisation protocol was now carried out in 0.4% glucose EZ MOPS at 30°C and 700rpm shaking with flow cytometry samples taken after 3 and 6 hours (instead of 3 and 5 previously). To test how much data could be gathered and to prepare for induction testing the interval between measurements was reduced from 30 minutes to 7 and a half minutes for the period between 120 and 200 minutes.

The characterisation results on the 6 promoters in MG1655 and MDS42 indicated that the level of output from the 6 promoters remained relatively the same in both strains (all 6 promoter RPUs were within the margin of error). This was also observed to be true for the flow cytometry data (not shown).

The dilutions settings used for Top10 *E. coli* worked reasonably well when applied to MG1655 and MDS42 although the absorbance achieved by the end of the characterisation run was still fairly low (figure 6.12 B). Manual intervention had been necessary on multiple occasions indicating that

the dilutions would need to be optimised for whichever of the two strains was chosen as the standard for characterisation. It was also observed that the increased measurement interval negatively affected cell growth as can be seen in the dips appearing sporadically during the rapid measurement region of the experiment particularly for MDS42 (figure 6.12 B).

It was considered that MG1655 would be more suitable for characterisation as this strain would be more genetically similar to the strains of *E. coli* by Synthetic Biologists and exhibited a higher growth rate which would make it easier to obtain data for a larger proportion of the growth curve and the genetics. Development of MDS42 characterisation could be returned to once a finished automated characterisation protocol had been developed.

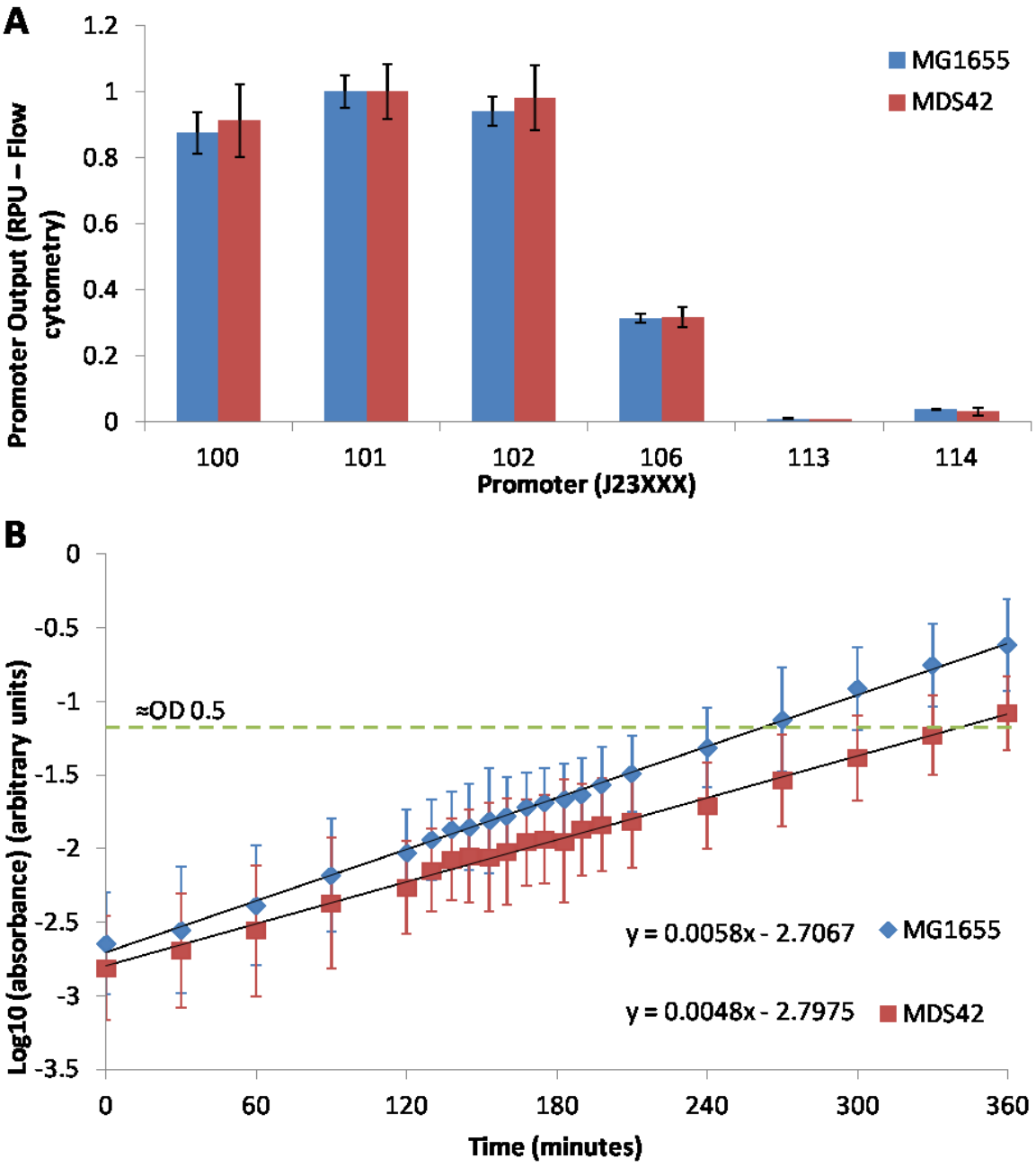


Figure 6.12 Results for the attempted characterisation of several Anderson promoters in MG1655 (Busby) and MDS42 *E. coli* using the assay protocol developed using TOP10 *E. coli*.

MG1655 and MDS42 were selected as the 2 standard *E. coli* strains for use as part of the ST-Flow project supported by EU FP7. The characterisation results for 6 promoters were used to compare to strains. For all 6 promoters the results for RPU output were within the margins of error indicating they had the same relative strengths (A). While the characterisation results remained the same the growth rate of MDS42 was significantly slower (B) and both strains took most of the characterisation run to reach 0.5 OD units (traditionally considered mid-log). For all experiments n=9 and error was one standard deviation.

6.3.4 Standardisation of MG1655 growth by optimised dilutions

To enable the platform to get the highest quality of data for BioParts characterised in MG1655 the dilution steps were critical. The aims for the dilutions will be 3 fold; firstly the dilutions must be robust and unlikely to fail when carried out, secondly the dilutions must provide samples for assay that are at very consistent population levels day to day and finally the cells provided for assay should be a suitable population size to yield suitable growth profiles across the length of the characterisation.

To rapidly iterate through the possible combinations of dilutions various pairs of absorbance based dilutions were tested on each run. As with the *E. coli* Top10 experiment (6.3.2) each column of wells was used to test at a different combination of dilutions with the curves produced examined according to the criteria above. MG1655 *E. coli* carrying plasmids SVc, SVc-101, SVc-112, SVc-114, SVc114a, SVc114b, SVc-16 and SVc-117 were grown overnight in 200µl of LB in a microplate at 30°C and 700rpm and used for dilution testing the next day. Cells were initially diluted in 0.4% glucose MOPS media before a second dilution step to a range of absorbance values. Outgrowth for 90 minutes was carried out and the samples diluted to a range of absorbance again. Samples were then grown and measured for fluorescence and absorbance every 15 minutes for 6 hours with the standard growth conditions of 30°C and 700rpm retained. This procedure was repeated many times to determine the best dilution values for the above criteria.

Growth phase optimisation resulted in a pair of dilution values which perfectly fit the optimal criteria. Dilution settings of 0.0117 absorbance units in 100µl MOPS media (blanked) initially and 0.024 absorbance units following outgrowth produced reproducible curves which covered the bulk of the MG1655 growth profile across the 6 hours of characterisation (figure 6.13A). The real results of the optimisation were hidden within the growth profiles of the *E. coli* and required model fitting to see the full detail in the form of parameters. The scripts developed by Catherine Ainsworth and myself were run on the generated data to obtain these growth parameters. The growth was analysed to determine the growth rate, carrying capacity and lag time and when these are plotted together (figure 6.13B) it can be observed that the optimisation had made these growth parameters significantly more consistent. This optimisation when applied to automated protocols using standard conditions will allow the reliable reproduction of growth conditions and may reduce any noise that would normally arise from this.

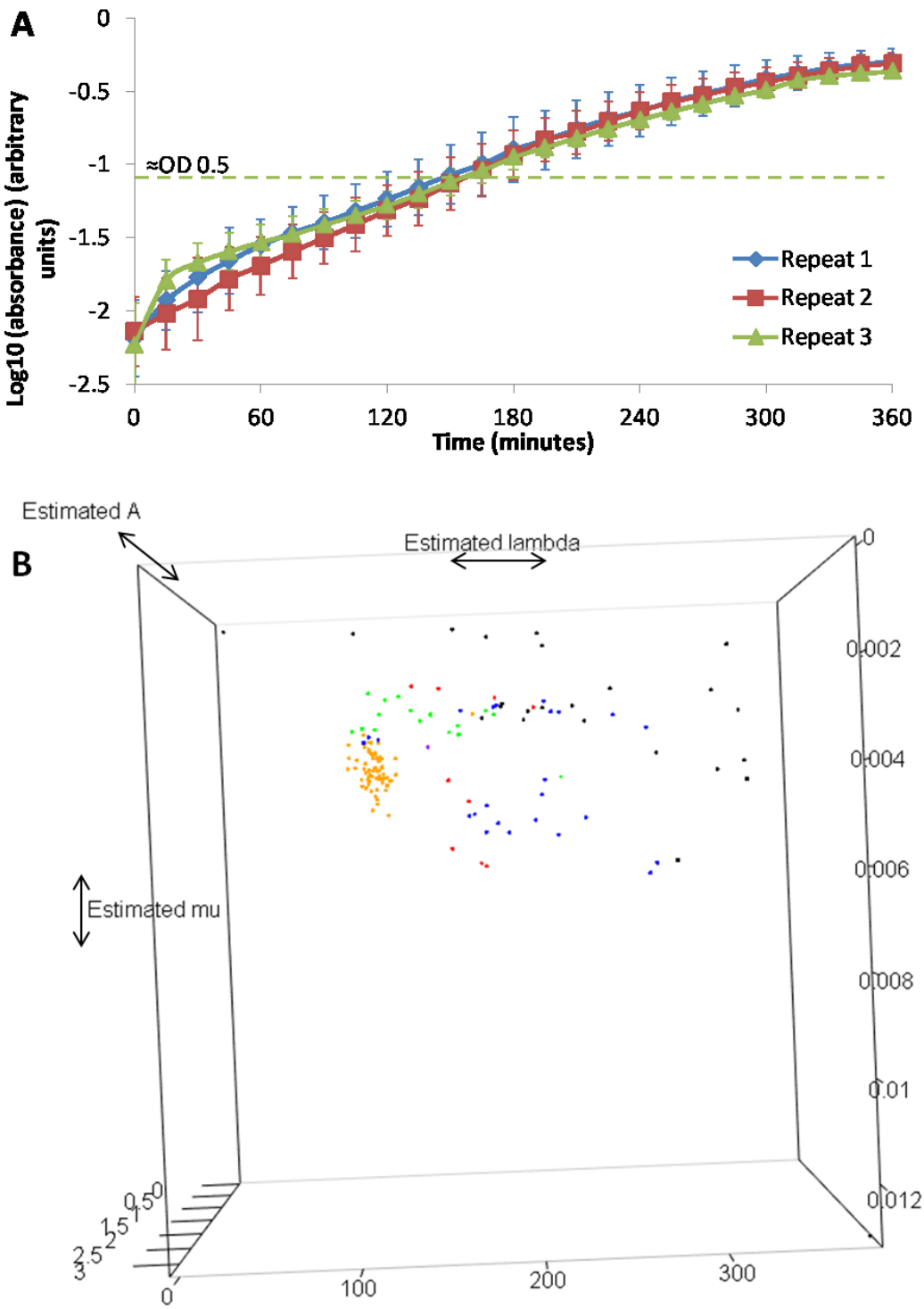


Figure 6.13 Optimisation of platform dilutions gives more appropriate growth profiles and highly reproducible standardised growth.

Accurate dilutions were necessary to ensure high quality reliable growth. Different dilutions were iterated until a combination that gave reliable production of growth curves that covered most of the *E. coli* growth profile was found (A, n=8). The growth of the cells in these experiments was tracked and analysed to fit these results to a model of *E. coli* growth with parameters of lag time (lambda), growth rate (mu) and carrying capacity (A). These parameters were plotted for several similar experiments in chronological order of Black, Red, Blue, Green and finally orange by which time the optimisation had been completed and the samples are tightly clustered in their growth characteristics (B, n > 6). A rotated version of B can be seen in appendix B.

6.3.5 Validation of fully aerobic growth during characterisation

6.3.5.1 Validation of FNR based oxygen sensors

For characterisation of BioParts entirely defined conditions were a necessity as unknown variables in an experiment could alter the results in many ways which may be difficult to quantify or understand. Where there was a factor that could not be controlled it was considered that the best option would be to attempt to understand and document the factor as far as possible. As previously discussed a particularly influential condition which could affect characterisation in a living chassis was the availability of oxygen. Understanding if the environment in which cells would trigger aerobic or anaerobic metabolism at given stages of characterisation was necessary but difficult *in vivo*. The *fnr* based biosensors previously tested in results 6.2.2 were ideal candidates to perform this role but required validation.

To validate the sensors a modified version of the original growth experiment was set up using the characterisation platform. The experimented was based on altering the level of oxygen available to a growing *E. coli* culture carrying the biosensor plasmids by growing the cells in a microplate until they reached a significant population level where upon shaking of the plate was interrupted . In a microplate the rate of oxygen diffusion is controlled by the surface area available for oxygen to diffuse into the media (Maier et al., 2004). When a microplate is subject to shaking the surface area increases and so stopping the shaking would limit the oxygen entry into the wells.

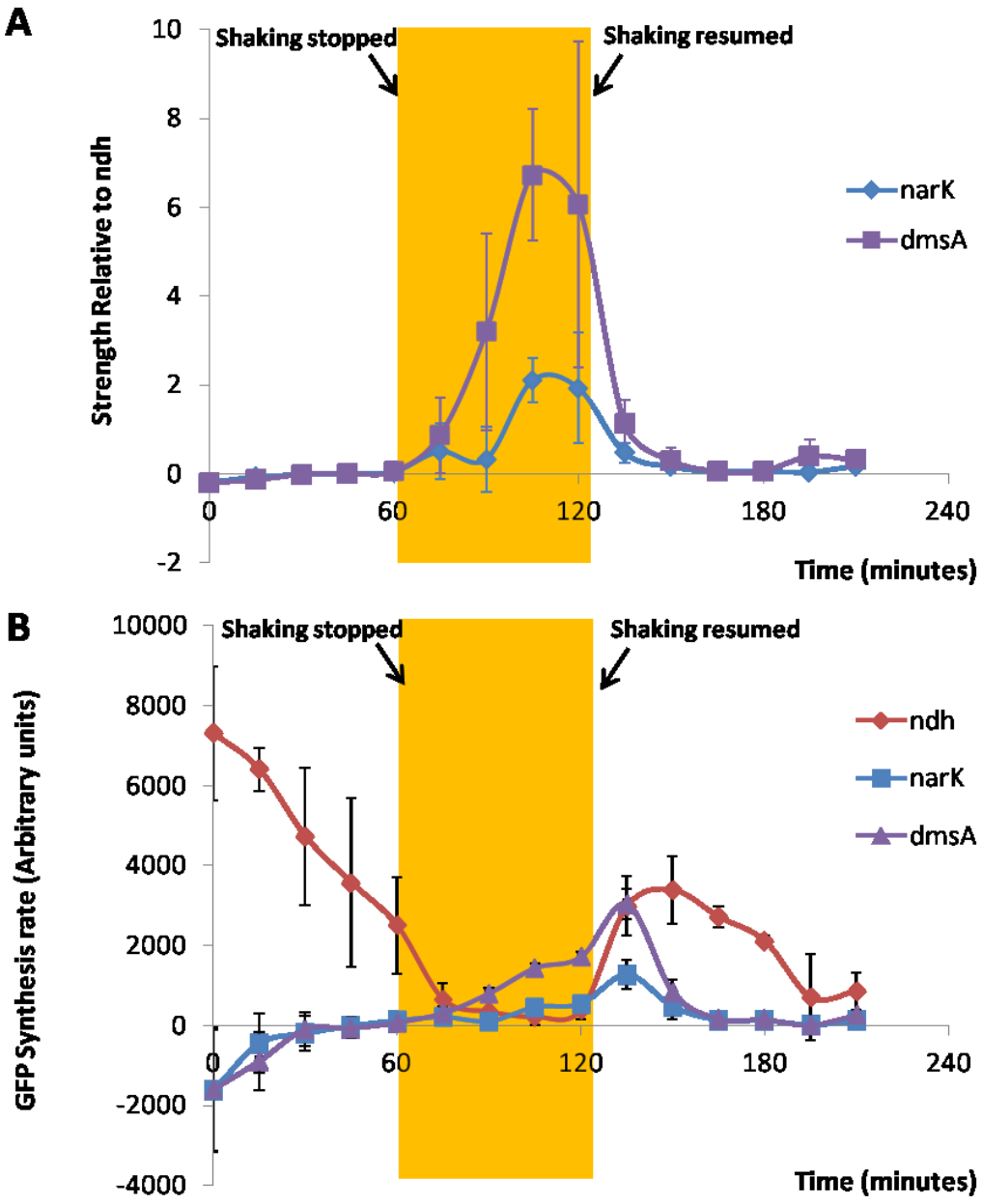
Top10 *E. coli* carrying plasmids with the *narK*, *dmsA* and *ndh* promoter based FNR biosensors were grown overnight before dilution into 200µl of LB with 1% glycerol in the robot incubator. The 1% glycerol was added to the LB to ensure that the cells would reach a high population density and a volume of 200µl was chosen so that the surface area to volume ratio would be even lower when shaking was interrupted. The *E. coli* samples were grown for around 3 hours with the incubator shaking at 37°C and 600rpm until a significant population had accumulated. At this point the incubator shaking was interrupted for 1 hour before shaking was resumed and the experiment continued for 2 further hours. From 2 hours into the growth onwards the fluorescence and absorbance were measured every 15 minutes.

The GFP synthesis rate for all 3 biosensors was calculated and the ratio of output from each FNR activated promoter relative to the FNR repressed promoter (*ndh*) was calculated for each time point. The ratio results can be seen in figure 6.14A and remain at a consistent level below 0.2 prior to the interruption of shaking. At the point immediately following shaking interruption there was only an insignificant increase in fluorescence from the FNR activated promoters relative to the

repressed promoter. Between 30 and 45 minutes the output of the activated promoters increased significantly compared to the *ndh* promoter indicating the environment inside the cells had radically changed. After an hour without shaking the output of the *narK* promoter was almost double that of the *ndh* promoter while the output of the *dmsA* promoter was significantly higher. Interestingly when the individual synthesis rates were considered it was clear that the majority of this signal change was caused by the repression of the *ndh* promoter. After 60 minutes the shaking was resumed and it was observed that within 30 minutes the ratio of output from activated promoters to the repressed promoter returned to the levels seen before shaking was stopped.

These results would indicate that the sensors were behaving as would be hoped as the ratio increased when shaking was turned off. The time taken for the cells to utilise the oxygen available in the media when the shaking was taken off was probably responsible for the observed slight delay. The approximately 30 minute return to the shaken levels would also be consistent with this as time would be required for fresh oxygen to reach the level in the media sufficient to oxidise the FNR proteins and return the output of the sensors to their pre-pause levels. These results suggest that the FNR biosensors are suitable for the task and that values for aerobic metabolism would be expected to be around 0.2 units. As the sensors used one of the key regulators for *E. coli* anaerobic metabolism it is possible that the results also reflect the metabolic state inside the cells.

Interestingly at the point immediately following resumption of shaking all 3 promoters exhibited a significant increase in synthesis rate suggesting that the level of oxygen present prior to the resumption may have been insufficient to allow formation of all GFP fluorophores and highlighting how oxygen may affect characterisations.



6.14 Monitoring aerobic and anaerobic metabolism by use of FNR regulated primer pairs.

The paired biosensors used earlier (6.2.2) were more thoroughly tested by an assay on the characterisation platform to verify their behaviour. The yellow shaded region corresponds with a 60 minute window where the incubator shaking was turned off. The ratios of the synthesis rates of the FNR activated (*narK* and *dmsA*) promoters and repressed (*ndh*) promoter were calculated to indicate the availability of oxygen (A). These ratios stayed consistently low while samples were shaken but rose rapidly when this was stopped. When shaking was resumed the ratios rapidly returned to the previously observed low level. When the individual synthesis rates were examined (B) it was clear that the FNR repressed promoter was significantly stronger when samples were shaken. For all sensors n=8. All values are the mean synthesis rate or ratio of synthesis rates and the error is one standard deviation.

6.3.5.2 Testing standard conditions using the FNR biosensors

As a final verification test for the standard characterisation conditions it was important to understand the likely impact of oxygen on the characterisation environment. The previous experiment with the FNR protein based biosensors had suggested that oxygen may have an impact on GFPmut-3b maturation in an *in vivo* environment making this critical. An environment that was confirmed to be reasonably oxygen saturated would be ideal but otherwise documentation of where the assay may be oxygen limited would offer significant insight into any results observed.

E. coli MG1655 carrying the *narK* and *ndh* promoter biosensor plasmids were used for the verification of standard conditions as the *dmsA* carrying MG1655 samples were prone to the population stability problems observed with the 3 module minimal plasmids observed by Casini (2010). Characterisation was carried out using the standard protocol for inducible promoter BioPart characterisation using the standard conditions of 30°C and 700rpm in 0.4% glucose MOPS media and the resulting ratio of *narK* and *ndh* promoter synthesis rates calculated for each time point.

The *narK* to *ndh* promoter ratio was consistently low for the first 5 hours of the characterisation procedure with the value staying below 1. In the previous assay the ratio only reached 1 when oxygen was being depleted from the media so this implied that oxygen was not depleted from the characterisation assay during the first 5 hours of the characterisation. The ratio of outputs only increased above a ratio of 1 at 5 hours and 30 minutes and continued to rise after this point. This indicated that oxygen may have been depleted from the media by this point and the FNR protein de-repressed however during this period of the assay the cells were beginning to leave exponential phase which may have impacted the biosensors in an unknown way. The aerobic nature of the assay was thus verified for at least 5 hours into the characterisation and the status of the following period of time remained unfortunately unclear.

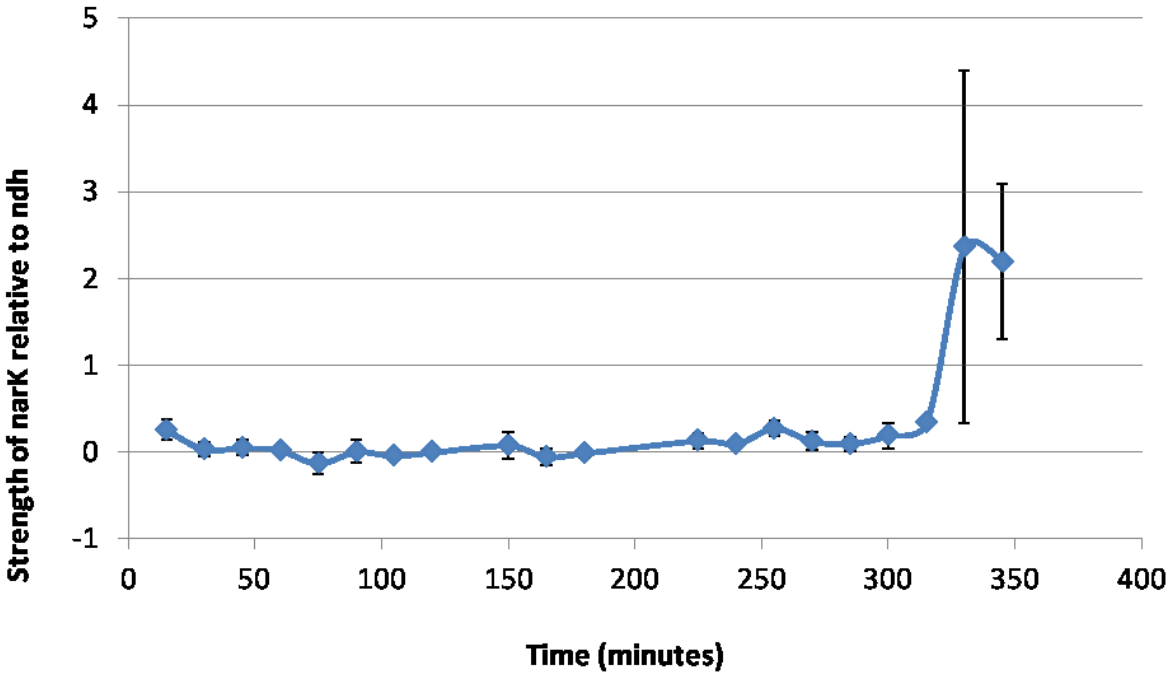


Figure 6.15 Output ratio of FNR regulated *narK*/*ndh* promoter pair under standard characterisation conditions (0.4% glucose MOPS media, 700rpm, 30°C and MG1655).

The *narK* and *ndh* promoter pair was used to test the standard characterisation conditions of the platform. The sensors were included in a characterisation run and the ratio of the *narK* and *ndh* promoter synthesis rates calculated. For the majority of the characterisation the ratio of output from these promoters remained low which indicated an aerobic environment. The ratio increased only during the last 45 minutes of the characterisation and this may imply insufficient oxygen for the samples in wells. Values displayed are the means and the error one standard deviation and the n number was 3.

6.4 Standardised automated characterisation results

6.4.1 The workflow of standard automated characterisation

A reliable protocol proven to work well for a suitable set of characterisation conditions had been developed and was now considered to be of high enough quality to be used for characterisation of promoter BioParts. This method was the standardised automated characterisation workflow and should be appropriate for testing promoter BioParts but also potentially many other expression related BioParts such as ribosome binding sites and riboswitches. The standard itself is comprised of 3 main features; the media and conditions, the method and the data calibration. These features are explained here in brief and used in all of the following experiments (where modifications will be stated).

Standard Characterisation Conditions:

Media: Rich EZ MOPS with 0.4% (w/v) glucose

Growth: 30°C at 700rpm shaking

E. coli: Strain used should be MG1655, from the Busby lab if possible.

The standard characterisation workflow is shown diagrammatically in figure 6.16 and is split into two experimental phases. The initial set-up phase required the platform to take samples grown overnight in LB media and sub-culture them for assay based on their absorbance. The exact mechanism varied but the absorbance values diluted to always remain the same and experiments were carried out under standard conditions. Following outgrowth and a second phase of dilution samples are transferred to the assay plate.

Samples were assayed for 6 hours with measurements taken every 15 minutes. These settings have been proven to be acceptable for MG1655 *E. coli* growth and aeration. If induction was to be carried out the inducer was added after 2 hours as cells will have reached a suitable population for induction. Flow cytometry samples should be taken at 3 hours and 6 hours as these samples should represent mid-log and late log growth phases.

Following assay the data was analysed and calibrated into absolute and Relative Promoter Units. Full details for all steps are included in the methods.

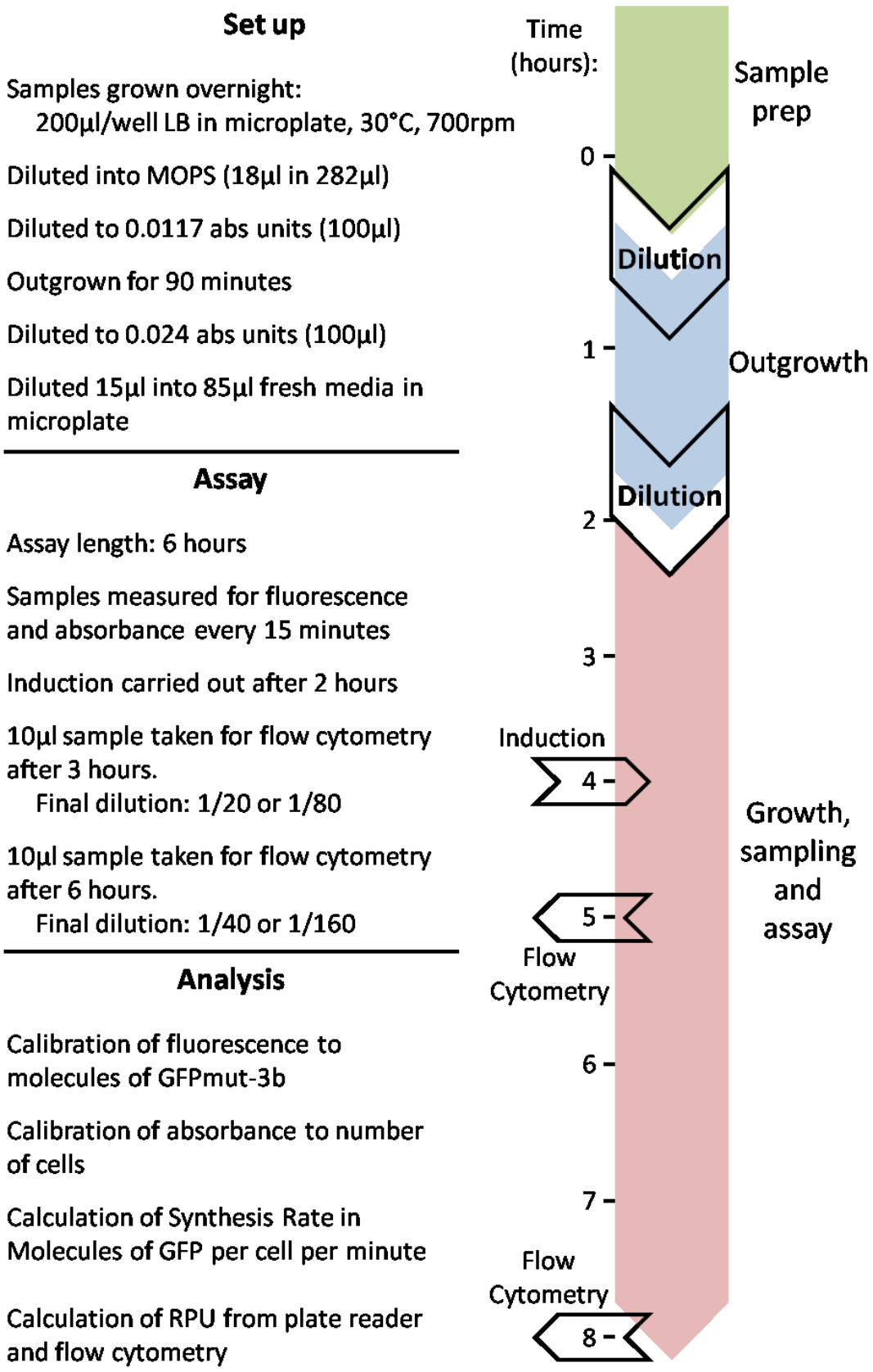


Figure 6.16 Overview of the automated standardised characterisation workflow.

The automated characterisation assay increased the reliability of characterisation and produced data in a standardised format. The assay workflow is shown diagrammatically and the key steps at each stage should be followed to obtain consistent results under standard conditions. For more details see methods section 5.7.

6.4.2 Data standardisation and Metrology

6.4.2.1 Absolute GFP calibration

A key difficulty identified during the BioFAB Bit to Atom to Bit experiment (6.2.1) was the difficulty of transferring results with meaningful units. While fluorescence data was normalised the result was still an arbitrary value of signal observed by the equipment. In a separate lab or a separate experiment the value was often relatively meaningless. While Relative Promoter Units avoided some of these issues by fixing a reference point that was utilised by many labs the results were still on a fairly arbitrary scale measured against a relative point. There was also no clear indication of just how useful RPU were as units for modelling or BioCAD tools. One significant unit set was put forward by Canton et al.(2008) for an absolute scale that had seen little use so far across the field but was the ultimate absolute units for the modelling of these experiments; molecules of GFP (per cell). While difficult to convert results into these units are by far the most appropriate for reporting.

To allow conversion into absolute units the quantification of fluorescence inside *E. coli* MG1655 cells was required. Wild-type GFP fluorescence is heavily pH dependant to the point of GFP being suggested as a sensor for pH changes in a range around physiological (Kneen et al., 1998). GFPmut-3b is relatively stable at cytoplasmic pH ranges but fluorescence has been shown to drop rapidly in conditions with a pH below 6.5 (Wilks and Slonczewski, 2007). As such the pH of the expressed environment may have a small but present effect on the fluorescence observed. As it is difficult to detect the pH state inside an *E. coli* cell the only way to deal with this unknown variable was to remove it entirely by lysing the cells to bring the GFPmut-3b into an environment which can be recreated. A useful calibration from fluorescent signal into molecules of GFP would require understanding the relationship between the fluorescence signal produced by GFP in cells before and after lysis and the fluorescent signal produced by known amounts of GFP in cell lysate.

B-PER II lysis buffer was chosen to carry out the lysis as this gave consistent lysis of >95% of cells and was amenable to work in microplates. A set of SVc plasmids containing Anderson promoters with a range of expression levels in MG1655 were grown in 0.4% glucose rich MOPS medium to an OD between 0.5 and 1. These cultures were centrifuged and washed before resuspension in fresh MOPS media at an OD of 0.5 and then split into two aliquots. One aliquot was lysed with an equal volume of lysis buffer (containing protease inhibitors) while the other was diluted with an equal volume of PBS and 200µl volumes of each were measured for fluorescence on the Synergy HT plate reader used for characterisation. On the same plate dilutions of purified GFP

which had been quantified by a modified lowry kit were mixed with negative control cell and the fluorescence measured.

Lysis of *E. coli* caused a significant reduction in the fluorescence signal observed from the solution. Following removal of background auto-fluorescence signal (SVc lysate and culture) a clear correlation (r^2 = 0.99) between the fluorescence signal before and after lysis was observed. The relationship was equal to an almost 40% drop in fluorescence signal regardless of pre-lysis fluorescence down to approximately 500 arbitrary fluorescence units (figure 6.17). This deviation from the consistent drop in fluorescence was for a signal below the level of auto-fluorescence observed for media in all experiments and so should not affect calibrations.

There was a strong correlation between the number of molecules of GFP diluted in cell lysate and the signal observed following background correction (figure 6.18). A linear approximation without intercept was used to describe this relationship because there should be no fluorescence following background correction at an Fl value of 0 and the fit suggested by fitting in Matlab resulted in no fluorescence signal equalling a large number of GFP molecules. As the fluorescence level where deviation occurred was so small it was considered than any inaccuracy would be unlikely to affect results.

It was now possible to convert fluorescence results obtained on the characterisation robot's Synergy HT plate reader into absolute units of molecules of GFP.

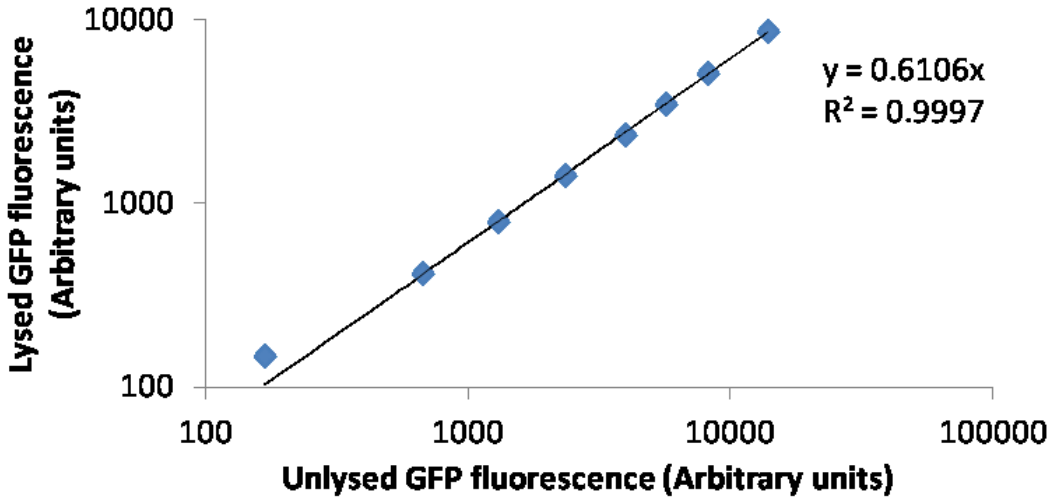


Figure 6.17 Relationship between fluorescence signals from lysed and unlysed MG1655 *E. coli*.

GFPmut-3b is a modified GFP that remains sensitive to pH and other environmental conditions which may alter its fluorescence. Lysing cells to release the GFP into assayable conditions altered the fluorescence produced by the protein. Following fluorescence background correction the effect is similar to an almost 40% drop in fluorescence on samples lysed compared to those left intact. All points are the average of three repeats and the error of one standard deviation is too small to be seen. The scale is logarithmic and so low values have little effect on the fit.

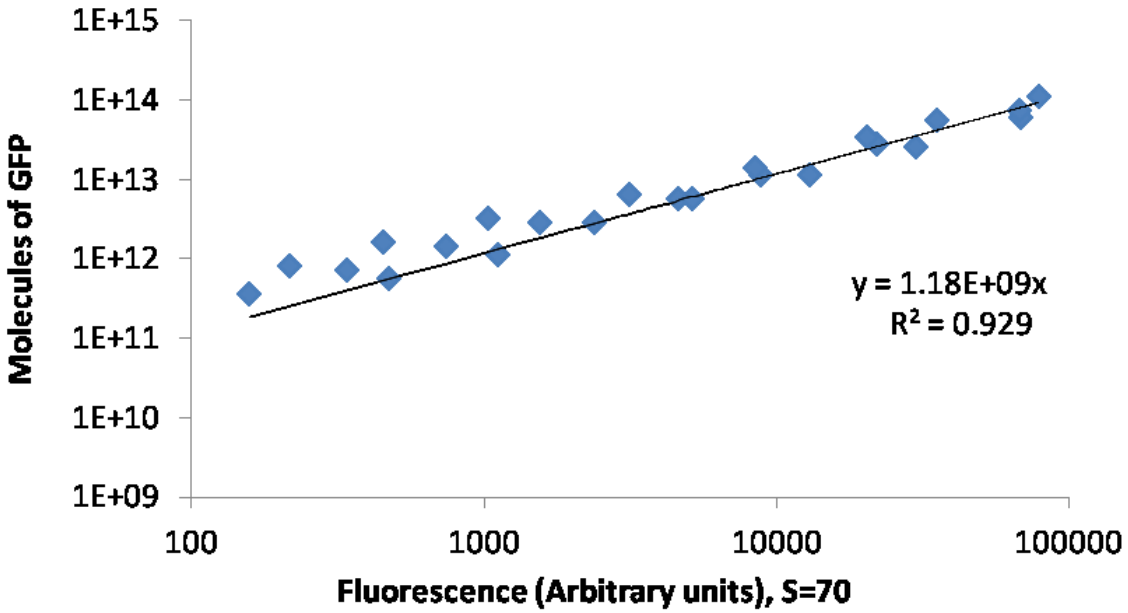


Figure 6.18 Fluorescence signal produced by purified GFPmut-3b protein diluted in cell lysate.

To complete the quantitation of *in vivo* GFPmut-3b the fluorescence produced by a known quantity of purified GFPmut-3b in cells lysate was determined. There was a clear relationship between the amount of purified GFPmut-3b protein and the fluorescence produced by those molecules following removal of background fluorescence. The relationship was linear until around 1000 fluorescence units. Results were pooled from 3 separate experiments. The scale is logarithmic and low values have little effect on the fit.

6.4.2.2 Fluorescein standardisation and calibration

While the equipment used with the characterisation platform can now have its measurements converted to units of molecules of GFPmut-3b previous use of this unit had not seen significant uptake by the field of Synthetic Biology. It was considered that a more convenient standardisation method would be advantageous for the field. A suitable standard was chosen in the chemical fluorophore sodium fluorescein which has an excitation and emission spectra similar to most green fluorescent proteins. Sodium fluorescein was readily available from chemical companies, was reasonably, could be stored as powder for long periods of time and gave consistent results following storage as long as it is kept away from sources of light.

Starting with an initial stock solution diluted in PBS (pH 7.35) dilutions were carried out until a solution with a fluorescent signal was observed that could be detected without saturation the plate readers PMT on the standard sensitivity settings. Stock solutions of 4µg/ml, 2µg/ml, 1µg/ml and 0.5µg/ml were made up in PBS and serial dilutions performed to generate a range of 19 fluorescein concentrations which were loaded onto a microplate and scanned at the standard sensitivities.

The resulting fluorescence values were background corrected using PBS to reveal the linear correlation between fluorescein concentration and fluorescence signal observed as shown in figure 6.19. The fluorescein standard data was used to allow conversion of data collected using a PMT sensitivity setting to values that would be observed at another and the same principle should apply across equipment at many institutions. The fluorescein standard data was also combined with the GFPmut-3b molecules calibration to allow a conversion to molecules of GFPmut-3b using the following equation (after auto-fluorescence subtraction):

$$1ng\ of\ sodium\ fluorescein\ in\ PBS\ \approx 2.01 \times 10^{12}\ molecules\ of\ GFPmut-3b\ (inside\ E.coli)$$

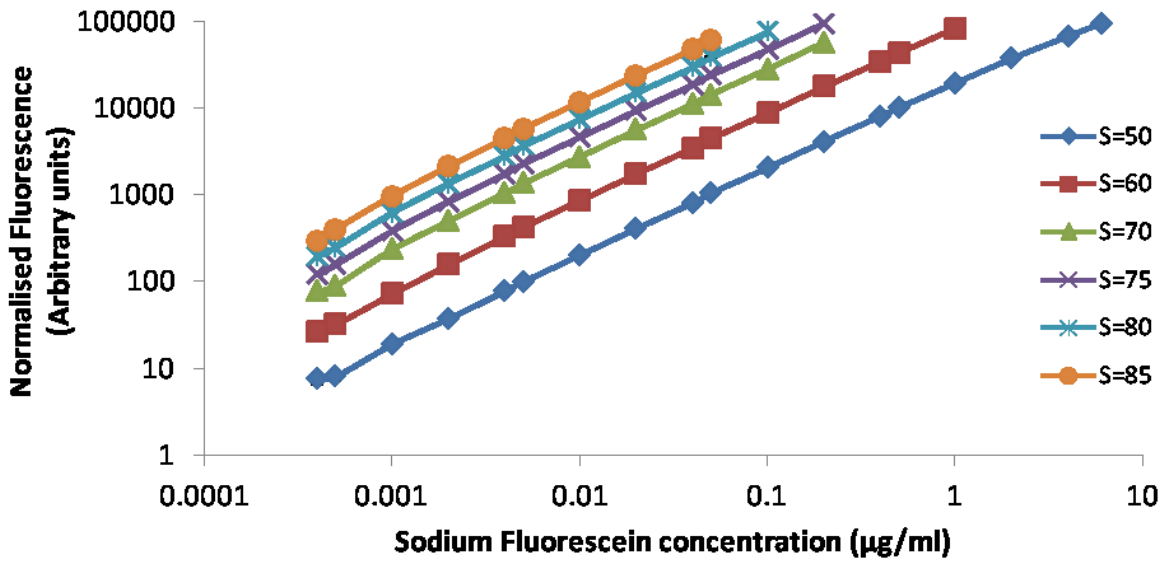


Figure 6.19 Fluorescence signal for a series of sodium fluorescein concentrations at various PMT sensitivities.

Sodium Fluorescein was chosen as a simple standardisation device for equipment between labs. The fluorescence signal for 100µl of a range of fluorescein concentrations was determined over a range of standard PMT sensitivities. Following background fluorescence removal there was a linear relationship between concentration and fluorescence signal. The readings were consistent across several repeat wells (single standard deviation error bars are present but are smaller than the point markers). Data points displayed were the mean of multiple wells.

6.4.2.3 OD to cell number calibration

While optical density is a measure of the absorption of light over a given path length it was used in assays to estimate the size of a bacterial population. While this was fairly appropriate for normalising fluorescence to account for population size it was not accurate enough for use in modelling or design and optical density is known to vary between pieces of equipment. To complete the conversion of characterisation data to units more suitable for an absolute standard an accurate calibration for the number of cells was required.

MG1655 *E. coli* carrying the SVc-101 plasmid were grown in 0.4% glucose EZ MOPS media in microplate at 30°C and 700rpm to mimic a standard characterisation experiment. Samples were grown for 6 hours and routinely scanned for absorbance every 15 minute using the characterisation platform's plate reader. Samples were taken every 45 minutes and diluted in water and kept on ice. These samples were then diluted to a level appropriate for assay and mixed with Accucount cell counting beads and run on the flow cytometer through the Automated Microplate Sampler (AMS).

Following data acquisition events in the flow cytometry files were initially gated in the FSC and SSC domains to remove the bulk of noise events before a second round of gating in the SSC and FL1 domains served to remove any remaining noise. The resulting events were gated according to their red to separate beads from cells. The number of cellular events relative to the number of bead events was calculated and along with the number of beads per µl and the dilution factor used to calculate the number of cells in the 100µl sample scanned for absorbance. This absorbance signal was converted to optical density by use of an absorbance to optical density conversion curve previously generated using a spectrophotometer. The resulting fit is shown in figure 6.20.

A highly accurate correlation between absorbance and cell number was obtained for samples in assay plates. This particular fit was used for calibration of automated characterisation data as any plate reader.

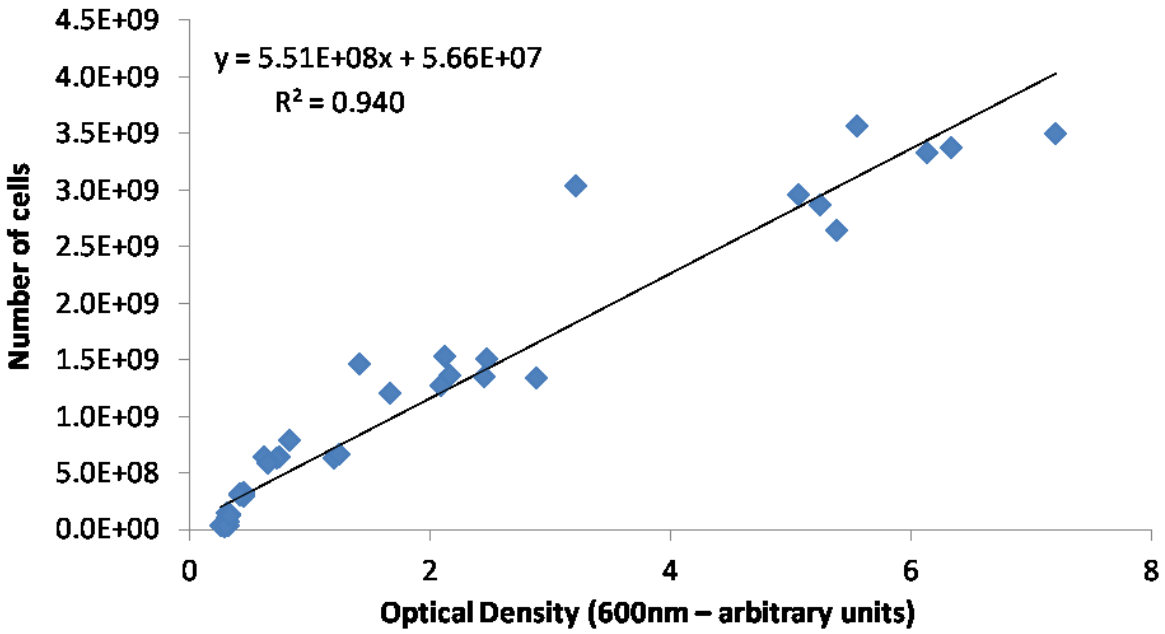


Figure 6.20 Relationship between optical density and number of *E. coli* MG1655 cells in a microplate well.

The number of cells was the absolute unit for a bacterial population but was generally roughly estimated by the absorbance or optical density of a solution. The number of cells grown under characterisation conditions at a range of absorbance values was counted using a flow cytometer and cell counting beads. Following the conversion of absorbance to optical density the numbers obtained were used to determine the relationship between optical density and the number of cells for *E. coli* MG1655 grown in EZ rich MOPS media at 30°C and 700rpm. Data includes pooled results for n=8 samples.

6.4.3 Characterisation of the constitutive Anderson promoter collection

The standard procedure had been developed and was ready for use so an experiment to test the throughput of the automated protocol was called for. Constitutive promoters were chosen as they are simple always expressed BioParts and so would be ideal for a test of throughput. Given the number of samples which can be housed in biological triplicate on a microplate only a library of constitutive promoters would be large enough to push the limit of the automated platform in terms of throughput. There were a few libraries available most notably the library constructed by Alper et al. (2005), some of the libraries built by Mutalik et al. (2013) and the Anderson promoter collection in the Parts Registry (n.d). At this time the Alper and Mutalik libraries were already well characterised whereas the third library had a few members well characterised but the others have very little in the way of characterisation. The Anderson library contained promoters with fairly well defined transcription start sites and has been widely distributed so was considered the ideal testing library

Promoters from the Anderson collection were ordered as oligos that would form EcoRI-SpeI cut Biobrick fragments when annealed and were restriction cloned into plasmid SVc. During this process 2 additional promoter were generated (J23114a and b) and retained to be part of a slightly expanded library. MG1655 were transformed with SVc plasmids carrying all the library members for use in characterisation. Single colonies were arrayed into a microplate and grown overnight in LB media for characterisation using the standard settings. In total 22 promoter containing plasmids and the negative control plasmid were set off with 72 samples loaded per day for assay (n=3 per day). The following day all of the samples were assayed according to the standard protocol in 0.4% glucose EZ MOPS media with conditions of 30°C and 700rpm for 6 hours. Measurements were carried out every 15 minutes as standard and samples were taken for analysis by flow cytometer after 3 and 6 hours using the AMS attachment. This was repeated for a further two days to bring the total number of samples to n=9 for 21 promoters, the reference promoter and the negative control plasmid.

The automated platform carried out the protocol as expected generating a huge amount of data. The data for one promoter (J23111) from the library characterised by the platform is shown as an example in figure 6.21. The data for each promoter comprised 9 fluorescence traces for the number of molecules in each sample well for the duration of the characterisation (figure 6.21 A), 9 cell number per well traces for the same promoter (6.21 B) and 18 fluorescence distributions

produced by the flow cytometer of which 2 are shown in figure 6.21 C. These traces allowed the calculation of synthesis rate across the entire growth curve of *E. coli*.

Taking measurements across the growth curve in this way allowed the observation that the output from the promoters changed during the growth as can be seen in the change in RPU results for the promoters between 180 minutes and 360 minutes in figure 6.22. This increase in RPU strength may have been a result of the output of the J23101 reference promoter itself dropping reducing compared to other promoters but indicated the importance of acquiring data across the culture's growth profile. The Anderson library contained promoters with a variety of strengths including members with and RPU output of almost 2.5 down to output of approximately 0.02. There was a slight a bias in the library towards lower output levels (13/22 had an RPU output of 0.5 or less). The RPU output level for each promoter in the library at 180m was consistent when measured by both the plate reader and flow cytometer, with a Pearson's coefficient of 0.996 and r^2 value of 0.993 when plate reader RPU outputs greater than 0 were compared to flow cytometry results (figure 6.22 A). At 360 minutes the flow cytometer and plate reader results were fairly consistent again although slightly more so in terms of rank as the Pearson's score remained the same but the r^2 dropped a little to 0.991. This seems to have been caused by the flow cytometer estimating higher outputs for most promoters (relative to J23101) compared to plate reader results (figure 6.22B) and the cause for this was unknown.

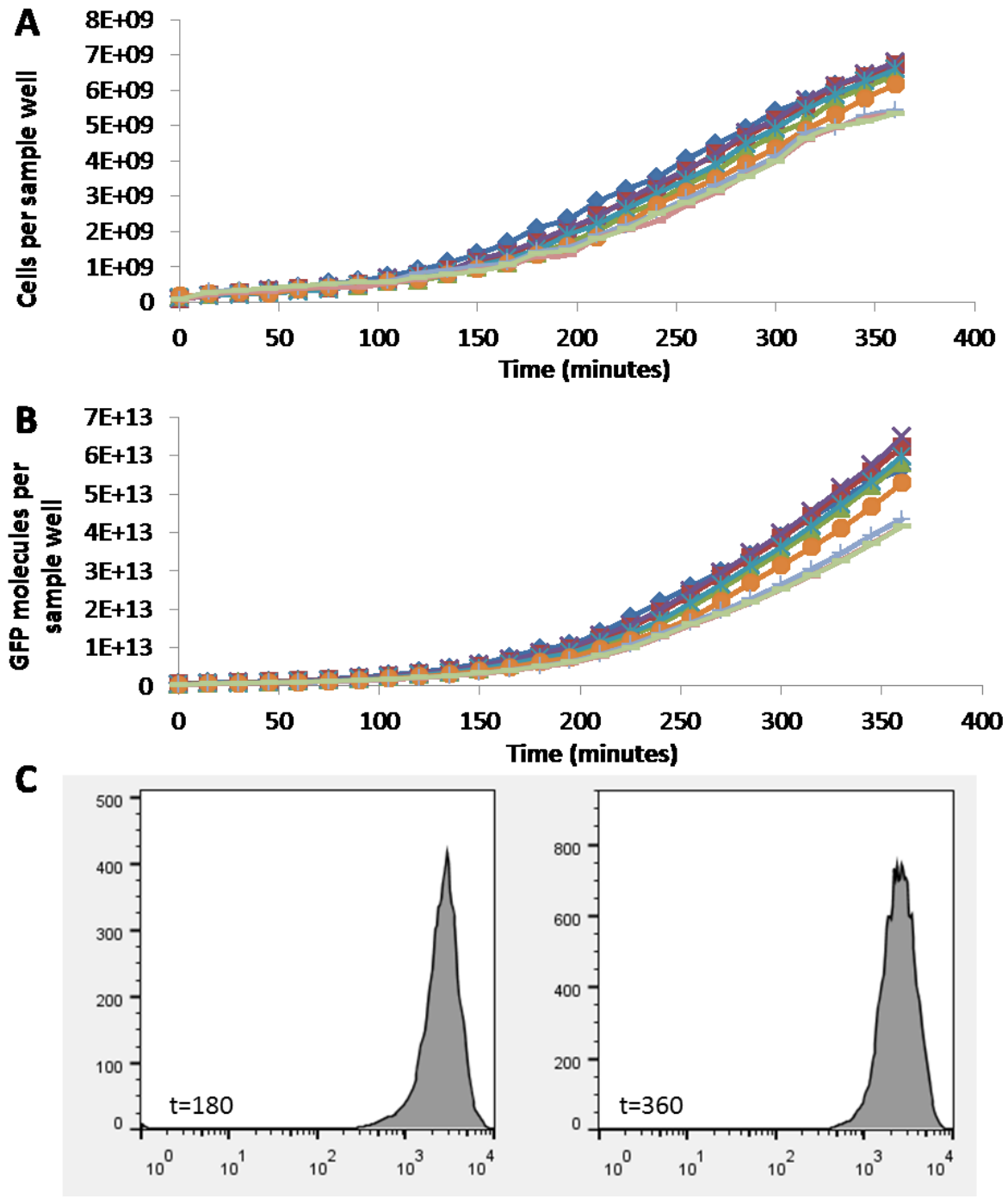


Figure 6.21 Data output for a single constitutive promoter (J23111) characterised using the standard workflow.

The automated platform and characterisation workflow produced a large amount of data for all characterised BioParts. The raw results included 9 cell number over time traces (A) which were used to accurately calculate growth rate and normalise populations, 9 fluorescence traces in molecules of GFPmut-3b (B) which allowed the detailed calculation of BioPart output over time and 18 flow distributions evenly split between the middle of logarithmic growth and during late logarithmic/early stationary growth – with an example of each shown (C).

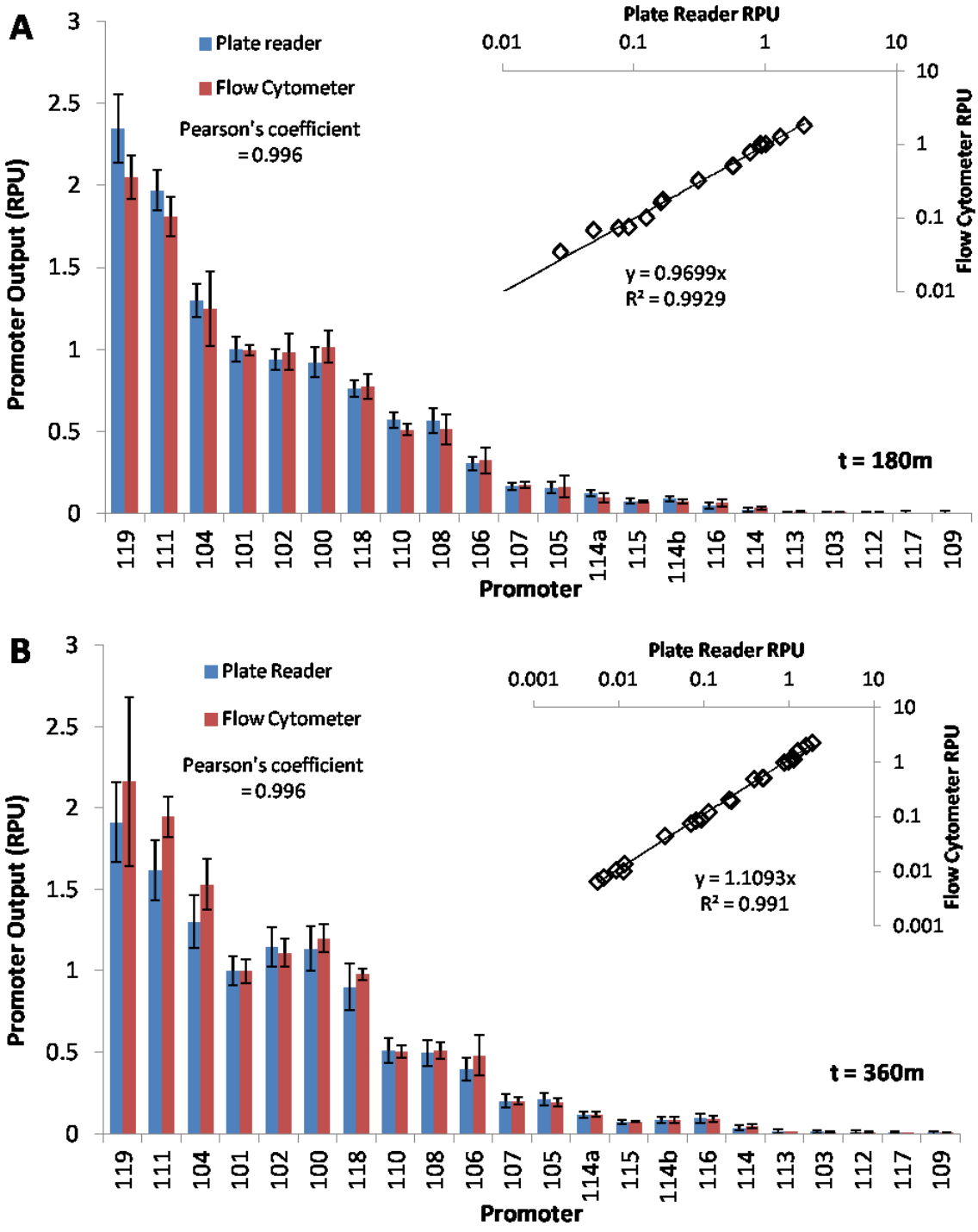


Figure 6.22 Characterisation results for the full Anderson constitutive promoter library (J23XXX series).

Constitutive promoters we chosen as they are simple systems and so have a constant output. The entire 20 member Anderson promoter library and 2 additional members were characterised and the average RPU and the standard deviation in observed RPUs from both plate reader and flow cytometry data calculated for both mid logarithmic and late logarithmic growth. The results were arranged in decreasing order of plate reader RPU output for the 180m time point with data from mid logarithmic growth displayed in A and late logarithmic growth in B. Results were displayed as columns for ease of comparison between the data types in terms of output and error. The plate reader and flow cytometer results were also plotted against each other in the top right of graph revealing a strong correlation between the flow cytometry and plate reader results that had not been seen to this degree before. The data suggested that the strength of the promoters may change during the course of a culture's growth. For all results n=9. Values are the mean and the error a single standard deviation

6.4.4 Impact of Carbon Source on characterisation results

The media context under which characterisation was carried out could potentially alter the results either by directly interacting with the BioParts or by changing chassis properties such as growth rate. It had been suggested that one of the benefits of RPU would be to remove these effects from characterisation results (Kelly et al., 2009). The MOPS media used for standard characterisation is both completely defined and also relatively modular and so made an ideal tool for testing the impact of components upon BioPart activity. The simplest component to switch was the carbon source and as this is generally used by a cellular chassis as fuel it made an ideal candidate to test how well RPU could cope with a change in conditions. Additionally it would be important to know if changing the carbon source could affect a BioPart characterisation in case a carbon source was used as an induce in future.

To test RPU as a method for removing condition context affects on characterisation 2 microplates were assayed simultaneously as this allowed multiple BioParts to be characterised in triplicate under a reasonable number of conditions. 4 promoters with outputs between approximately 0.1 and 2 RPU were selected as suitable to allow testing of 8 different carbon sources. The automated characterisation protocol was modified to effectively run two characterisations simultaneously with 4 characterisation conditions tested on each plate and a time shift of approximately 8 minutes. The temperature the assay was carried out was increased to 37°C to reduce the chance of samples not reaching suitable population sizes during the assay while conditions of 700rpm shaking, measurements every 15 minutes and 0.4% carbon source EZ MOPS media were kept the same.

The 8 carbon sources chosen were glucose, xylose, glycerol, rhamnose, arabinose, sucrose, mannose and maltose as these should give a suitable range of growth rates while also including carbon sources likely to be useful as inducers (xylose, rhamnose and arabinose).The flow cytometry sampling times were moved to 2 hours and 4 hours into the characterisation assay to reflect the changed temperature. The characterisation was carried out and the data analysed as normal with growth rates calculated using the Growth Rates programme written by Hall et al. (2014).

For each promoter the output in molecules of GFP per cell per minute varied depending upon the carbon source used (figure 6.23 A). Interestingly glucose was the poorest carbon source in terms of absolute GFP production. The flow cytometry data taken at 120 minutes was initially used to test the hypothesis that RPU would be consistent across carbon sources. While the results in figure 6.23 B suggested that the expression of results in RPU were the similar in all conditions, when each

promoter's results were tested for significance at 5% by ANOVA the results indicated that the RPUs were likely to be separate (F scores>F crit of 3.44>2.16, 2.51>2.16, 2.56>2.16 and 5.03>2.15 for J23106, J23110, J23111 and J23115 respectively). Kelly et al. (2009) had suggested that RPU output for a promoter should ideally be calculated from the mid-logarithmic portion of bacterial growth and the timing of this was observed to vary across the 8 carbon sources. When the output in RPU for each promoter was calculated from the plate reader data during the mid-logarithmic growth phase of each condition the results (figure 6.23 C) again suggested that RPU was relatively consistent across all conditions. When ANOVA analysis was carried out on these results the differences in RPU output were found to be not statistically significant at 5% for J23106, J23110, J23115 (F scores < F crit of 0.947<2.16, 0.474<2.16 and 1.08<2.17). The RPU differences for J23111 across all 8 carbon sources were statistically significant (F score of 2.95 and an F crit of 2.16) but when results for maltose and mannose were excluded the results again indicated that RPU values were not significantly different (1.25<2.41).

While the output of the promoters could be expressed in RPU to remove the effect of the carbon source the effect itself was potentially interesting. Unfortunately J23115 had to be excluded from this analysis as many results possessed standard deviations larger than the observed synthesis rate. For the other promoters there appeared to be a relatively common change in absolute output for all the promoter BioParts compared to the absolute output for each promoter in 0.4% glucose media (figure 6.24 A). The hypothesis that the increase in output was proportionally the same for each promoter for a given carbon source was again tested by ANOVA and shown to be not statistically significant at 5% for all but glycerol and rhamnose (F score > F crit of 6.48>2.90 and 4.21>2.90 respectively). These results may be useful as a rough guide for how much a BioParts absolute characteristics may change in certain carbon sources.

The hypothesis that the growth rate was responsible for the change in absolute output was also tested but this only indicated weak correlation between growth rate and GFP production (figure 6.24 B) and only if glucose was excluded (r^2=0.615 v r^2=0.061).

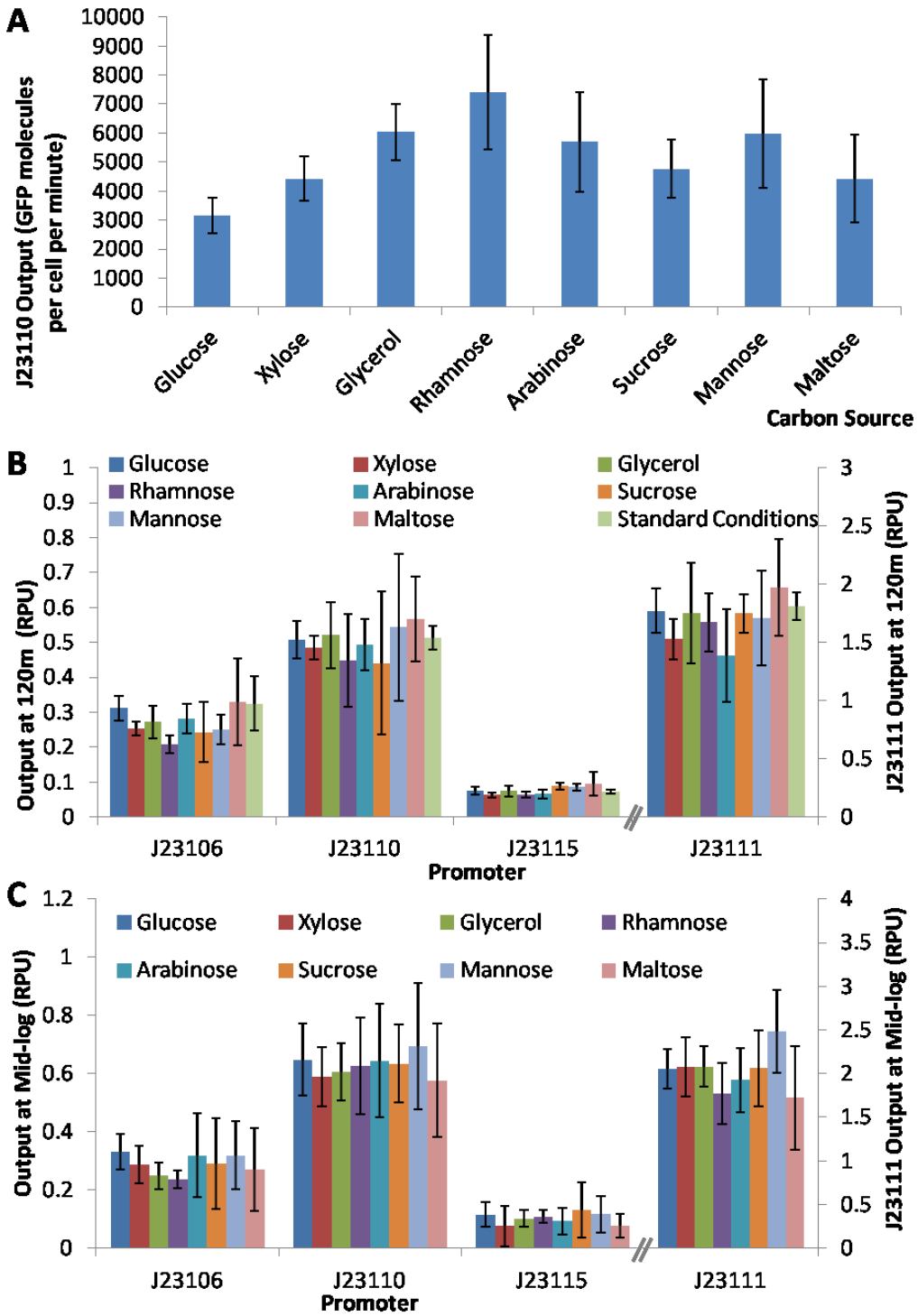


Figure 6.23 Expression of characterisation results in RPU can remove the effect of media context.

The media in which BioParts are characterised can alter their observed characteristics. The GFP expression results in absolute units for J23110 were dependent upon the carbon source of the media during characterisation (A). When J23101 was used as a reference the characterisation results in RPU were more consistent but there was still variation between measurements taken at the same time point (B). These results were even more consistent when the synthesis rates at mid-log were used for calculation of output in RPU (C) rather than a single time point. For all results n=8 or 9. All results were means and the error is a single standard deviation.

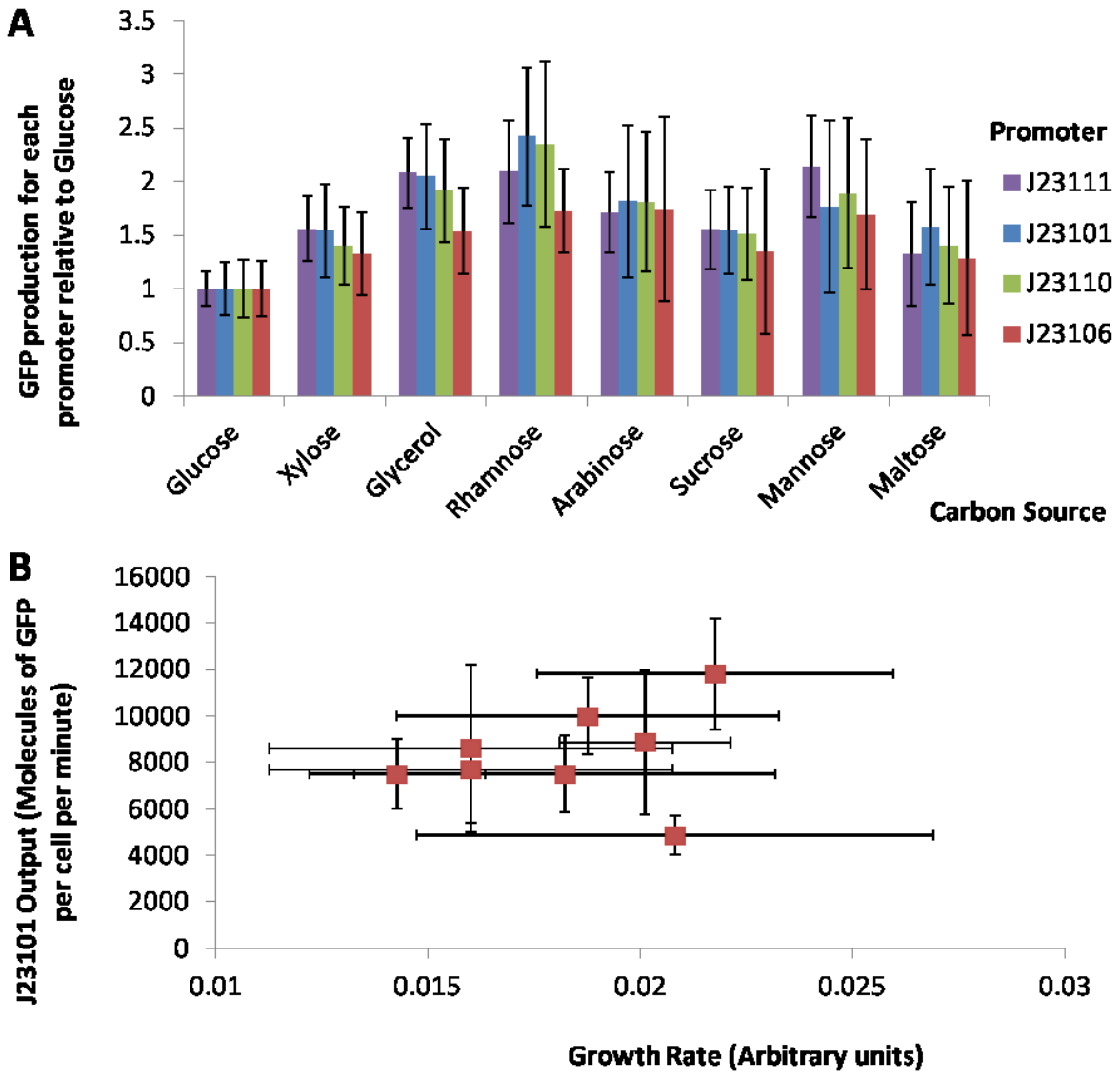


Figure 6.24 The absolute effect of carbon source change on GFP production and relationship to growth rate.

Changing the carbon source used for BioPart characterisation and system operation can alter the output of the BioPart. When the carbon source of the media was changed a reasonably consistent increase in GFP production from multiple different constitutive promoters was observed (A). The relationship between the expression of GFP and growth rate was unclear across the 8 carbon sources tested though there was weak correlation between the growth rate and absolute GFP output if the glucose was considered an outlier is excluded (B). For all results, n=8 or 9. Values are the mean for the samples and the error the sample deviation.

6.4.5 Characterisation of inducible promoters

6.4.5.1 The xylF xylose inducible promoter

Inducible promoters are more complicated BioParts which have been sued as control elements and so may be more interesting to circuit designers. The inducible automated protocol was designed to characterise an inducible promoter to a very high level of detail and pick up any characteristics which could be easily missed. Many induction systems had already been well used and tested by both Synthetic Biologists and the wider biology community and so a less well tested system was sought out. *E. coli* had a wide range of sensing systems for many metabolites which could make useful BioParts and the xylose responsive system based around the activator protein XylR and one of the promoters it regulates *xylF* was chosen for the first inducible characterisation.

The minimised *xylF* promoter with no CAP binding sites carried by the SVc-X plasmids was characterised with the *xylR* gene over expressed in the regulator module to produce large amounts of protein to activate the promoter. The negative control was the SVc-X plasmid and the reference plasmid was SVc-X-101 to ensure accurate results in RPU. LB cultures were set up for overnight growth at 30°C and 700rpm in microplates as before. The characterisation assay itself was carried out on the characterisation platform at 30°C and 700rpm using 0.5% glucose MOPS. The glucose in the MOPS media was increased to compensate for the large volume of liquid added during induction and the inducer was diluted in carbon source free MOPS to maintain the standard characterisation conditions. Induction was carried out with 24 concentrations of xylose in a range from 0mM to 300mM after 2 hours while the rest of the protocol remained the same (see methods for details).

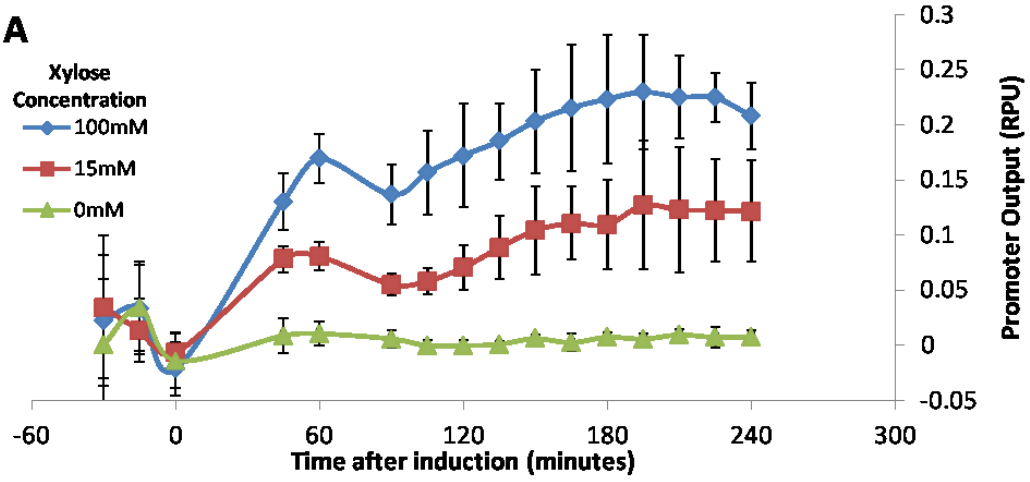
The key characterisation results that would be suited for a datasheet are shown in figure 6.25. At maximum output with the XylR protein over expressed *xylF* is a weak to medium strength output promoter only reaching an RPU of around 0.3 (figure 6.25C). While *xylF* was turned on almost immediately by the addition of xylose it can be seen in figure 6.25A that it took at least 45 minutes to reach full output for a given level of induction. The uninduced level of output was approximately 0 following induction (<0.02). The input-output relationship of the promoter is shown in figure 6.25 B where it can be observed that the promoter was induced by xylose starting from approximately 1mM. Promoter output increased with increasing xylose induction up to at least 150mM. At concentrations above 150mM the promoter may have reached maximum output but there are too few data points above this concentration to be clear.

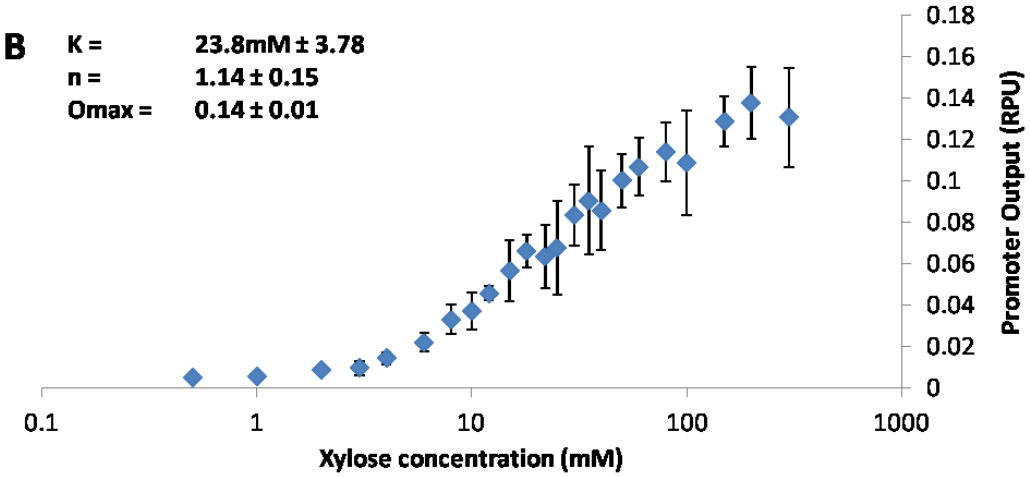
Also included in figure 6.25B are the parameters for a hill function fit to the data in Matlab. The hill function has been used to model activation (Canton et al., 2008) type processes in biology

and may be the most suitable model from activation without a specific XylR protein *xylF* promoter model being created. The parameters indicated low cooperativity in the XylR proteins and that the affinity governing the input-output function was approximately 23.8mM. This information was also included in a key results table in figure 6.25C containing the key characteristics.

Hidden within the data were also a subtle set of transient characteristics that could be missed by less rigorous characterisations. Figure 6.26 A and B displays the input-output relationship as observed in the plate reader data at 60 minutes and 105 minutes after induction. When the two input-output curves were compared there was a distinct hump to the curve at 60 minutes and this appeared to indicate that the time taken to reach full induction for the given concentration increased as the inducer concentration increased. Whether this was because it was related to the concentration or it simply too longer for more output to be achieved is unclear.

Interestingly when the hill equation was fit to these two datasets there was a significant difference in the n cooperativity and Km parameters with the fit in favour of regulator coopertivity but lower inducer affinity at 60 minutes before the trend is reversed by 105 minutes. The trend observed at 105 minutes was also evident in the data at 360 minutes (figure 6.26 C) indicating that this effect may be related to how quickly full homogenous expression is reached. A paper published by Ni et al. (2013) indicated that XylR acted as a dimer and exhibited cooperativity at much lower xylose concentrations however this dimer was shown to bind to the operator sites spread over both the *xylA* and *xylF* promoters simultaneously. As the *xylF* promoter characterised did not carry the *xylA* operators this would suggest induction of the *xylF* promoter was due to an unusual induction mechanism.





C

Key Characteristics			
Inducer	Xylose	**Inductive range**	Early: 1.5mM to 50mM
Response time	~60 minutes		Full: Above 1mM
Maximum output	~0.28 RPU	**On:off ratio**	~14
Basal Output	<0.02 RPU	**Growth Impact**	Negligible
Model/equation		**Parameters**	
$Ouput = \frac{O_{max}.[Xylose]^n}{K_m{}^n + [Xylose]^n}$		$Km = 23.8\,\text{mM} \pm 3.78$ $n = 1.14 \pm 0.15$ $O_{max} = 0.14\,RPU \pm 0.01$	

Figure 6.25 Characterisation results for the xylose inducible *xylF* promoter.

Inducible promoters were more complex and had an array of key characteristic which were documented in 2 figures and a key characteristics table. The first key characteristic was the dynamic response of the promoter (A) where the output in RPU for the *xylF* promoter is displayed for 3 inducer concentrations. The input-output relationship was detailed for the 180m time point from flow cytometry data in (B) with the promoter output for a given concentration of xylose and the parameters of the hill equation fit the data. All this data was summarised as key characteristics values in a quick reference table with the model and parameters included (C). For all results n=9. Data points shown were means and the error was one standard deviation.

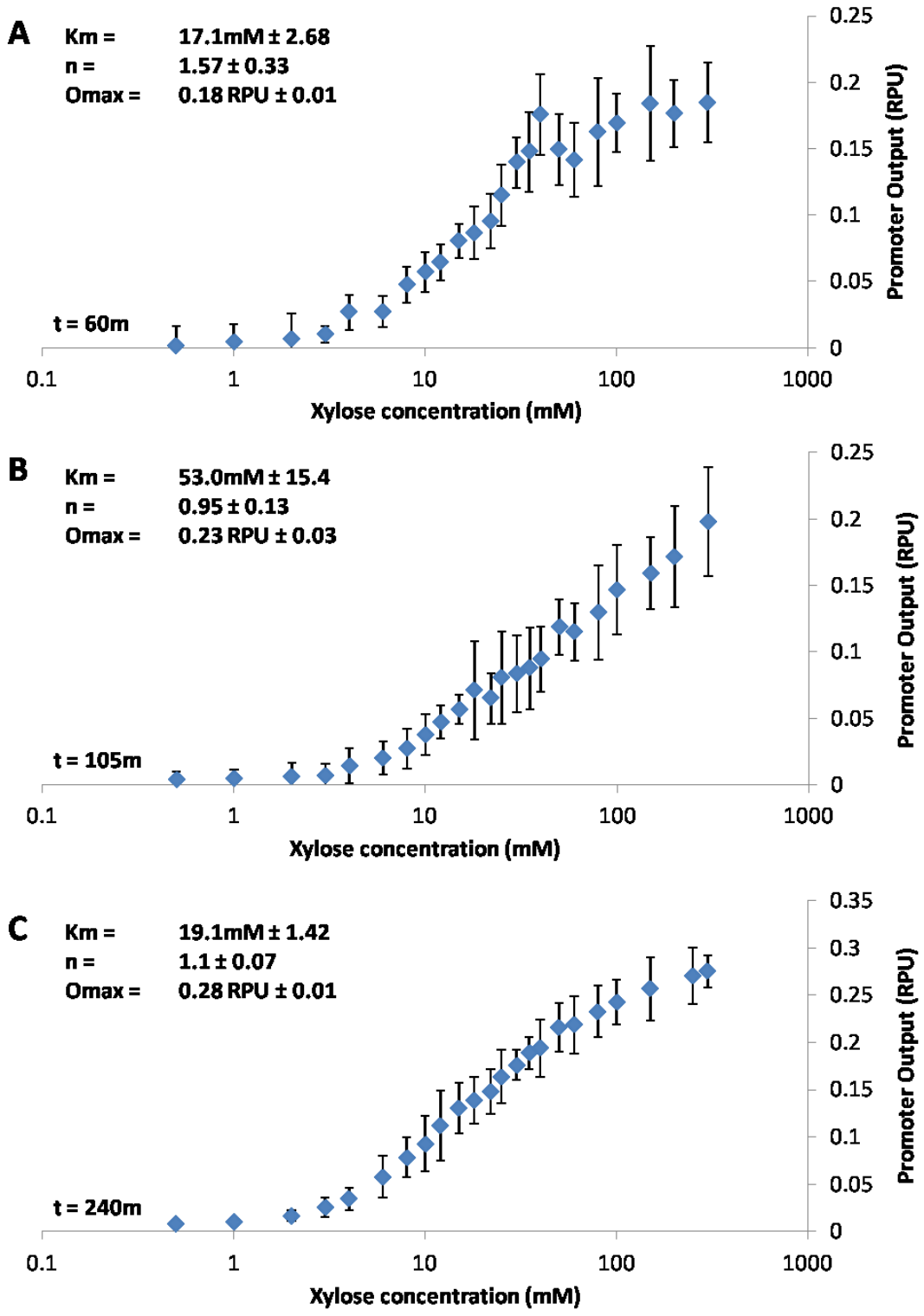


Figure 6.26 Dynamic changes in the input-output relationship of the *xylF* promoter following induction.

The input-output response was not uniform over time and the changes in it were observed in the plate reader data. The input-output graphs in (A) and (B) were taken from the plate reader data taken at the 60minute post induction timepoint and 105 minute post-induction time point and demonstrate how the relationship evolved over time. After 60 minutes induction promoter output in the range of approximately 50 to 150mM appears to be higher than expected from the flow cytometry data. By 105 minutes post-induction this effect had disappeared. By 360 minutes (C) the promoter output as measured by flow cytometry indicated that the input-output relationship may be approaching a hill function when the promoter was full activated for the concentration. For all data n=9. Data points shown were means and the error was one standard deviation.

6.4.5.2 Rhamnose inducible rhaB promoter

A second inducible promoter system was sought for characterisation and this time th rhamnose inducible RhaS protein and *rhaB* promoter system was chosen. The *rhaB* system was chosen because the *rhaBAD* promoter had previously been used a small nubmer of expression vectors but had not reached the level of use that similar systems such as *araBAD* have.

To allow characterisation of the *rhaB* promoter the *rhaS* gene was amplified from the MG1655 genome and transferred into the standard regulator cassette creating SVc-S. The *rhaB* promoter was also amplified from the genome and had its CAP binding sites removed prior to transfer into SVc-S. A reference vector containing the J23101 promoter and *rhaS* gene was generated for use in characterisation. The RBS in front of the *rhaS* gene in all 3 plasmids was converted to a new design before characterisation. The characterisation protocol was carried out exactly as for the *xylF* characterisation with growth in 0.5% glucose MOPs medium at 30°C and 700rpm with induction at 2 hours and rhamnose added to give concentrations in the range of 0mM to 250mM.

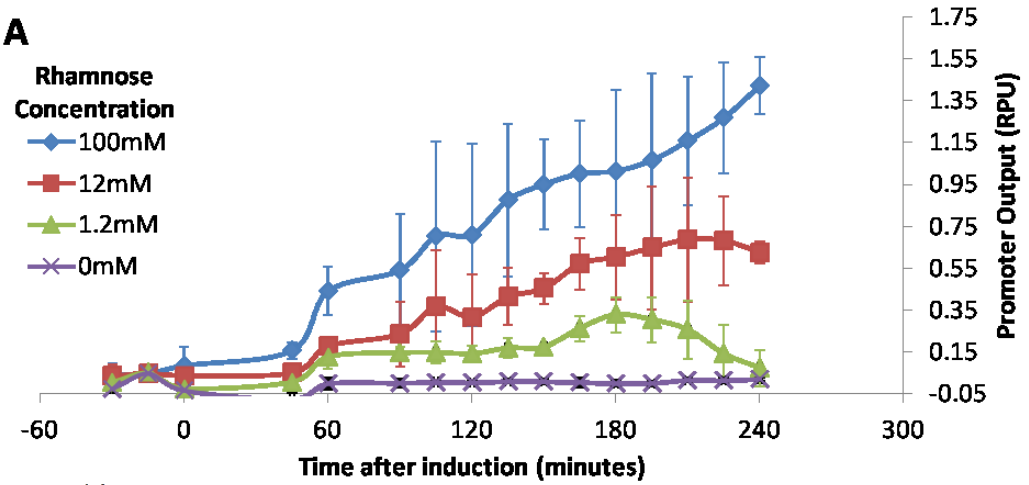
While also being an activated promoter the *rhaB* promoter has radically different characteristics to the *xylF* promoter. Unlike *xylF* which reach a stable maximum output level the output from *rhaB* under the highest inducer concentration increased for the entire duration of the characterisation. This was not however uniform for all inducer concentrations as while the high induction conditions demonstrated increasing levels of output many middle range concentrations reached a plateau after a period of time and many low concentrations began to reduce output towards the end of the characterisation (figure 6.27A). This effect was also observed in the flow cytometry data with rhamnose concentrations which had triggered promoter output at 180 minutes no longer demonstrating promoter activity by the 360 minute flow cytometry samples (figure6.28 A and figure 6.27 B respectively). This drop in production from low induction concentrations suggested that rhamnose may have been metabolised by the chassis.

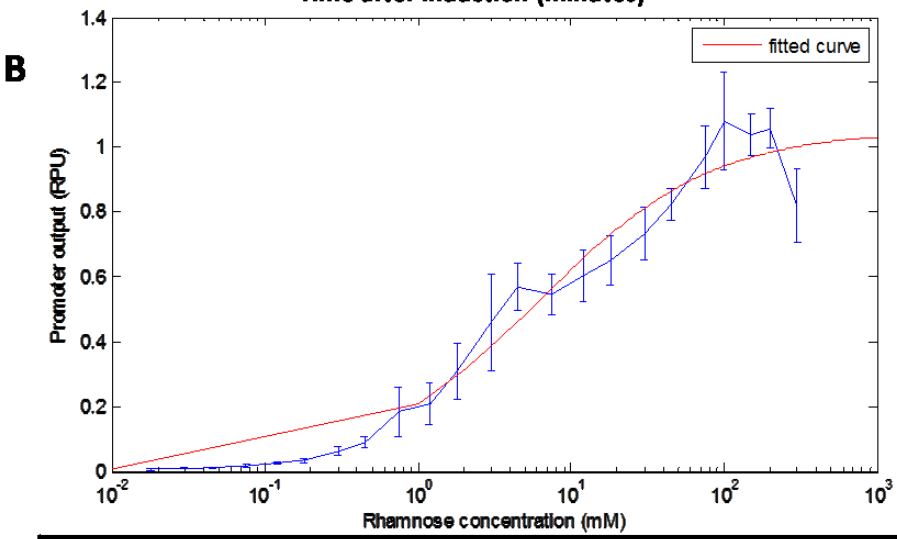
The input-output relationship for this promoter at the 360 minute time point indicated which concentrations maintained output for a significant period of time (figure 6.27B). The hill function fitting for the 360 minute time point was preferred for this promoter because at the 180 minute time point the hill function struggled to fit the curve (figure 6.27A).

The fluorescence distributions of the induced samples when assayed by flow cytometry data indicated that the cells within the samples were not very homogenous following induction. While the coefficient of variation for a highly induced sample was only in the region of 50% the same

coefficient for the low induction concentrations reached in excess of 250% (figure 6.28B). The cause of this homogeneity is unknown and may be caused by some level of interaction with the inducer or host.

While the *rhaB* promoter exhibited a lot of curious behaviour it was capable of reaching a very high level of output compared to the *xylF* promoter and had a much higher on:off ratio which may make it useful in certain circumstances (figure 6.27 C).





C

Key Characteristics			
Inducer	Rhamnose	**Inductive range**	Early: 0.1mM to ~75mM
Response time	N/A – Output increase for high induction		Rhamnose is consumed over time
Maximum output	1.41 RPU	**On:off ratio**	Up to 60
Basal Output	~0.02 RPU	**Growth Impact**	Above 200mM
Model/equation		**Parameters**	
$Ouput = \frac{O_{\max}.[Rhamnose]^n}{K_m{}^n + [Rhamnose]^n}$		$Km = 6.19\ \text{mM} \pm 3.30$ $n = 0.77 \pm 0.221$ $O_{\max} = 1.05\ RPU \pm 0.131$	

Figure 6.27 Characterisation results for the rhamnose inducible *rhaB* promoter.

The *rhaB* promoter had particularly interesting characteristics with different dynamic profiles generated by different rhamnose induction concentrations (A). This was reflected in the input-output relationship where by the end of the characterisation there was no induction where less approximately 0.1mM rhamnose had been used (B). Some of the key characteristics for this promoter were very good however with it possessing a high maximum output and low basal level (C). For all results, n=6 or 9. Data points shown were means and the error was one standard deviation.

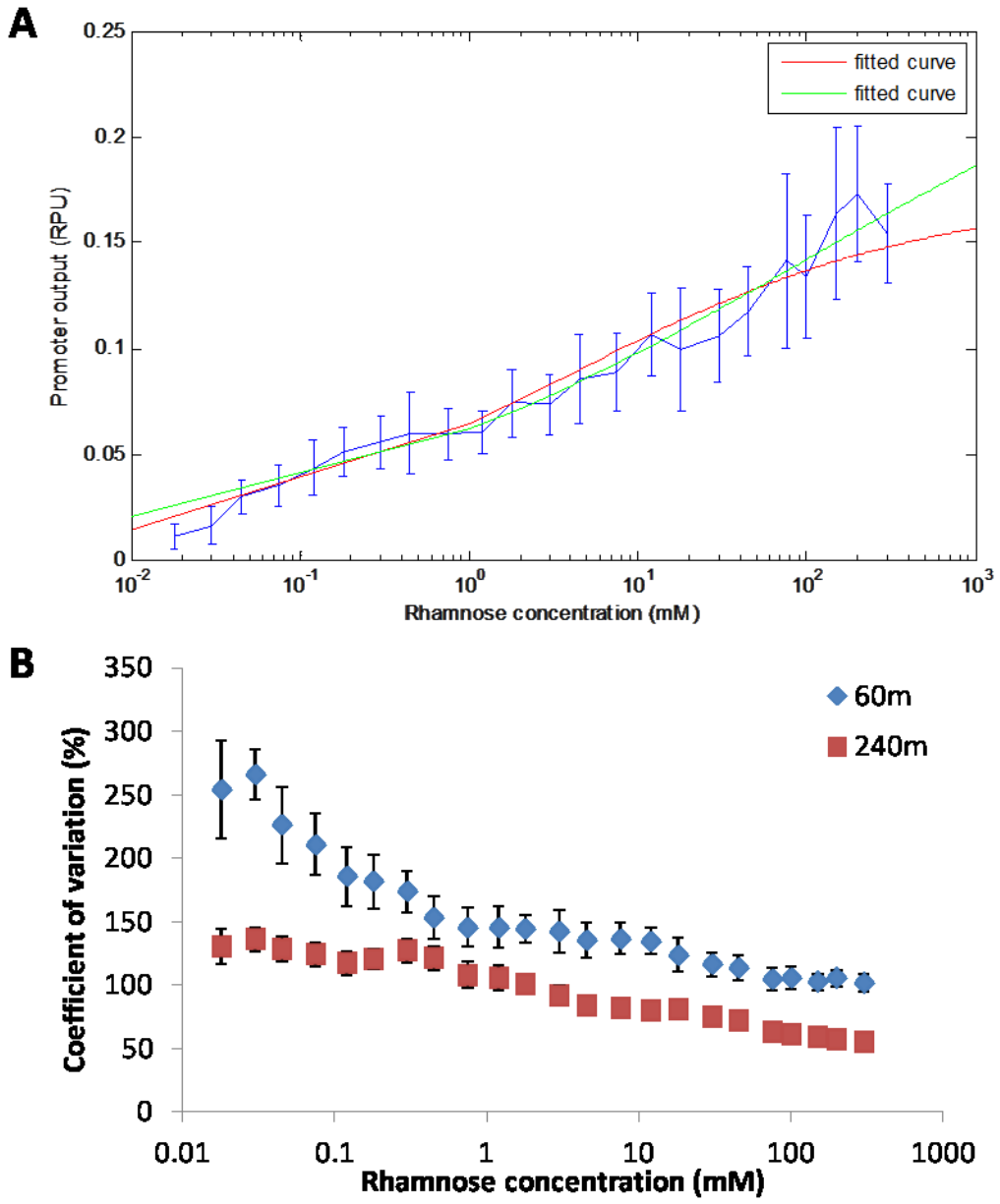


Figure 6.28 Flow cytometry reveals hidden details of the response of *rhaB* to induction.

The flow cytometry results for the *rhaB* promoter indicated some of its more unusual behaviour. The input-output curve at 180 minutes (A) was significantly different to the one observed 3 hours later with a much more even response to inducer that did not fit the hill equation well (fit lines with no limits (green) or Omax limited to maximum RPU (red)). There was also a huge variation in fluorescence distributions for *rhaB* (B). This was particularly true for the 180 minute flow cytometry sampling point and at lower induction concentrations where the variation in fluorescence was more than double the average fluorescence. For all results n=9. Data points shown were means and the error was one standard deviation.

6.5 Characterisation of synthetically generated inducible promoters

While inducible promoters were characterised this was a much lower throughput process that the characterisation of constitutive promoters. In order to characterise more inducible promoters more quickly a promoter engineering approach was undertaken to determine whether it would be possible to forward engineer activation based inducible promoters. Previous experiments by Cox et al. (2007) had demonstrated that it was possible to add operator sites for an activator to a core promoter sequence to generate a new promoter that could respond to the inducer. Oddly they had determined that while it was possible to predict in advance the properties of repressible promoters based on the promoter sequence and the added operator this had not been possible with activation. For the promoters they engineered they had started with a set of high strength promoter cores and then seemed to hit an activation threshold beyond which the promoters could be induced no further.

From the characterisation of the Anderson promoter collection there was data for a large set of fairly weak promoters available for engineering into novel synthetic activation based inducible promoters. For the promoter engineering a set of operators was required that would ideally give strong consistent output when bound by the associated regulator. The binding sites from *xylF* promoter would be an ideal choice as the induction using these sites had been consistent over a useful range and the original *xylF* promoter itself was very weak. To generate libraries of xylose inducible promoters 8 weak to medium constitutive promoters were chosen from the Anderson collection exhibiting outputs between 0 and 0.4 RPUs (figure 6.29 A). The sequence of the *xylF* and *xylA* promoters was compared to the Anderson promoters to identify where best to placement the operators relative to the promoter sequence. The operators and constitutive promoters overlap by 3 bases pairs (figure 6.29 A) and for these sequences for the promoters were considered to be over-ruling and was thus retained.

The Anderson promoters were transferred from SVc into SVc-X and the upstream sequences were added to generate 3 sets of 8 promoters each with a different arrangement of operators (figure6.29 B). Set X1 had the operators from *xylF*, set X2 had and a hybrid of the two with operator 1 from *xylA* and operator 2 from *xylF* and set X3 had the operators from *xylA*. Following assembly each set was tested in turn.

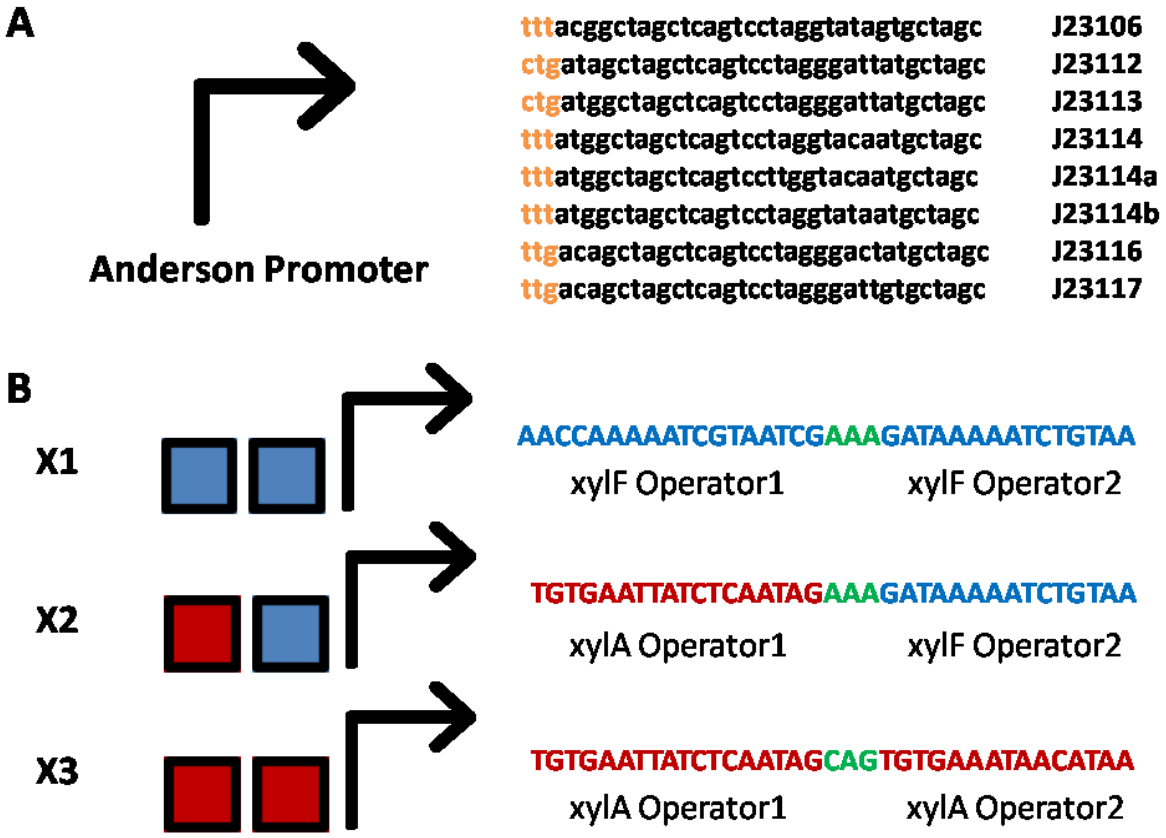


Figure 6.29 Generation of synthetic xylose inducible promoters by fusion of xylR binding sites to 8 weak Anderson promoters.

3 sets of potential xylose inducible promoters were generated based upon combinations of the *xylF* and *xylA* binding sites. Coloured sequences: orange is the operator/promoter overlap, green is the gap between operators, blue a *xylF* operator sequence and red a *xylA* operator sequence.

The last 3 base pairs of the second operator overlap the first 3 base pairs of the constitutive promoters. For both the *xylA* and *xylF* operators this sequence was ttg which matched 2 of the 8 constitutive sequences perfectly and 2 out of 3 bases for 4 others. Where the promoter did not match the promoter sequence was retained to retain consistency for the forward engineering.

The promoters were named beginning with their BioBrick code and finally the designator of the operator set, for example; J23114 with set X1 became 114X1.

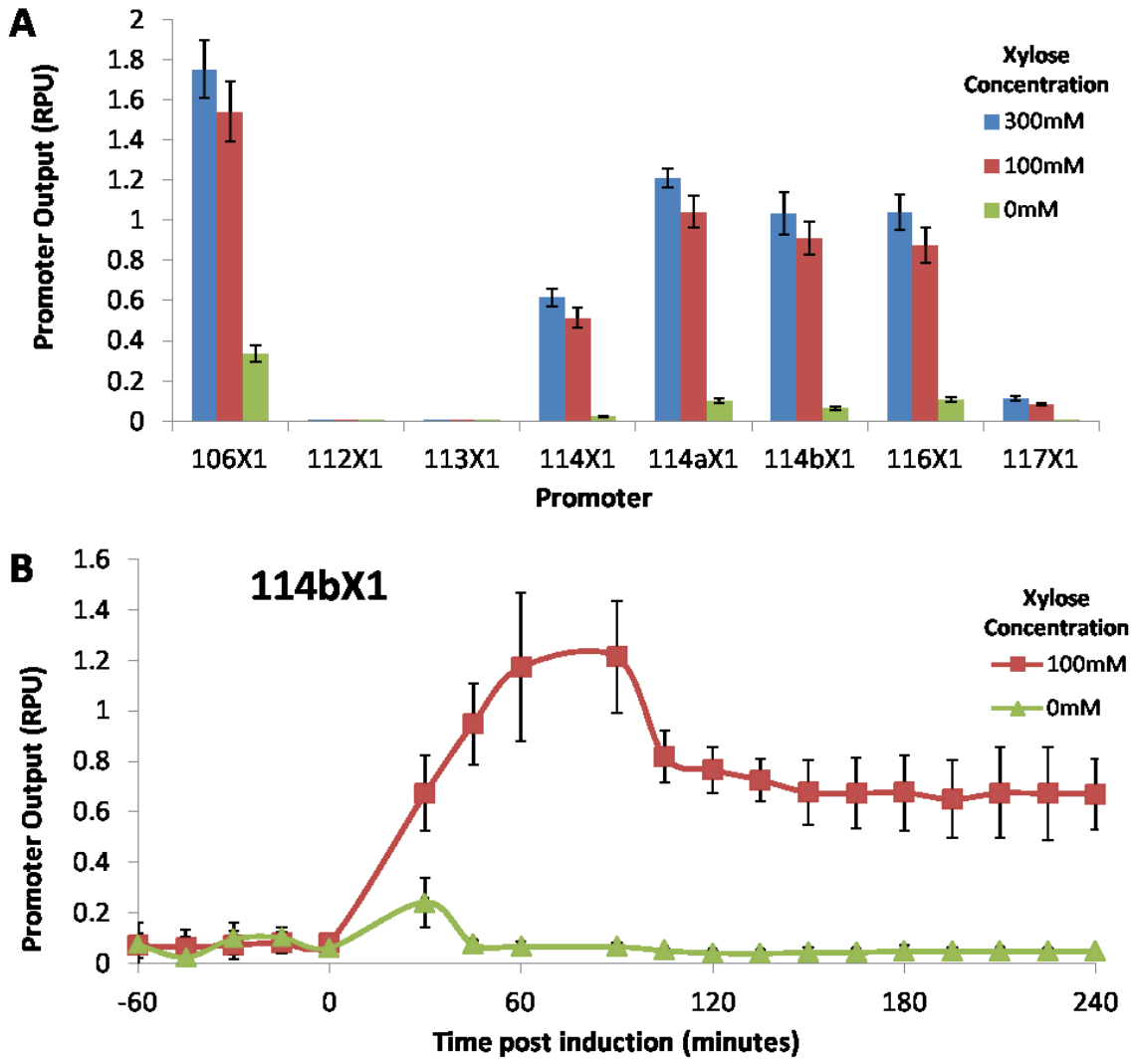
6.5.1 Characterisation of xylose inducible promoter set X1 (*xylF* operators)

The X1 library was the first set to be tested to see if addition of xylose binding sites would enable the promoters to be activated by xylose induction. The secondary aim was to demonstrate whether or not it would be possible to predict the output for a promoter when induced before it was tested and ideally this would be possible from the initial constitutive Anderson collection characterisation experiment.

Testing was carried out using 3 xylose concentrations (300mM, 100mM and 0mM) as these represented the maximum inducer concentrations and a no induction concentration. As before the stock solutions were dissolved in carbon source free MOPS media. The samples were set up for overnight growth in LB in a microplate with a film lid. The assay was carried out under standard conditions of 30°C and 700rpm in EZ MOPS media with an initial concentration of 0.5% glucose as had been used previously for the *xylF* promoter characterisation.

The results for the X1 set were surprising. Of the 8 promoters which were fused to the operator sites from the *xylF* promoter 6 responded positively to induction with xylose (figure 6.30 A). The other 2 showed negligible difference between induced and uninduced. All 6 of the inducible promoters were induced more by 300mM than by 100mM xylose which was to be expected. An expression profile similar to the *xylF* promoter was observed with around an hour required to reach full induction before a small drop to a long period of very regular expression (figure 6.30 B). Interestingly this period of expression was more consistent for at least some of the promoters than observed for *xylF* indicating that an interaction which affected expression over time may have been lost in the transfer to a Synthetic Promoter.

The on:off ratio of the set of promoters bore only a limited relation to the observed uninduced strength (figure 6.30 C) with the third smallest induction ratio being observed for the promoter with the strongest constitutive expression (106/106X1). The strongest on:off ratio came from the second weakest inducible promoter (114/114X1). When the uninduced promoter output was compared to the induced promoter output observed during the initial Anderson collection characterisation (figure 6.31) there was a clear trend that as uninduced activity increased the induced activity also increased but with diminishing returns. The diminishing returns caused the poor on:off ratio result observed with 106/106X1 as the induced output had increased but not proportionally to the uninduced output compared to 114/X114X1. The diminishing returns also suggested that the hypothesis of Cox et al. (2007) that there was an activation threshold above which increased expression could not be achieved was likely to be correct.



C

	106X1	112X1	113X1	114X1	114aX1	114bX1	116X1	117X1
On:Off Ratio (300mM)	5.22	1.16	0.96	27.38	12.01	15.92	9.71	17.30
On:Off Ratio (100mM)	4.59	1.04	1.01	22.93	10.35	14.02	8.20	12.88
Basal Output (RPU)	0.335	0.006	0.001	0.022	0.101	0.065	0.107	0.007

Figure 6.30 Characterisation of 8 Anderson promoters fused with the X1 (*xylF*) operator sites.

Addition of XylR operator sites upstream of constitutive promoter enabled the induction of most of these promoters by xylose. Of the 8 promoters engineered with binding sites 6 respond positively to induction by xylose with 106 producing the most output and 117 the least where induction was observed (A). The expression profile for the synthetic promoters was very uniform. Output rapidly reached the maximum level before relaxing to a stable level (B). The on:off ratios of the 8 synthetic promoters indicated that the strongest promoter may not be the best promoter as its on:off ratio was actually poor comparatively and the best ratios belong to the two weakest inducible promoters (C). For all plots n=9. Data points shown were means and the error was one standard deviation.

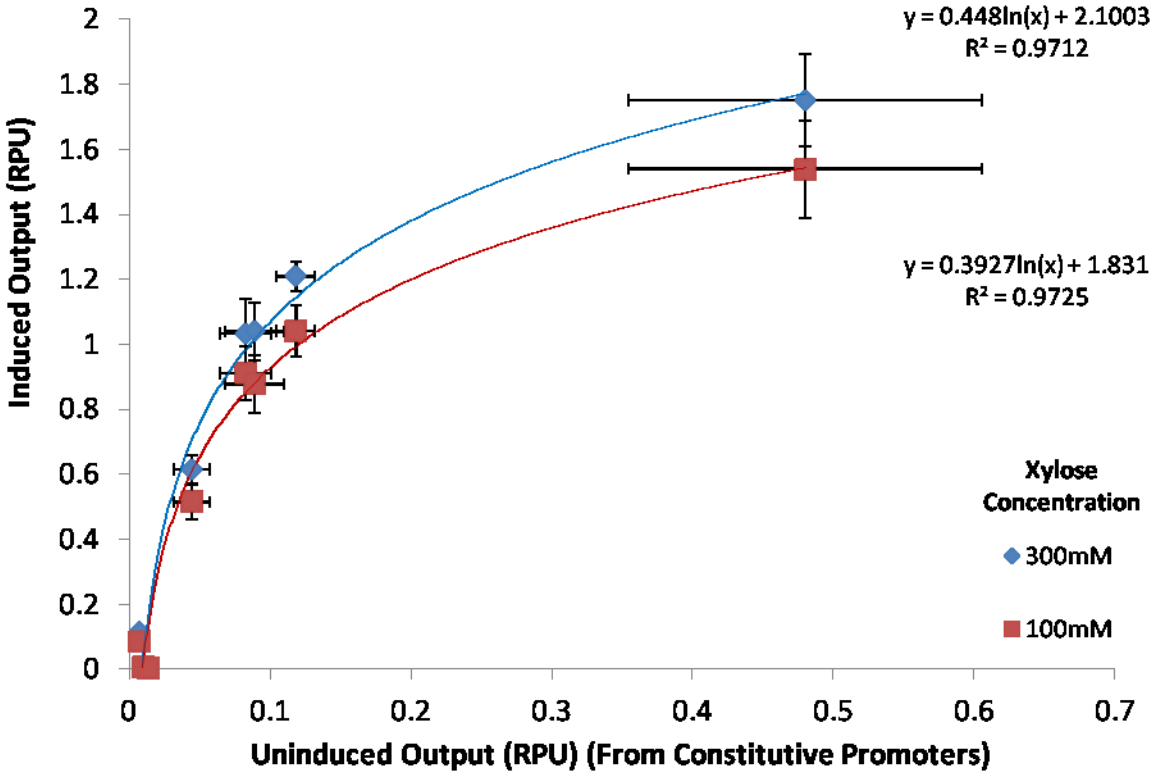


Figure 6.31 Output of the X1 promoters induced with xylose correlates with output for the Anderson promoter used to build them.

There was a clear relationship of increasing induced output with increasing uninduced output from synthetic xylose promoters although the relationship suffered from diminishing returns. The uninduced output results were from the constitutive Anderson promoter collection characterisation (see 6.5.2). For all samples n=9. Data points shown are means and the error is the standard deviation.

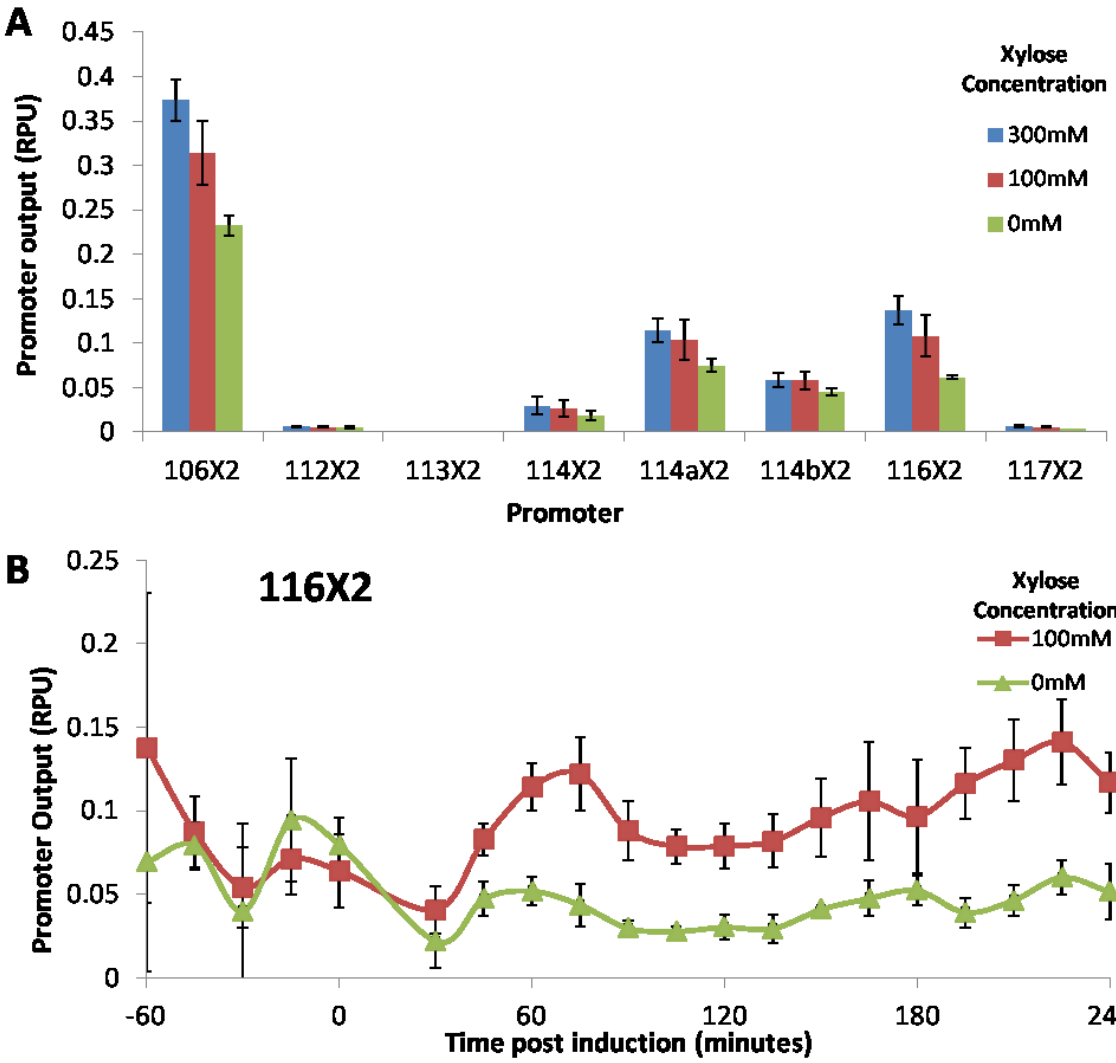
6.5.2 Characterisation of xylose inducible promoter set X2 (hybrid operators)

The first synthetic xylose promoter set was a success and it was important to see if the methodology could be repeated to generate more inducible promoters with new characteristics. The X2 set was designed to see if the binding sites could be used as independent modules or if the needed to be kept in pairs. Where the X1 set had the binding sites from an existing promoter the second set contained the hybrid pairing of one binding site each from *xylA* and *xylF*.

The X2 set was characterised in exactly the same way as X1 set with induction concentrations of 300mM, 100mM and 0mM xylose. The characterisation was carried out following overnight outgrowth in LB in a microplate at 30°C. The characterisation itself used the standard conditions of MG1655 *E. coli*, 30°C at 700rpm and in EZ MOPS media which initially contained 0.5% glucose.

Of the 8 X2 promoters 7 exhibited very little in the way of induction in either xylose concentration and only promoters 106X2, 114aX2 and 116X2 showing any clear induction (figure 6.32 A). Where induction was observed there was not the robust output that had been observed with the X1 set although this was probably partially the result of the very weak output strength of approximately 0.1 RPU (figure 6.32 B). The on:off ratios for all the promoters was poor with only 116X2 having an induced output more than double the uninduced output (figure 6.32 C).

Mixing the operator sites had clearly interfered with the operators sites either by altering their interaction with the promoter or with the XylR regulator. This was backed up by the paper by Ni et al. (2013) which indicated that each operator pair was bound by a single monomer and so swapping these operators may well interfere with the binding.



C

	106X2	112X2	113X2	114X2	114aX2	114bX2	116X2	117X2
On:Off Ratio (300mM)	1.61	1.25	1.03	1.58	1.53	1.32	2.22	1.80
On:Off Ratio (100mM)	1.35	1.07	1.06	1.41	1.38	1.30	1.76	1.52
Basal Output (RPU)	0.232	0.005	0.001	0.019	0.075	0.045	0.062	0.004

Figure 6.32 Characterisation of 8 Anderson promoters fused with the X2 (hybrid *xylF* and *xylA*) XylR operator sites.

The use of hybrid sites for the binding of the XylR protein was not particularly effective. Few of the promoters showed significant response to xylose (A) and those that did produced weak and fairly noisy signals (B). The on:off ratio for these promoters was poor with only one obtaining double the amount of induced output compared to uninduced output (C). For all results n=9. Data points shown were means and the error was one standard deviation.

6.5.3 Characterisation of xylose inducible promoter set X3 (*xylA*)

As the *xylA* promoter has not been characterised and as the X2 set which included one operator sequence and had failed so it was important to identify if using the operators from the *xylA* promoter would allow induction when applied to the promoters. As J23112 and J23113 had both failed in sets X1 and X2 it was also important to test if they would pick up any activity as failure again may indicate they were potentially non-functional.

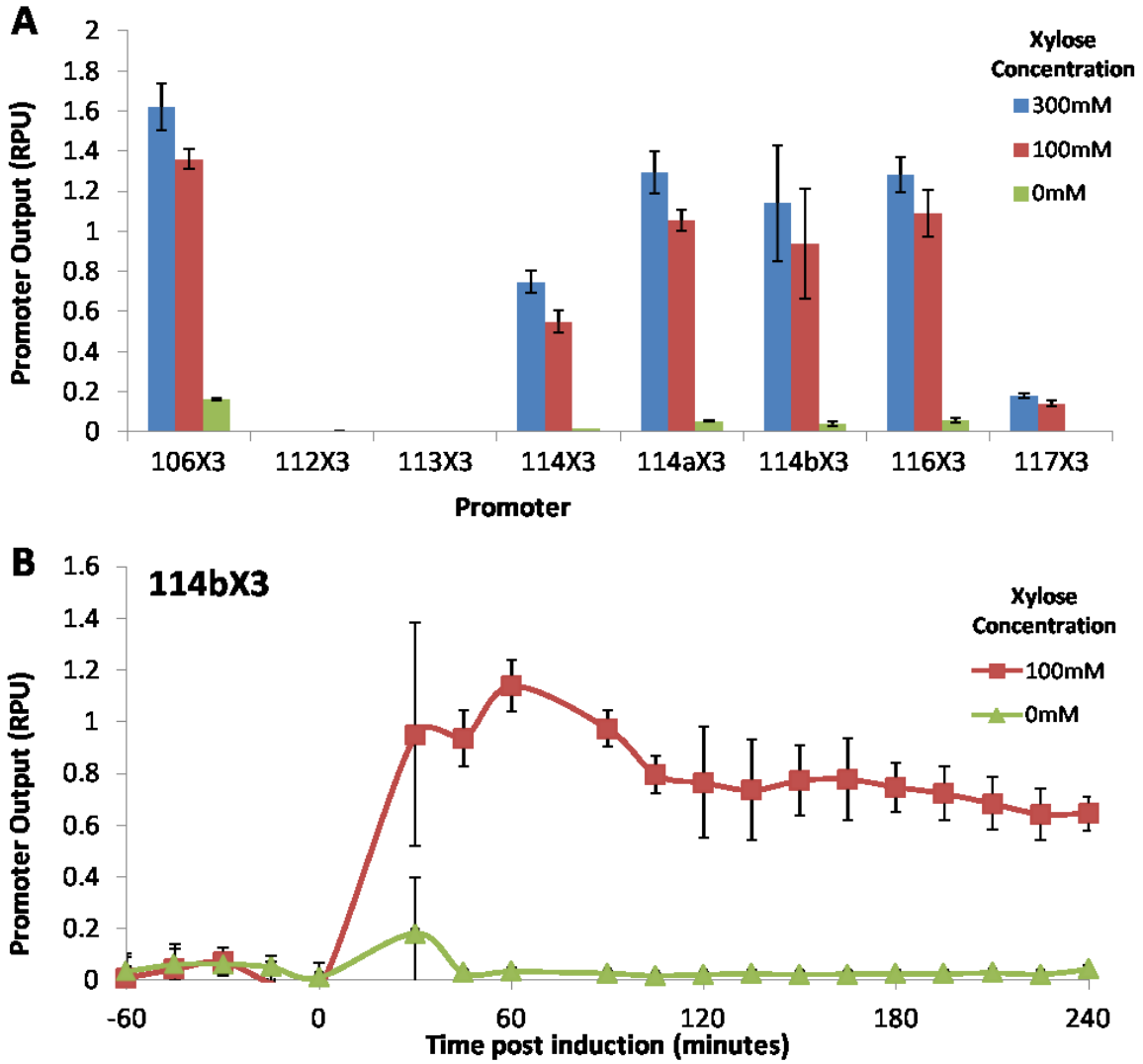
The X3 set was characterised using the same protocol as the X1 and X2 sets using the standard characterisation conditions of MG1655 *E. coli* at 30°C and 700rpm in EZ MOPS media which initially contained 0.5% glucose. The xylose inducer was again diluted in carbon source free MOPS media so final concentrations would be 300mM, 100mM and 0mM.

The X3 set of inducible promoters again contained 6 xylose inducible promoters and 2 failures (figure 6.33 A). The two failures were again promoters based on J23112 and J23113. These promoters are the 2 which has 2 base pair mismatches in the three base pair shared sequence between promoter and operator and this may have interfered with XylR protein binding or the two promoters may simply be non-functional. For the promoters which were xylose inducible the dynamic characteristics were again impressive with full output reached in around an hour before relaxing to a level that was probably less stable than the X1 set but more stable than the *xylF* promoter (figure 6.33 B).

The on:off ratio of the X3 promoter set was significantly better than their X1 counterparts however with the lowest ratio for an xylose inducible promoter being approximately 10 fold and 3 of the X3 promoters possessing on:off ratios of 30 or higher with the highest a ratio of almost 50 (6.33 C). The relationship of uninduced output to induced output was less clear compared to the X1 set. While there was a relationship between the uninduced output observed in the constitutive characterisation experiment and the induced expression a logarithmic function fit the data fairly poorly in comparison to the X1 set (r^2=0.928 v r^2=0.971). In particular the three 114 promoters formed an odd grouping some distance from the fit line (figure 6.34 A).

When the induced expression was compared to the uninduced expression of the X3 set of promoters directly (i.e. the uninduced data collected as part of this experiment) the clear relationship was restored (figure 6.34 B) with a better fit (r^2=0.983) to the curve. When the uninduced samples from all 3 synthetic promoter experiments were compared to the original constitutive data (figure 6.34 C) there was a reduction in the uninduced level of output from all of

the synthetic promoters. This drop in basal level was particularly strong in the X3 set and caused the previously obtained data to no longer predict the induced output as accurately.



C

	106X3	112X3	113X3	114X3	114aX3	114bX3	116X3	117X3
On:Off Ratio (300mM)	10.02	1.30	1.07	48.08	24.08	30.64	22.64	36.95
On:Off Ratio (100mM)	8.41	1.17	1.03	35.25	19.68	25.16	19.23	29.05
Basal Output (RPU)	0.162	0.002	0.002	0.016	0.054	0.037	0.057	0.005

Figure 6.33 Characterisation of 8 Anderson promoters fused with the X3 (*xylA* promoter) operator sites.

The X3 synthetic xylose inducible promoter set contained 6 promoters highly responsive to xylose but also 2 that failed to respond (A). The expression profile of these promoters was similar to the X1 set but with a less robust long term output (B). The on:off ratios of the X3 set were particularly impressive however with one promoter reaching an on:off ratio 3 times higher than the *xylF* promoter and nearly the level of 60 obtained by the *rhaB* promoter (C). For all results n=9, values were means and error was one standard deviation.

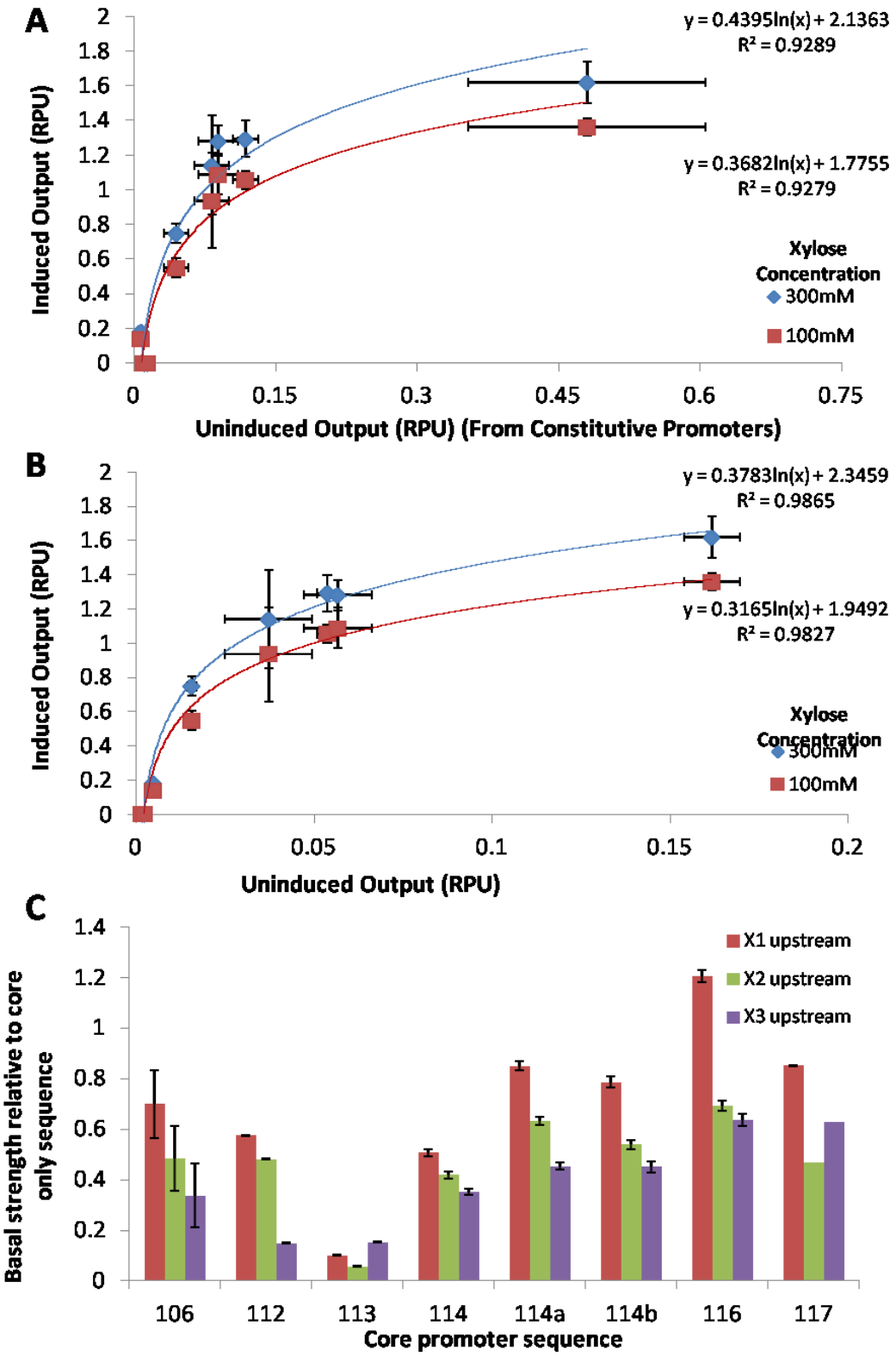


Figure 6.34 Relationship between the induced and uninduced output of the X3 synthetic promoter set and the effect caused by upstream binding sites on uninduced output.

Ideally there would be a clear relationshipbetween the induced output and the constiutive output for the X3 promoter set however the trend while clear was imprecise with the constitutive data (A). When the X3 induced data was compared to X3 uninduced data the relationship was much more clear and accurate than before (B). It was observed that the upstream operator sequences reduced the basal expression levels of the promoters and this effect was much larger for the binding sites taken from *xylA* (X3) than for the binding sites from *xylF* (X1) or the hybrid site (C). This effect was responsible for the imprecision in A. For all results, n=9. Data points shown were means and the error was one standard deviation.

6.5.4 High resolution xylose library member (X114aX) characterisation

While the synthetic promoter library sets clearly responded to xylose it was unknown how these modifications had changed the input-output for the various sets. If the relationship did not change it would be possible to forward engineering large numbers of promoters and test only a few to understand this relationship which would massively increase throughput. To determine if this would be the case three of the promoters from the sets were chosen to characterise in greater detail. Additionally as the X1 set was built using the *xylF* promoter operators it would be possible to compare the results to get an indication of how possible this would be. One of the promoters was also taken from the X2 set in case an in detail analysing could provide insight into why the X2 set had failed.

For this characterisation the synthetic xylose promoters 114aX1, 114aX2 and 114X3 were chosen to keep a level of consistency in the experiment as the core would remain the same. These promoters were characterised using the same set up used to characterise the *xylF* promoter using the same induction concentrations of xylose and standard conditions of MG1655 *E. coli* grown at 30°C and 700rpm in EZ MOPS media with initially 0.5% glucose.

When repeated with the high detail inducible promoter characterisation protocol the 100mM expression profile results for 114aX1 and 114aX3 were almost exactly the same as observed in the library characterisation experiment and this expression profile was be replicated at other inducer concentrations (figure 6.35). For 114aX1 the robust profile was conserved and observed at other inducer concentrations with the slow drop toward the end of the characterisation (figure 6.35 A) replicated at most inducer concentrations. The expression profile patterns were also similar for 114aX3 with the profile replicated across most inducer concentrations (figure 6.35 B).

The Flow cytometry data at the 6 hour time point had been the most reliable for modelling or curve fitting the data for these points was analysed to allow the fitting of the hill equation to the input-output plots obtained for the 114aX promoters (figure 6.36). The input-output relationship of 114aX1 formed a sigmoid consistent with a hill function heading towards a plateau at around 1.2 RPU. The Km obtained for fitting to 114aX1 and the *xylF* promoter gave results with the confidence interval if data for 114aX1 at 6 hours and *xylF* promoter at 3 hours with the *xylF* promoter's 6 hour output falling just outside the confidence interval. Interestingly the hill cooperativety parameter was significantly different in 114aX1 compared to the *xylF* promoter which was surprising given they shared the same XylR operator sequences (figure 6.36 A). There was a very rough input-output relationship for the 114aX2 promoter although a large amount of the signal may have been noise or

error making it unclear if the XylR protein was binding and inducing the promoter (figure 6.36 B). There was a very strong input-output relationship for promoter X114aX3 which is closer for many of the parameters to the *xylF* promoter than the X114aX1 promoter was (figure 6.36C). Any cause for this other than the lower basal fluorescence from X114aX3 was unknown.

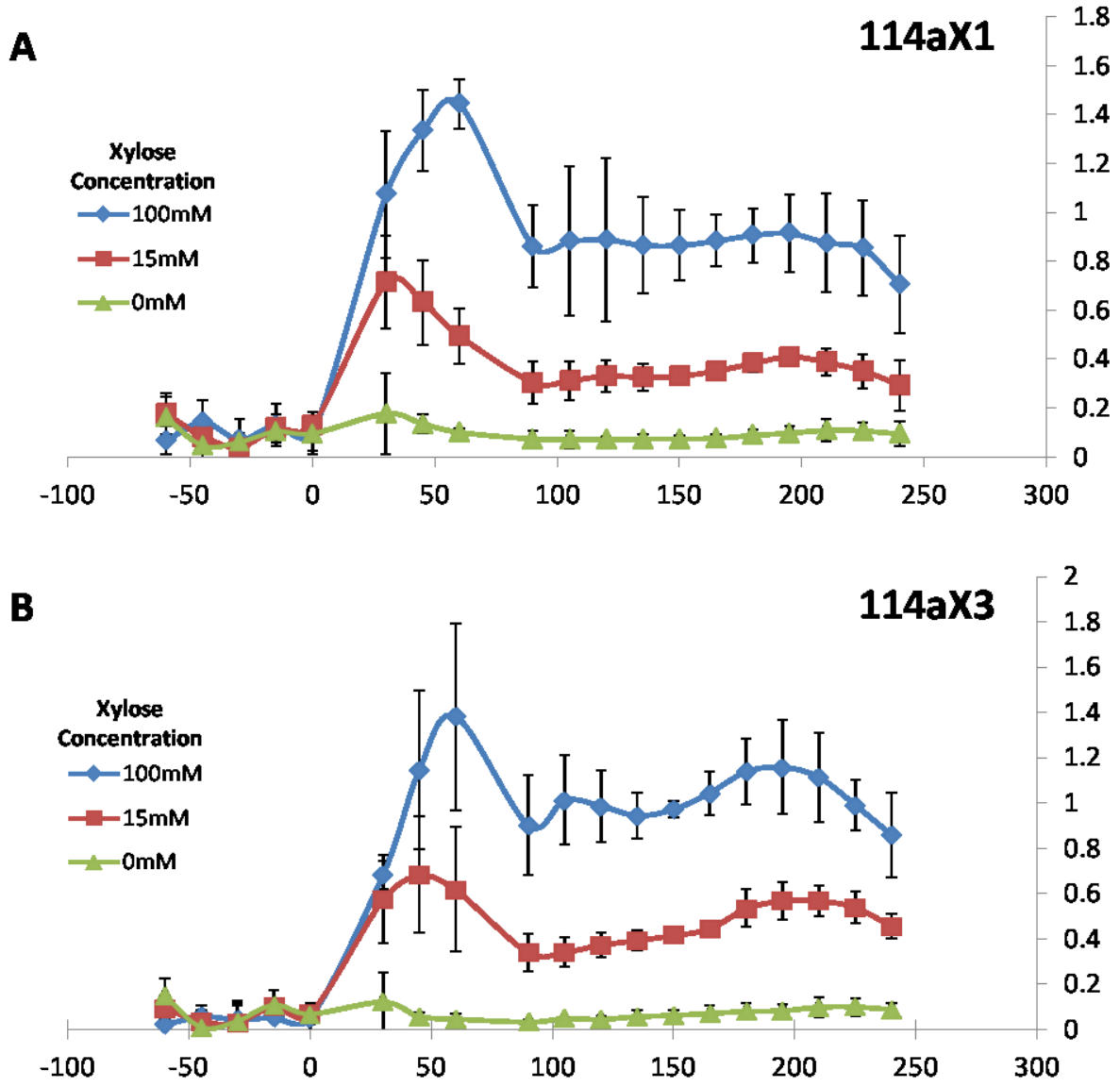


Figure 6.35 Comparison between the dynamic outputs for 3 induction concentrations for the 114a promoter fused to X1 or X3 upstream sequences.

The X114aX1 and X114aX3 have very consistent expression profiles across xylose concentrations. The X114aX1 expression profile (A) is very flat for the bulk of the characterisation whereas the profile of X114X3 (B) is more similar to the profile of the *xylF* promoter. For all results, n=3. Data points shown were means and the error was one standard deviation.

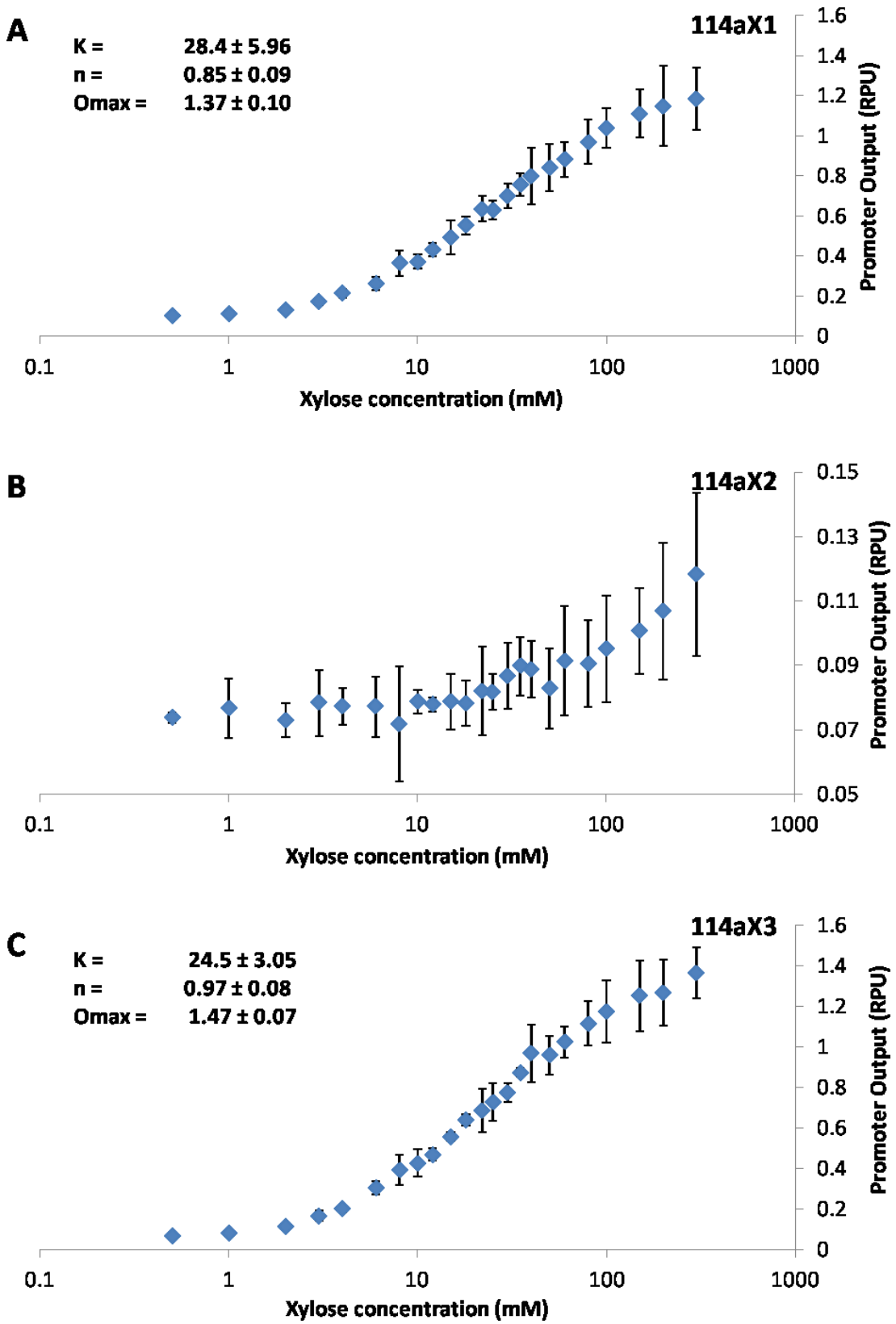


Figure 6.36 Detailed input-output relationship for the 114a synthetic promoters at 4 hours post induction.

The input-output relationship for 114aX1 (A) is approaching sigmoid and appears superficially similar to that of *xylF* with which it shares operator sequence but the parameters for the hill equation (inset) suggest that the input-output relationship of X114aX3 (C) may be closer to that of the xylF promoter. The input-output relationship for X114aX2 may have indicated a slight response to xylose but had too much variation to display statistical significance. For all results n=3. Data points shown were means and the error is one standard deviation.

7 Discussion

7.1 Automated workflow characterisation

The automated BioPart characterisation was designed to enable the gathering of high quality characterisation data high through-put. It achieved this by using the advantages of robotics in being able to rapidly and reliably dilute and subculture cells as part of repeating protocols which were consistent every time. Small variations in timing, temperatures and other factors affected data in surprising ways and automating the procedure removed many of those issues by standardising them. The protocol was design around ensuring high quality growth of the standard characterisation strain and the consistency in growth parameters was an indication of how much the results benefited. This could be beneficial for the field as variation levels in characterisation experiments have been found to be up to 50% and related to lots of variables in the environmental and chassis scenario that could not be controlled (Kelly et al., 2009). As the platform controls and standardises growth many of those uncontrollable factors will probably be standardised and as result largely removing their influence.

Automated tools have the potential to be transformative for Synthetic Biology by allowing the throughput of work to be significantly increased without requiring large increases in manpower for repetitive tasks. With the drop in synthesis costs and advances in assembly technologies, both manual and automated, characterisation had become an increasing bottleneck. The automated characterisation workflow required minimal interaction from a researcher while carrying out characterisation assays and thus allowed the researcher to perform other task such as cloning or designing experiments or devices. Hopefully as labs and groups increasingly utilise automation tools it will be more than possible for a single researcher to look after multiple platforms while they are operating and as a result massively increasing through put.

The workflow was demonstrated characterising a library of constitutive promoters, inducible promoters in very high detail and even the effects of carbon source on promoter output. These applications showed that the platform was capable of collecting large amounts of data while retaining a high level of quality across the whole of bacterial growth. The workflow had been validated in terms of assay robustness and ensuring that the assay did not interfere inappropriately with the samples. This went so far as to suggest that an aerobic characterisation environment was present for the majority of the protocol. In addition the edge effect encountered in microplate based assay for many years was not present in the automated characterisation protocols and while the

reason for this is unknown it may be related to the use of plate lids which have been shown to reduce the edge effect previously (Burt et al., 1979).

The workflow was also used to provide some high detail characterisations for the field of Synthetic Biology. An accurate characterisation of the entire Anderson Collection of promoters had been carried out for the first time which may be of help to various groups, both academic and student based, who use or have used these BioParts as the collection is one of the most widely distributed set of Biological Parts thanks to the parts registry and iGEM (Parts Registry, n.d). Many labs have used one or more of the Anderson collection of BioParts and it may be possible for labs to use their own characterisation data in addition the characterisation results here to aid in the design new of biological devices and systems.

The characterisations of the *xylF* and *rhaB* promoters have demonstrated the application of RPU units to an inducible system and have demonstrated the reasons behind using such an integrated and thorough approach to characterisation in the exploration of the interesting behaviour of the *rhaB* promoter in particular. The data obtained may provide interesting insight into the biological interactions surrounding rhamnose induction when a synthetic construct adding additional RhaS protein is employed. The platform had also been used to confirm one of the untested uses of the RPU standard as the data produced demonstrated that promoter expression levels remain relatively consistent between the MG1655 and MDS42 strains and in media containing different carbon sources thus supporting the arguments of Kelly et al. (2009).

Beyond this the developed automated BioPart characterisation workflow is an opportunity for the field of Synthetic Biology. Characterisation has suffered from a lack of a standardisation drive even though time has been invested significantly in this area (Kelly et al., 2009, Canton et al., 2008). Many of the characterisations carried out in Synthetic Biology seemed to be done in a relatively ad hoc manner often seemingly pulling in the most convenient tools (chassis) and materials (media) available. While this has allowed the publication of significant amounts of characterisation data up to now for the future of Synthetic Biology adoption of standard procedures and tools will enable the tracking of sources of biological and contextual interactions that can only be detected in large volumes of data. The automated characterisation workflow is an opportunity for labs to buy into a common characterisation standard to a level they feel is appropriate.

For some groups adapting to a new standard could be hard but by using the workflow this has been simplified to some degree. The core of the workflow was the choice of media and strain around which everything else was built. The protocol itself was the method to get the data of the

best possible quality out of the tools of glucose based MOPS media and MG1655 *E. coli*. The automation side of the standard is the ideal best practice but while this is the case the protocol can be carried out by hand with a smaller number of samples. Where there are equipment limitations, such as access only to a plate reader, tools have been developed to help understand the characterisation environment and data obtained from one source is still relatively acceptable. While designed and tested in the characterisation of promoter BioParts the core was the consistent reliable growth of *E. coli* which should be a component of all *in vivo* characterisation experiments. Furthermore expression BioPart characterisation (for example Ribosome Binding Site and some UTR BioParts) may be possible using the developed protocol with the right data models or reporters. Ultimately it should be possible to build out from the skeleton documented here to add increasingly difficult but insightful procedures that allow the characterisation of a whole range of BioParts.

The workflow has a head start already in that it has been developed around 2 key components which have declared standard for a significant group of labs working collaboratively in Europe. For those labs updating their protocols to fit with the standard characterisation procedure should be a small step and with a small group pushing forward with a standard momentum may build. Many of the methods put forward by Canton et al. (2008) were briefly picked up before momentum was lost and the RPU standard created by Kelly et al. (2009) has convinced a few groups to use a small set of changes to conform to the standard. With a group adopting elements of the standard designed to get the best out of those elements there is hope the standard automated characterisation workflow can begin to gain some traction.

7.2 Data standardisation and datasheets

Data standards are vitally important for Synthetic Biology and the transfer of data between individuals in the field. This work attempted to cherry pick the best current standardisation and documentation techniques while supplementing them with new standards and ideas where something wasn't working.

The first data standard that was utilised and was very useful was the RPU standard by Kelly et al. (2009). Biology is full of context and interactions can influence the results of a characterisation experiment or the effectiveness of a design of a biological system. The automated promoter BioPart characterisation workflow was set up to produce data in RPU as one of its outputs in order to account for a large chunk of the contextual influence on characterisation.

For experiments using the automated characterisation workflow a stricter interpretation of the appropriate reference construct was used with the entire plasmid identical except for the promoter being tested. The original standard only required the 5' region of the transcript to be kept consistent but by keeping the entire plasmid consistent this should account for other unknown context effects that may be caused by sequences on the plasmid. One such affect was noted during characterisation when the presence of a different sequence upstream of the synthetic xylose promoters altered their constitutive output. The drawback with using RPU as the lone standard for data was that the results would be strictly relative and may hide potentially interesting data.

For this reason RPU was useful for data display and use but was not the sole type of data that could be generated by the protocol. A complementary approach with a second unit set based on absolute values was incorporated into the experimental design. A prime candidate was the molecules of GFP per cell units set proposed by Canton et al. (2008). These units were particularly appropriate because they were responsible for the signal that was being observed. While carrying out the initial calibrations to convert observed data into the absolute units was tricky and time consuming it only had to be carried out once. This may be a double edged sword as the calibrations are unlikely to be repeated unless clearly anomalous results were to be obtained and it would be wise to attempt to double check the results. To make these issues more clear a simpler calibration was sought for labs to use that could allow the checking of these calibrations my multiple groups. This could be achieved by carrying out the calibrations experiments with collaborative labs and then use the fluorescein calibration described in 6.4.2.2 as a standard measure to allow comparison of the calibration.

The fluorescein standard calibration was one area where a common tool was lacking. There were a large number of labs working in Synthetic Biology characterising and testing design in plate readers made by various manufacturers. Even 2 plate readers from the same manufacturer gave slightly different results in arbitrary units when compared to each other. With this being the case there was a need for a simple and easy to use standard for all labs. Unlike the GFP lysate experiments required by Canton et al. (2008) the fluorescein standard curve required only a small amount of time to set up and was accurate for the duration of this project. The data acquired during calibration also suggested it may be possible to use fluorescein directly to calibrate into molecules of GFPmut-3b however this would again be best attempted as a collaborative effort. Fluorescein as a standard for plate reader measurements may also be more likely to receive uptake as fluorescein is already used in the standardisation and calibration beads used with a flow cytometer where data is converted to molecules of equivalent fluorescence based on fluorescein.

Hand in hand with data standardisation was the development of datasheets. The datasheets displayed in this work had in a very different niche to either of the other two datasheet produced by the field so far. The BglBrick datasheets by Lee et al. (2011) were very much designed to document an expression vector and focused on the amount of protein which would be generated under many conditions while also paying attention to the cross reactivity of inducers. The datasheet by Canton was much closer to the one presented here but was still in a separate discrete niche as it was designed to primarily present the data for a characterisation of a simple activator device not too dissimilar to the *xylF* or *rhaB* promoter characterisations from this work.

The datasheet displayed here and the datasheet format results for the *xylF* and *rhaB* promoters were very much centred on the characterisation itself and of key note is the array of metadata associated with the datasheets. The datasheets contain all the information about the characteristics you could want to see, whether analysed or the raw traces of the data. The key difference was the second page which gave the user an overview of the entire characterisation from the media used to the equipment settings and plasmid the experiment was carried out using right the way through to data analysis settings. Following these first version of datasheets newer prototypes have been developed and the characterisation metadata articles have been expanded upon further. For inducible systems the key characteristics figures for *xylF* and *rhaB* promoters and each of the engineered promoter sets are similar to the layouts and data for the current prototype datasheet versions.

7.3 Oxygen Biosensors

As part of the verification of the automated platform a paired biosensor was developed to tap into an oxygen responsive protein and the anaerobic metabolic pathways of *E. coli*. One of the reasons for the development of this tool was the lack of serious alternatives with very few tools published or available for monitoring oxygen in microplate well let alone *in vivo*. In terms of *in vivo* biosensors there are few examples and they are primarily developed to detect the free oxygen compounds which are frequently associated with infection.

The only other biosensor clearly capable of detecting the presence of oxygen *in vivo* was a fret based biosensor with a normal fluorescent protein and an anaerobic fluorescent protein fused together so that resonance can only occur in the presence of oxygen (Potzkei et al., 2012). This biosensor had a far wider application range as it could be expressed in a variety of cells and with a

decent likelihood of correct function. The FNR based biosensors developed here were a solution designed specifically for *E. coli* and may require some *E. coli* proteins to recycle the FNR protein into a re-usable state regularly or to refresh it following triggering. They were also designed to be multipurpose and detect the transition into anaerobic metabolic conditions as the sensors were based on the *E. coli* system responsible for regulating anaerobic metabolism which was the real concern of the experiment. Unfortunately it was not possible to determine if the biosensors were detecting a drop in localised oxygen levels or if they were reporting a change in the overall metabolic state inside the cells however the clear switch to an 'on' state (low oxygen) suggests that they did meet the original objective of indicating when an assay was likely to be aerobic. The affect of the sensors themselves had upon the cells was unknown because although cultures carrying the *dmsA* promoter spontaneously died when growing rapidly in MOPS media the same was noted for the pAC plasmids in LB media anyway (Casini, 2010).

While the sensors were also designed as a pair in order to reduce the impact dropping levels of oxygen may have on the production of the fluorophore it still may be the case that there was a hidden effect that may only be removed by the successful switch of the reporter gene to an oxygen insensitive reporter. Only upon testing with the fret oxygen biosensor or in conditions where the level of oxygen can be completely controlled will it be possible understand exactly which signal the biosensors are relaying.

7.4 Effect of carbon source on BioPart activity

The automated characterisation workflow was demonstrated to be capable of the testing context effects on BioPart functions. While it was possible to collect this data the volume of results produced by this characterisation assay was large enough to prove difficult to analyse without dedicated scripts beyond those currently in use. A more efficient automated method for analysing these results and dealing with the variations between samples under the similar conditions is required to make the most of such a dataset. To achieve significant goals in terms of understanding the effect of carbon sources and other context components on BioParts larger models and datasets will be required and the tools to analyse them in particular will need significant improvement.

The change in carbon source was observed to alter the level of BioPart activity and also the growth rate of the *E. coli* MG1655 chassis. The change in growth rate was however relatively small and there was no significant correlation between the growth rate and the BioPart activity though

this may be a result of the relatively small changes in growth rate across the different carbon sources rather than the lack of a correlation. There was however an indication that for simple constitutive BioParts any change in conditions from the standard characterisation conditions will affect each BioPart relatively equally based on their characteristics under standard conditions. As such it may be possible to take the carbon source dataset and use this to predict how the observed activity of the promoter BioPart would change when in a new context.

Significantly the carbon source dataset suggested that RPU may remain consistent in different context scenarios for simple BioParts as originally suggested by Kelly et al. (2009). This fitted with the earlier observation as it suggested that the activity of the BioParts would stay at a level consistent with each other i.e. that if the output of a given BioPart was increased by 20% generally the output of the others would increase by 20%. There was however limitations to this as it was observed that the change in output for the strongest promoter characterised in the experiment did not match particularly well when the absolute values were considered even though the RPU at mid-log displayed no significant difference. It was also important to remember that there was only no significant difference in RPU values across the carbon sources when the mid-log synthesis rate rather than a single time point was chosen. As pointed out by Kelly et al. (2009) this may be caused by the peak in synthesis rate at this point and care about highlights the care that needs to be take with the conditions of characterisation.

7.5 Characterisation and forward engineering of inducible promoters

BioPart engineering had become an increasingly useful tool for developing large numbers of BioParts for use in biological designs. Recently a variety of engineering promoters had been developed for various processes including large libraries of the commonly required but simple constitutive promoters (Mutalik et al., 2013) and also a library of repressible promoters ideal for designing logic circuits (Stanton et al., 2013). Engineered promoters had advantages over natural ones as they can be completely synthetic sequences where biological context had been removed which made them more suitable for engineering as there was a reduced likelihood of cross talk or hidden interaction with a cell chassis. While this was the case pretty much all engineered promoters were constitutive or repressible and while these promoters were a large part of the base control and logic toolbox the ability to engineer all the tools would be an advantage. Promoters which work by activation have a few key advantages in that they are capable of adopting an ‘off’ state that is almost

completely off (RPU < 0.02) and do not require large numbers of repressor proteins in order to keep expression down. Activated promoters only need large amounts of protein if they need to be fully on and producing a lot of output.

The *rhaB* and *xylF* promoters are good examples for this having basal states of next to 0. The promoters used in this research were also semi-engineered as they have had all the unnecessary sequence removed in an attempt to decouple them from the *E. coli* context particularly the interaction both wild type promoters had with the CAP protein. The promoters were as decoupled as they can get inside an *E. coli* cell and for any further decoupling orthogonal regulators and promoters were required and orthogonal systems are a likely next target. Theoretically both promoters can be activated using the genomic copy of their regulator gene present but in practice the native levels of the proteins are probably too low to achieve a useful output.

The synthetic xylose based inducible promoters were only the second recent project to attempt to predictably engineer promoters inducible by an activation mechanism. The previous attempt by (Cox et al., 2007) was an attempted to generate promoters from modules with fairly high basal levels and so could not produce data appropriate for forward engineering. Instead their data suggested the existence of an activation threshold level above which activation can no longer boost a promoter. The synthetic xylose promoters represented an alternative strategy using very weakly constitutive promoters to generate new promoters that were below the activation threshold and the data indicated that the hypothesis was likely to be correct. The 12 high quality members of the sets had an induced output highly predictable by their uninduced output although only 6 of these could be predicted from the data for the constitutive sequence. An important factor on the forward engineering of activation based promoters was the sequence immediately upstream of the promoter as these sequences have been shown to affect the strength of the downstream promoter (Ellinger et al., 1994). This affect had a large impact on xylose set 3 resulting in a lower uninduced output level which rendered data gathered on the constitutive promoters inaccurate.

The predictive design of these promoters went one step further. The X1 set was designed using the XylR binding sites from the *xylF* promoter and it appears that when those binding sites were transferred the robust expression profile first observed with the *xylF* promoter was shared with the new synthetic promoters implying that some of the more subtle properties of the promoters may also be predictable. These synthetic promoters were capable of reaching strong levels of output (~RPU1.3) and mostly had very low basal levels with the highest on:off ratio at almost 50:1 for one of the promoters. The only real drawback to the design of these promoters was that they were not orthogonal and so their output may be modified by the *E. coli* host processes.

8 Future Work

8.1 Automated Characterisation and Data Standardisation

The automated characterisation workflow has been demonstrated and validated for the standard conditions. The most important piece of work to accomplish in this regard should hopefully be the documentation and spread of the workflow and conditions as a new standard for Synthetic Biology. It would be useful for other labs to attempt to utilise the standard in their procedure if possible or to report back any problems which occur so updates can be made.

Fluorescein calibration of the equipment of other labs and re-testing of the GFP calibration experiments at other sites and by other labs would be extremely desirable to really to increase the accuracy of the calibration.

Beyond this it would be useful to further work with the protocol to tweak the few remaining issues such as cell settling and also to begin to optimise dilutions for more *E. coli* strains and conditions allowing these to also be tested as robustly as possible. The workflow itself should be treated as a skeleton to which new assays should be added to allow the characterisation of more BioParts. In particular these include the capability to carry out assay such as rtPCR from *in vivo* samples and to use new reporters such as the spinach aptamer thus allowing the characterisation of more BioParts by the same overall protocol.

Development of datasheets is ongoing as part of the SynBIS overall project and will centre on further improvements to documentation of characterisation to form part of an online data repository and as an exchange standard.

8.2 Oxygen Biosensors

While the oxygen biosensors appear to indicate that growth is aerobic it would be good to fully test their functionality. In order to do this the reporter gene would need to be swapped for the anaerobic fluorescent protein or another oxygen insensitive reporter before further testing.

At this point an experiment to compare the FNR based biosensors to the fret based biosensor made by Potzkei et al. (2012) *in vivo* would be useful for the *E. coli* biosensors in order to properly understand exactly which mechanism or effect the biosensor is detecting or if it could be a

combination of both. Alternatively it may be possible to tweak the system to be far more sensitive to lower levels of the FNR protein as the current versions require the gene to be over expressed. The simplest method for doing this would be to either search the literature then test more sensors or to increase the RBS strength for the reporter gene.

For the aerobic testing of the standard characterisation workflow could now also be repeated with the fret oxygen biosensor (Potzkei et al., 2012) to have a second indicator of the oxygen available to cells.

8.3 Context Effects on BioPart Activity

With the development of SynBIS continuing rapidly the data storage and analysis are almost at the stage where experiments to test a range of context effects can be carried out. Following on from the carbon source texting results there are plenty of targets for further testing including other major carbon sources such as lactose, other media components particularly amino acid sources and base sources, a variety of temperatures and even other chassis. Carrying out these experiments has the potential to produce large amounts of useful data for the field and particularly large dataset which will be ideal for data mining when they are properly stored inside the SynBIS database.

8.4 Forward Engineering of inducible promoters

The data obtained in this research demonstrated that forward engineering of activation based inducible promoters should be possible. As a first step more members of the Anderson collection should be engineered to respond to inducer to generate more data for the production of the models which will enable forward engineering. Beyond this techniques such as genomic mining (Stanton et al., 2013) should be employed to identify new activators which may be used for development of orthogonal synthetic promoters and further contributing to models to allow their forward engineering.

9 Bibliography

AIBA, H., HANAMURA, A. & YAMANO, H. 1991. Transcriptional terminator is a positive regulatory element in the expression of the Escherichia coli crp gene. *J Biol Chem,* 266**,** 1721-7.

ALPER, H., FISCHER, C., NEVOIGT, E. & STEPHANOPOULOS, G. 2005. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A,* 102**,** 12678-83.

ANDERSEN, J. B., STERNBERG, C., POULSEN, L. K., BJORN, S. P., GIVSKOV, M. & MOLIN, S. 1998. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol,* 64**,** 2240-2246.

ANDERSON, J. C., CLARKE, E. J., ARKIN, A. P. & VOIGT, C. A. 2006. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol,* 355**,** 619-27.

ANDERSON, J. C., DUEBER, J. E., LEGUIA, M., WU, G. C., GOLER, J. A., ARKIN, A. P. & KEASLING, J. D. 2010. BglBricks: A flexible standard for biological part assembly. *J Biol Eng,* 4**,** 1.

ANDERSON, J. C., VOIGT, C. A. & ARKIN, A. P. 2007. Environmental signal integration by a modular AND gate. *Mol Syst Biol,* 3**,** 133.

ANDRIANANTOANDRO, E., BASU, S., KARIG, D. K. & WEISS, R. 2006. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol,* 2**,** 2006 0028.

ARKIN, A. 2008. Setting the standard in synthetic biology. *Nat Biotechnol,* 26**,** 771-4.

BEAL, J., LU, T. & WEISS, R. 2011. Automatic Compilation from High-Level Biologically-Oriented Programming Language to Genetic Regulatory Networks. *PLoS One,* 6.

BEAL, J., WEISS, R., YAMAN, F., DAVIDSOHN, N., AND ADLER, A. 2012. A Method for Fast, High-Precision Characterization of Synthetic Biology Devices. *MIT-CSAIL-TR,* 2012.

BLOUNT, B. A., WEENINK, T., VASYLECHKO, S. & ELLIS, T. 2012. Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One,* 7**,** e33279.

BRASIER, A. R. & RON, D. 1992. Luciferase Reporter Gene Assay in Mammalian-Cells. *Methods Enzymol,* 216**,** 386-397.

BURT, S. M., CARTER, T. J. N. & KRICKA, L. J. 1979. Thermal-Characteristics of Microtitre Plates Used in Immunological Assays. *J Immunol Methods,* 31**,** 231-236.

BUSBY LAB. 2013. Available: http://www.birmingham.ac.uk/staff/profiles/biosciences/busby-steve.aspx.

CAMBRAY, G., GUIMARAES, J. C., MUTALIK, V. K., LAM, C., MAI, Q.-A., THIMMAIAH, T., CAROTHERS, J. M., ARKIN, A. P. & ENDY, D. 2013. Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res,* 41**,** 5139-5148.

CANTON, B., LABNO, A. & ENDY, D. 2008. Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol,* 26**,** 787-93.

CARDINALE, S., JOACHIMIAK, M. P. & ARKIN, A. P. 2013. Effects of genetic variation on the E. coli host-circuit interface. *Cell Rep,* 4**,** 231-7.

CAROTHERS, J. M., GOLER, J. A., JUMINAGA, D. & KEASLING, J. D. 2011. Model-driven engineering of RNA devices to quantitatively program gene expression. *Science,* 334**,** 1716-9.

CARRIER, T., JONES, K. L. & KEASLING, J. D. 1998. mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system. *Biotechnol Bioeng,* 59**,** 666-72.

CASINI, A. 01 September 2010. *RE: pAC Plasmid characteristics.* Type to HIRST, C. D.

CASINI, A., MACDONALD, J. T., JONGHE, J. D., CHRISTODOULOU, G., FREEMONT, P. S., BALDWIN, G. S. & ELLIS, T. 2014. One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res,* 42**,** e7.

CERONI, F., FURINI, S., GIORDANO, E. & CAVALCANTI, S. 2010. Rational design of modular circuits for gene transcription: A test of the bottom-up approach. *J Biol Eng,* 4**,** 14.

CHAN, C. L. & GROSS, C. A. 2001. The anti-initial transcribed sequence, a portable sequence that impedes promoter escape, requires sigma(70) for function. *Journal of Biological Chemistry,* 276**,** 38201-38209.

CHANDRAN, D., BERGMANN, F. T. & SAURO, H. M. 2009. TinkerCell: modular CAD tool for synthetic biology. *J Biol Eng,* 3**,** 19.

CHAPPELL, J., JENSEN, K. & FREEMONT, P. S. 2013. Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res,* 41**,** 3471-3481.

CHEN, Y. J., LIU, P., NIELSEN, A. A. K., BROPHY, J. A. N., CLANCY, K., PETERSON, T. & VOIGT, C. A. 2013. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Methods,* 10**,** 659-+.

CHOE, M. & REZNIKOFF, W. S. 1991. Anaerobically expressed Escherichia coli genes identified by operon fusion techniques. *J Bacteriol,* 173**,** 6139-46.

CHOI, Y. J. & LEE, S. Y. 2013. Microbial production of short-chain alkanes. *Nature,* 502**,** 571-4.

COOKSON, N. A., MATHER, W. H., DANINO, T., MONDRAGON-PALOMINO, O., WILLIAMS, R. J., TSIMRING, L. S. & HASTY, J. 2011. Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol Syst Biol,* 7.

COX, R. S., 3RD, SURETTE, M. G. & ELOWITZ, M. B. 2007. Programming gene expression with combinatorial promoters. *Mol Syst Biol,* 3**,** 145.

DANINO, T., MONDRAGON-PALOMINO, O., TSIMRING, L. & HASTY, J. 2010. A synchronized quorum of genetic clocks. *Nature,* 463**,** 326-330.

DAVIS, J. H., RUBIN, A. J. & SAUER, R. T. 2011. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res,* 39**,** 1131-41.

DE MORA, K., JOSHI, N., BALINT, B. L., WARD, F. B., ELFICK, A. & FRENCH, C. E. 2011. A pH-based biosensor for detection of arsenic in drinking water. *Anal Bioanal Chem,* 400**,** 1031-1039.

DEKEL, E. & ALON, U. 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature,* 436**,** 588-592.

DENSMORE, D., HSIAU, T. H. C., KITTLESON, J. T., DELOACHE, W., BATTEN, C. & ANDERSON, J. C. 2010. Algorithms for automated DNA assembly. *Nucleic Acids Res,* 38**,** 2607-2616.

DOIG, S. D., PICKERING, S. C. R., LYE, G. J. & BAGANZ, F. 2005. Modelling surface aeration rates in shaken microtitre plates using dimensionless groups. *Chemical Engineering Science,* 60**,** 2741-2750.

DUEBER, J. E., WU, G. C., MALMIRCHEGINI, G. R., MOON, T. S., PETZOLD, C. J., ULLAL, A. V., PRATHER, K. L. & KEASLING, J. D. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol,* 27**,** 753-9.

EGAN, S. M. & SCHLEIF, R. F. 1993. A Regulatory Cascade in the Induction of Rhabad. *J Mol Biol,* 234**,** 87-98.

EICHENBERGER, M., CHAPPELL, J., JENSEN, K. E. & ROUILLY, V. 9th of March 2009. *RE: Personal Communication.* Type to HIRST, C.

ELLINGER, T., BEHNKE, D., KNAUS, R., BUJARD, H. & GRALLA, J. D. 1994. Context-Dependent Effects of Upstream a-Tracts - Stimulation or Inhibition of Escherichia-Coli Promoter Function. *J Mol Biol,* 239**,** 466-475.

ELLIS, T., WANG, X. & COLLINS, J. J. 2009. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol,* 27**,** 465-71.

ELOWITZ, M. B. & LEIBLER, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature,* 403**,** 335-8.

ENDY, D. 2005. Foundations for engineering biology. *Nature,* 438**,** 449-53.

ENGLER, C., KANDZIA, R. & MARILLONNET, S. 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One,* 3.

FORTMAN, J. L., CHHABRA, S., MUKHOPADHYAY, A., CHOU, H., LEE, T. S., STEEN, E. & KEASLING, J. D. 2008. Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol,* 26**,** 375-81.

FUKAZAWA, H., MIZUNO, S. & UEHARA, Y. 1995. A microplate assay for quantitation of anchorage-independent growth of transformed cells. *Anal Biochem,* 228**,** 83-90.

GALDZICKI, M., RODRIGUEZ, C., CHANDRAN, D., SAURO, H. M. & GENNARI, J. H. 2011. Standard biological parts knowledgebase. *PLoS One,* 6**,** e17005.

GARDNER, T. S., CANTOR, C. R. & COLLINS, J. J. 2000. Construction of a genetic toggle switch in Escherichia coli. *Nature,* 403**,** 339-42.

GIBSON, D. G. 2011. Enzymatic Assembly of Overlapping DNA Fragments. *Synthetic Biology, Pt B,* 498**,** 349-361.

GIBSON, D. G., GLASS, J. I., LARTIGUE, C., NOSKOV, V. N., CHUANG, R. Y., ALGIRE, M. A., BENDERS, G. A., MONTAGUE, M. G., MA, L., MOODIE, M. M., MERRYMAN, C., VASHEE, S., KRISHNAKUMAR, R., ASSAD-GARCIA, N., ANDREWS-PFANNKOCH, C., DENISOVA, E. A., YOUNG, L., QI, Z. Q., SEGALL-SHAPIRO, T. H., CALVEY, C. H., PARMAR, P. P., HUTCHISON, C. A., SMITH, H. O. & VENTER, J. C. 2010. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science,* 329**,** 52-56.

GLASS, J. I., ASSAD-GARCIA, N., ALPEROVICH, N., YOOSEPH, S., LEWIS, M. R., MARUF, M., HUTCHISON, C. A., 3RD, SMITH, H. O. & VENTER, J. C. 2006. Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A,* 103**,** 425-30.

GRUNBERG-MANAGO, M. 1999. Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu Rev Genet,* 33**,** 193-227.

GULATI, S., ROUILLY, V., NIU, X., CHAPPELL, J., KITNEY, R. I., EDEL, J. B., FREEMONT, P. S. & DEMELLO, A. J. 2009. Opportunities for microfluidic technologies in synthetic biology. *J R Soc Interface,* 6 Suppl 4**,** S493-506.

GUPTA, S., BRAM, E. E. & WEISS, R. 2013. Genetically Programmable Pathogen Sense and Destroy. *ACS Synth Biol*.

HALL, B. G., ACAR, H., NANDIPATI, A. & BARLOW, M. 2014. Growth Rates Made Easy. *Mol Biol Evol,* 31**,** 232-238.

HANAHAN, D., JESSEE, J. & BLOOM, F. R. 1991. Plasmid Transformation of Escherichia-Coli and Other Bacteria. *Methods Enzymol,* 204**,** 63-113.

HEIM, R., PRASHER, D. C. & TSIEN, R. Y. 1994. Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *Proc Natl Acad Sci U S A,* 91**,** 12501-12504.

HILLSON, N. J., ROSENGARTEN, R. D. & KEASLING, J. D. 2012. j5 DNA Assembly Design Automation Software. *ACS Synthetic Biology,* 1**,** 14-21.

HOOK-BARNARD, I. G. & HINTON, D. M. 2007. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul Syst Bio,* 1**,** 275-93.

KAHL, L. J. & ENDY, D. 2013. A survey of enabling technologies in synthetic biology. *J Biol Eng,* 7**,** 13.

KAPLAN, S., BREN, A., ZASLAVER, A., DEKEL, E. & ALON, U. 2008. Diverse two-dimensional input functions control bacterial sugar genes. *Mol Cell,* 29**,** 786-792.

KELLY, J. R., RUBIN, A. J., DAVIS, J. H., AJO-FRANKLIN, C. M., CUMBERS, J., CZAR, M. J., DE MORA, K., GLIEBERMAN, A. L., MONIE, D. D. & ENDY, D. 2009. Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng,* 3**,** 4.

KITNEY, R. & FREEMONT, P. 2012. Synthetic biology - the state of play. *FEBS Lett,* 586**,** 2029-2036.

KITTLESON, J. T., CHEUNG, S. & ANDERSON, J. C. 2011. Rapid optimization of gene dosage in E. coli using DIAL strains. *J Biol Eng,* 5**,** 10.

KLUMPP, S. 2011. Growth-rate dependence reveals design principles of plasmid copy number control. *PLoS One,* 6**,** e20403.

KLUMPP, S. & HWA, T. 2008. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci U S A,* 105**,** 20245-50.

KLUMPP, S., ZHANG, Z. & HWA, T. 2009. Growth rate-dependent global effects on gene expression in bacteria. *Cell,* 139**,** 1366-75.

KNEEN, M., FARINAS, J., LI, Y. X. & VERKMAN, A. S. 1998. Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophys J,* 74**,** 1591-1599.

KOSURI, S., GOODMAN, D. B., CAMBRAY, G., MUTALIK, V. K., GAO, Y., ARKIN, A. P., ENDY, D. & CHURCH, G. M. 2013. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc Natl Acad Sci U S A,* 110**,** 14024-14029.

KUDLA, G., MURRAY, A. W., TOLLERVEY, D. & PLOTKIN, J. B. 2009. Coding-Sequence Determinants of Gene Expression in Escherichia coli. *Science,* 324**,** 255-258.

KWOK, R. 2010. Five hard truths for synthetic biology. *Nature,* 463**,** 288-90.

LAXMINARAYAN, R., DUSE, A., WATTAL, C., ZAIDI, A. K., WERTHEIM, H. F., SUMPRADIT, N., VLIEGHE, E., HARA, G. L., GOULD, I. M., GOOSSENS, H., GREKO, C., SO, A. D., BIGDELI, M., TOMSON, G., WOODHOUSE, W., OMBAKA, E., PERALTA, A. Q., QAMAR, F. N., MIR, F., KARIUKI, S., BHUTTA, Z. A., COATES, A., BERGSTROM, R., WRIGHT, G. D., BROWN, E. D. & CARS, O. 2013. Antibiotic resistance-the need for global solutions. *Lancet Infect Dis,* 13**,** 1057-98.

LEE, T. S., KRUPA, R. A., ZHANG, F., HAJIMORAD, M., HOLTZ, W. J., PRASAD, N., LEE, S. K. & KEASLING, J. D. 2011. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng,* 5**,** 12.

LEGUIA, M., BROPHY, J., DENSMORE, D. & ANDERSON, J. C. 2011. Automated Assembly of Standard Biological Parts. *Synthetic Biology, Pt B,* 498**,** 363-397.

LINSHIZ, G., STAWSKI, N., POUST, S., BI, C., KEASLING, J. D. & HILLSON, N. J. 2013. PaR-PaR laboratory automation platform. *ACS Synth Biol,* 2**,** 216-22.

LOOGER, L. L., DWYER, M. A., SMITH, J. J. & HELLINGA, H. W. 2003. Computational design of receptor and sensor proteins with novel functions. *Nature,* 423**,** 185-190.

LOU, C. B., STANTON, B., CHEN, Y. J., MUNSKY, B. & VOIGT, C. A. 2012. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat Biotechnol,* 30**,** 1137-+.

LU, T. K. & COLLINS, J. J. 2007. Dispersing biofilms with engineered enzymatic bacteriophage. *Proc Natl Acad Sci U S A,* 104**,** 11197-202.

LUNDHOLT, B. K., SCUDDER, K. M. & PAGLIARO, L. 2003. A simple technique for reducing edge effect in cell-based assays. *J Biomol Screen,* 8**,** 566-570.

MAIER, U., LOSEN, M. & BUCHS, J. 2004. Advances in understanding and modeling the gas-liquid mass transfer in shake flasks. *Biochemical Engineering Journal,* 17**,** 155-167.

MARTIN, L., CHE, A. & ENDY, D. 2009. Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. *PLoS One,* 4**,** e7569.

MEDINTZ, I. L., GOLDMAN, E. R., LASSMAN, M. E., HAYHURST, A., KUSTERBECK, A. W. & DESCHAMPS, J. R. 2005. Self-assembled TNT biosensor based on modular multifunctional surface-tethered components. *Anal Chem,* 77**,** 365-372.

MIZOGUCHI, H., MORI, H. & FUJIO, T. 2007. Escherichia coli minimum genome factory. *Biotechnology and Applied Biochemistry,* 46**,** 157-167.

MOON, T. S., CLARKE, E. J., GROBAN, E. S., TAMSIR, A., CLARK, R. M., EAMES, M., KORTEMME, T. & VOIGT, C. A. 2011. Construction of a Genetic Multiplexer to Toggle between Chemosensory Pathways in Escherichia coil. *J Mol Biol,* 406**,** 215-227.

MOSER, F., BROERS, N. J., HARTMANS, S., TAMSIR, A., KERKMAN, R., ROUBOS, J. A., BOVENBERG, R. & VOIGT, C. A. 2012. Genetic Circuit Performance under Conditions Relevant for Industrial Bioreactors. *ACS Synthetic Biology,* 1**,** 555-564.

MOSER, F., HORWITZ, A., CHEN, J., LIM, W. A. & VOIGT, C. A. 2013. Genetic Sensor for Strong Methylating Compounds. *ACS Synthetic Biology,* 2**,** 614-624.

MUTALIK, V. K., GUIMARAES, J. C., CAMBRAY, G., LAM, C., CHRISTOFFERSEN, M. J., MAI, Q.-A., TRAN, A. B., PAULL, M., KEASLING, J. D. & ARKIN, A. P. 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods,* 10**,** 354-360.

MUTALIK, V. K., QI, L., GUIMARAES, J. C., LUCKS, J. B. & ARKIN, A. P. 2012. Rationally designed families of orthogonal RNA regulators of translation. *Nat Chem Biol*.

MYERS, C. J., BARKER, N., JONES, K., KUWAHARA, H., MADSEN, C. & NGUYEN, N. P. D. 2009. iBioSim: a tool for the analysis and design of genetic circuits. *Bioinformatics,* 25**,** 2848-2849.

NEIDHARDT, F. C., BLOCH, P. L. & SMITH, D. F. 1974. Culture Medium for Enterobacteria. *Journal of Bacteriology,* 119**,** 736-747.

NI, L., TONTHAT, N. K., CHINNAM, N. & SCHUMACHER, M. A. 2013. Structures of the Escherichia coli transcription activator and regulator of diauxie, XylR: an AraC DNA-binding family member with a LacI/GalR ligand-binding domain. *Nucleic Acids Res,* 41**,** 1998-2008.

OLIVER, D. G., SANDERS, A. H., HOGG, R. D. & HELLMAN, J. W. 1981. Thermal-Gradients in Microtitration Plates - Effects on Enzyme-Linked Immunoassay. *J Immunol Methods,* 42**,** 195-201.

OLIVER, M. H., HARRISON, N. K., BISHOP, J. E., COLE, P. J. & LAURENT, G. J. 1989. A Rapid and Convenient Assay for Counting Cells Cultured in Microwell Plates - Application for Assessment of Growth-Factors. *J Cell Sci,* 92**,** 513-518.

PARK, C., CAMPBELL, J. L. & GODDARD, W. A., 3RD 1993. Design superiority of palindromic DNA sites for site-specific recognition of proteins: tests using protein stitchery. *Proc Natl Acad Sci U S A,* 90**,** 4892-6.

PARTS REGISTRY. n.d. *Regstry of Standard Biological Parts* [Online]. Available: www.partsregistry.org [Accessed 4 February 2014].

PATWARDHAN, R. P., LEE, C., LITVIN, O., YOUNG, D. L., PE'ER, D. & SHENDURE, J. 2009. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol,* 27**,** 1173-1175.

PECCOUD, J., ANDERSON, J. C., CHANDRAN, D., DENSMORE, D., GALDZICKI, M., LUX, M. W., RODRIGUEZ, C. A., STAN, G. B. & SAURO, H. M. 2011. Essential information for synthetic DNA sequences. *Nat Biotechnol,* 29**,** 22; discussion 22-3.

PILPEL, Y. 2011. Noise in biological systems: pros, cons, and mechanisms of control. *Methods Mol Biol,* 759**,** 407-25.

POSFAI, G., PLUNKETT, G., FEHER, T., FRISCH, D., KEIL, G. M., UMENHOFFER, K., KOLISNYCHENKO, V., STAHL, B., SHARMA, S. S., DE ARRUDA, M., BURLAND, V., HARCUM, S. W. & BLATTNER, F. R. 2006. Emergent properties of reduced-genome Escherichia coli. *Science,* 312**,** 1044-1046.

POTHOULAKIS, G., CERONI, F., REEVE, B. & ELLIS, T. 2013. The Spinach RNA Aptamer as a Characterization Tool for Synthetic Biology. *ACS Synthetic Biology*.

POTZKEI, J., KUNZE, M., DREPPER, T., GENSCH, T., JAEGER, K. E. & BUCHS, J. 2012. Real-time determination of intracellular oxygen in bacteria using a genetically encoded FRET-based biosensor. *BMC Biol,* 10.

QI, L., HAURWITZ, R. E., SHAO, W. J., DOUDNA, J. A. & ARKIN, A. P. 2012. RNA processing enables predictable programming of gene expression. *Nat Biotechnol,* 30**,** 1002-+.

QUAN, J. Y. & TIAN, J. D. 2009. Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways. *PLoS One,* 4.

QUINN, J., BEAL, J., BHATIA, S., CAI, P., CHEN, J., CLANCY, K., HILLSON, N., GALDZICKI, M., MAHESHWARI, A. & POCOCK, M. 2013. Synthetic Biology Open Language Visual (SBOL Visual), version 1.0. 0.

RAJENDRAN, M. & ELLINGTON, A. D. 2008. Selection of fluorescent aptamer beacons that light up in the presence of zinc. *Anal Bioanal Chem,* 390**,** 1067-1075.

RHODIUS, V. A., SEGALL-SHAPIRO, T. H., SHARON, B. D., GHODASARA, A., ORLOVA, E., TABAKH, H., BURKHARDT, D. H., CLANCY, K., PETERSON, T. C., GROSS, C. A. & VOIGT, C. A. 2013. Design of orthogonal genetic switches based on a crosstalk map of rs, anti-rs, and promoters. *Mol Syst Biol,* 9.

RO, D. K., PARADISE, E. M., OUELLET, M., FISHER, K. J., NEWMAN, K. L., NDUNGU, J. M., HO, K. A., EACHUS, R. A., HAM, T. S., KIRBY, J., CHANG, M. C., WITHERS, S. T., SHIBA, Y., SARPONG, R. &

KEASLING, J. D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature,* 440**,** 940-3.

RODRIGUEZ-VILLALON, A., PEREZ-GIL, J. & RODRIGUEZ-CONCEPCION, M. 2008. Carotenoid accumulation in bacteria with enhanced supply of isoprenoid precursors by upregulation of exogenous or endogenous pathways. *J Biotechnol,* 135**,** 78-84.

RONEN, M., ROSENBERG, R., SHRAIMAN, B. I. & ALON, U. 2002. Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci U S A,* 99**,** 10555-10560.

ROSENFELD, N., YOUNG, J. W., ALON, U., SWAIN, P. S. & ELOWITZ, M. B. 2007. Accurate prediction of gene feedback circuit behavior from component properties. *Mol Syst Biol,* 3**,** 143.

ROSSGER, K., CHARPIN-EL-HAMRI, G. & FUSSENEGGER, M. 2013. A closed-loop synthetic gene circuit for the treatment of diet-induced obesity in mice. *Nat Commun,* 4.

ROYAL ACADEMY OF ENGINEERING 2009. Synthetic Biology: scope, applications and implications.

SALIS, H. M. 2011. The ribosome binding site calculator. *Methods Enzymol,* 498**,** 19-42.

SALIS, H. M., MIRSKY, E. A. & VOIGT, C. A. 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol,* 27**,** 946-50.

SALMON, K., HUNG, S. P., MEKJIAN, K., BALDI, P., HATFIELD, G. W. & GUNSALUS, R. P. 2003. Global gene expression profiling in Escherichia coli K12. The effects of oxygen availability and FNR. *J Biol Chem,* 278**,** 29837-55.

SCOTT, M. & HWA, T. 2011. Bacterial growth laws and their applications. *Curr Opin Biotechnol,* 22**,** 559-65.

SHARMA, P., ASAD, S. & ALI, A. 2013. Bioluminescent bioreporter for assessment of arsenic contamination in water samples of India. *Journal of Biosciences,* 38**,** 251-258.

SHEKARCHI, I. C., SEVER, J. L., LEE, Y. J., CASTELLANO, G. & MADDEN, D. L. 1984. Evaluation of Various Plastic Microtiter Plates with Measles, Toxoplasma, and Gamma-Globulin Antigens in Enzyme-Linked Immunosorbent Assays. *J Clin Microbiol,* 19**,** 89-96.

SILVA-ROCHA, R., MARTINEZ-GARCIA, E., CALLES, B., CHAVARRIA, M., ARCE-RODRIGUEZ, A., DE LAS HERAS, A., PAEZ-ESPINO, A. D., DURANTE-RODRIGUEZ, G., KIM, J., NIKEL, P. I., PLATERO, R. & DE LORENZO, V. 2013. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res,* 41**,** D666-75.

SOLE, R. V., MUNTEANU, A., RODRIGUEZ-CASO, C. & MACIA, J. 2007. Synthetic protocell biology: from reproduction to computation. *Philos Trans R Soc Lond B Biol Sci,* 362**,** 1727-39.

SONG, S. G. & PARK, C. 1997. Organization and regulation of the D-xylose operons in Escherichia coli K-12: XylR acts as a transcriptional activator. *Journal of Bacteriology,* 179**,** 7025-7032.

STANTON, B. C., NIELSEN, A. A., TAMSIR, A., CLANCY, K., PETERSON, T. & VOIGT, C. A. 2013. Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat Chem Biol*.

STUEBER, D. & BUJARD, H. 1982. Transcription from Efficient Promoters Can Interfere with Plasmid Replication and Diminish Expression of Plasmid Specified Genes. *Embo Journal,* 1**,** 1399-1404.

SUN, Z. Z., YEUNG, E., HAYES, C. A., NOIREAUX, V. & MURRAY, R. M. 2013. Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an Escherichia coli Based TX-TL Cell-Free System. *ACS Synthetic Biology*.

TABOR, J. J., SALIS, H. M., SIMPSON, Z. B., CHEVALIER, A. A., LEVSKAYA, A., MARCOTTE, E. M., VOIGT, C. A. & ELLINGTON, A. D. 2009. A synthetic genetic edge detection program. *Cell,* 137**,** 1272-81.

TAMSIR, A., TABOR, J. J. & VOIGT, C. A. 2011. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature,* 469**,** 212-5.

TEMME, K., HILL, R., SEGALL-SHAPIRO, T. H., MOSER, F. & VOIGT, C. A. 2012a. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res,* 40**,** 8773-8781.

TEMME, K., ZHAO, D. H. & VOIGT, C. A. 2012b. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proc Natl Acad Sci U S A,* 109**,** 7085-7090.

TRINH, C. T., UNREAN, P. & SRIENC, F. 2008. Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. *Appl Environ Microbiol,* 74**,** 3634-3643.

UMENHOFFER, K., FEHER, T., BALIKO, G., AYAYDIN, F., POSFAI, J., BLATTNER, F. R. & POSFAI, G. 2010. Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb Cell Fact,* 9**,** 38.

UNDEN, G., ACHEBACH, S., HOLIGHAUS, G., TRAN, H. G., WACKWITZ, B. & ZEUNER, Y. 2002. Control of FNR function of Escherichia coli by O2 and reducing conditions. *J Mol Microbiol Biotechnol,* 4**,** 263-8.

VIDAL-AROCA, F., GIANNATTASIO, M., BRUNELLI, E., VEZZOLI, A., PLEVANI, P., MUZI-FALCONI, M. & BERTONI, G. 2006. One-step high-throughput assay for quantitative detection of beta-galactosidase activity in intact Gram-negative bacteria, yeast, and mammalian cells. *Biotechniques,* 40**,** 433-+.

WANG, B., KITNEY, R. I., JOLY, N. & BUCK, M. 2011. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat Commun,* 2**,** 508.

WANG, H. H., ISAACS, F. J., CARR, P. A., SUN, Z. Z., XU, G., FOREST, C. R. & CHURCH, G. M. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature,* 460**,** 894-U133.

WEGERER, A., SUN, T. Q. & ALTENBUCHNER, J. 2008. Optimization of an E. coli L-rhamnose-inducible expression vector: test of various genetic module combinations. *BMC Biotechnol,* 8.

WESTERS, H., DORENBOS, R., VAN DIJL, J. M., KABEL, J., FLANAGAN, T., DEVINE, K. M., JUDE, F., SEROR, S. J., BEEKMAN, A. C., DARMON, E., ESCHEVINS, C., DE JONG, A., BRON, S., KUIPERS, O. P., ALBERTINI, A. M., ANTELMANN, H., HECKER, M., ZAMBONI, N., SAUER, U., BRUAND, C., EHRLICH, D. S., ALONSO, J. C., SALAS, M. & QUAX, W. J. 2003. Genome engineering reveals large dispensable regions in Bacillus subtilis. *Mol Biol Evol,* 20**,** 2076-90.

WIELAND, M. & FUSSENEGGER, M. 2012. Engineering Molecular Circuits Using Synthetic Biology in Mammalian Cells. *Annu Rev Chem Biomol Eng*.

WILKS, J. C. & SLONCZEWSKI, J. L. 2007. pH of the cytoplasm and periplasm of Escherichia coli: rapid measurement by green fluorescent protein fluorimetry. *J Bacteriol,* 189**,** 5601-7.

WREGHITT, T. G., NAGINGTON, J. & GRAY, J. 1982. An Elisa Test for the Detection of Antibodies to Legionella-Pneumophila. *J Clin Pathol,* 35**,** 657-660.

XIE, Z., WROBLEWSKA, L., PROCHAZKA, L., WEISS, R. & BENENSON, Y. 2011. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science,* 333**,** 1307-1311.

YOON, J. & RILEY, M. R. 2009. Grand challenges for biological engineering. *J Biol Eng,* 3**,** 16.

ZASLAVER, A., KAPLAN, S., BREN, A., JINICH, A., MAYO, A., DEKEL, E., ALON, U. & ITZKOVITZ, S. 2009. Invariant Distribution of Promoter Activities in Escherichia coli. *PLoS Comput Biol,* 5.

ZHANG, X. G. & BREMER, H. 1995. Control of the Escherichia-Coli Rrnb P1 Promoter Strength by Ppgpp. *Journal of Biological Chemistry,* 270**,** 11181-11189.

ZUCCA, S., PASOTTI, L., MAZZINI, G., DE ANGELIS, M. G. & MAGNI, P. 2012. Characterization of an inducible promoter in different DNA copy number conditions. *BMC Bioinformatics,* 13 Suppl 4**,** S11.

ZWIETERING, M. H., JONGENBURGER, I., ROMBOUTS, F. M. & VAN 'T RIET, K. 1990. Modeling of the bacterial growth curve. *Appl Environ Microbiol,* 56**,** 1875-81.

10 Appendices

10.1 Appendix A: SVc Promoter-less sequences

Plasmid	Full verified Sequence
SVc	CGCGCCGGTAGCGTTCAGACTCCTCGACGCATCTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAA TCACCAACTGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTATGGCTGGATGATGGGGCGATTCG GGCCTGGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTT GGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATG TGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTT ACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCG TTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGAC TATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATT CCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGT ATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATG CAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAA ACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACC TTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTCTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAA GATCATCTTATTAAGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGA CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCTGACGCAGCCGAGTTCCGATGGCCGGCCCTCGAGTCCCGTCAA GTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACT GCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAG GCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCC CCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTAT GCATTTCCTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTC ATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAAC CGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTT CCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAA ATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAAC TCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCC ATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTGGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCC CTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGA TTTTGAGACACAACGTGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGGGTAACTCCCGCTGTAGAC GATTTCGCGATTTAAATAAATAATAAAAAAGCCGGATTAATAATCTGGCTTTTTATATTCTCTCTCTAGTATATAAACGC AGAAAGGCCCACCCGAAGGTGAGCCAGTGTGAAGAGGTTCAGCCCGTAGTCCTTATTAAGCACCGGTGGAGAGGTTC AGCCCGTAGTCCTTATTAAGCACCGGTGGAGAGGTTCAGCCCGTAGTCCTTATTAAGCACCGGTGGAGTGACGACCTT

	CAGCACGTTCGTACTGTTCAACGATGGTGTAGTCTTCGTTGTGGGAGGTGATGTCCAGTTTGATGTCGGTTTTGTAAGC ACCCGGCAGCTGAACCGGTTTTTTAGCCATGTAGGTGGTTTTAACTTCAGCGTCGTAGTGACCACCGTCTTTCAGTTTCA GACGCATTTTGATTTCACCTTTCAGAGCACCGTCTTCCGGGTACATACGTTCGGTGGAAGCTTCCCAACCCATGGTTTTT TTCTGCATAACCGGACCGTCGGACGGGAAGTTGGTACCACGCAGTTTAACTTTGTAGATGAACTCACCGTCTTGCAGG GAGGAGTCCTGGGTAACGGTAACAACACCACCGTCTTCGAAGTTCATAACACGTTCCCATTTGAAACCTTCCGGGAAG GACAGTTTCAGGTAGTCCGGGATGTCAGCCGGGTGTTTAACGTAAGCTTTGGAACCGTACTGGAACTGCGGGGACAG GATGTCCCAAGCGAACGGCAGCGGACCACCTTTGGTAACTTTCAGTTTAGCGGTCTGGGTACCTTCGTACGGACGACC TTCACCTTCACCTTCGATTTCGAACTCGTGACCGTTAACGGAACCTTCCATACGAACTTTGAAACGCATGAACTCTTTGA TAACGTCTTCGGAGGAAGCCATCTAGTATTTCTCCTCTTTCTCTAGTATGGAGCCGAATACCTGCCGTGCTAGCATAATA CCTAGGACTGAGCTAGCTGTAAAAGACCTTTACGCCGCTGGAGACTAGTCACAGCTAACACCACGTCGTCCCTATCTGC TGCCCTAGGTCTATGAGTGGTTGCTGGATAACGAATTCGCGGCCGCTTCTAGAGAAAGAGGAGAAATACTAGATGCGT AAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTG TCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTG TTCCATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCGAGATACCCAGATCATATGAAACAGCAT GACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAG ACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGAT GGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAAT GGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAAT ACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAA CGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAA ATAATAATACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTT GTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGG
SVc-X	CGCGCCGGTAGCGTTCAGACTCCTCGACGCATCTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAA TCACCAACTGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTATGGCTGGATGATGGGGCGATTCG GGCCTGGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTT GGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATG TGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTT ACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCG TTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGAC TATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATT CCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGT ATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATG CAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAA ACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACC TTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTCTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAA GATCATCTTATTAAGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGA CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCTGACGCAGCCGAGTTCCGATGGCCGGCCCTCGAGTCCCGTCAA

	GTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACT GCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAG GCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCC CCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTAT GCATTTCCTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTC ATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAAC CGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTT CCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAA ATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAAC TCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCC ATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTGGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCC CTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGA TTTTGAGACACAACGTGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGGGTAACTCCCGCTGTAGAC GATTTCGCGATTTAAATAAATAATAAAAAAGCCGGATTAATAATCTGGCTTTTTATATTCTCTCTCTAGTATATAAACGC AGAAAGGCCCACCCGAAGGTGAGCCAGTGTGAAGAGGTTCAGCCCGTAGTCCTTATTACAACATGACCTCGCTATTTA CATCGCGATACTCTTTTGGCGTCGTGTCATATGCTTTTTTAAAAACAGAGTAGAAATATTGCAGCGATGGATAACCGCA CATTTGCGATATCTCATTGATCGACAAGGTGGTTGAAATCAGCAGACTGCGCGCTTTCTCCAGCTTCTCGGCATGAATC ATGGCATGGATGGTTTCACCCACCTCTTCTTTAAAACGCTTCTCAAGATTGGAGCGCGAGATCCCGACCGCATCCAGTA CCTGATCCACTTTAATCCCTTTACAGGCGTGATTACGAATGTAATGCATGGCCTGAATAACGGCGGGATCGGTCAGCGA GCGATAATCTGTTGAGCGCCGTTCAATGACGCGAACTGGTGGGACCAAAATTCGCTGTAGCGGCATTTCTTCTTTATCT AATAATCGATGCAACAGTTTTGCCGCCTGATAGCCCATTTGCCGCGCGCCCTGAGCGACCGAAGAAAGGGCGACACGC GACAGATAGCGGGTCAGTTCTTCGTTATCGATGCCAATCACGCATAATTTTTCCGGTACGGGAATATGTAGATGTTCAC ATACTTGCAGAATATGCCGCGCTCGGGCGTCAGTAACGGCAATAATCCCGGTTTGCGGTGGTAGCGTTTGTAGCCAGT CTGCCAGCCGATTTTGCGCGTGTTGCCAGTTCTCTGGCGCGGTTTCTAACCCCTGATAAACCACTCCGCGATACTTTTCT TCGGCGACAAGCTGACGAAATGCATATTCGCGCTCAGTGGCCCAACGTTTGCCGCTTGATTCCGGAAGACCATAAAAA GCAAAGCGGTTAACGCCTTTCTCTTTTAAATGCAAAAATGCGCTTTCAACCAGCGCATAGTTATCGGTGGCAATGTAAT GAACGGGTGGGTAACTTTCTGCAAGGTGATACGAGCCGCCAACCCCAACAATGGGGACGTCGACATCAGCCAGCGCT TGCTCGATCTGTTTGTCGTCGAAGTCGGCAATGACGCCATCTCCTAACCAGTCCTTGATTTTATCAATGCGGGCGCGGA AATCTTCTTCAATGAAAATATCCCATTCCGATTGTGACGCCTGTAAATATTCCCCTACGCCTTCTACTACCTGCCGGTCAT AGGCTTTATTGGCATTGAACAGTAATGTGATGCGGTGACGTTTAGTAAACAAGGTGGGGATATTGGATGAGTTACGTG GAGCCGAATACCTGCCGTGCTAGCATAATACCTAGGACTGAGCTAGCTGTAAAAGACCTTTACGCCGCTGGAGACTAG TCACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTTGCTGGATAACGAATTCGCGGCCGCTT CTAGAGAAAGAGGAGAAATACTAGATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTA GATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTT AAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTT GCGAGATACCCAGATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACT ATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATC GAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATG

	TATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCG TTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCC ACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGA TTACACATGGCATGGATGAACTATACAAATAATAATACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAA GACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCC TTTCTGCGTTTATAGG
SVc- S25II	CGCGCCGGTAGCGTTCAGACTCCTCGACGCATCTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAA TCACCAACTGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTATGGCTGGATGATGGGGCGATTCG GGCCTGGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTT GGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATG TGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTT ACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCG TTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGAC TATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATT CCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGT ATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATG CAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAA ACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACC TTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTCTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAA GATCATCTTATTAAGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA AGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGA CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCTGACGCAGCCGAGTTCCGATGGCCGGCCCTCGAGTCCCGTCAA GTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACT GCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAG GCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCC CCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTAT GCATTTCCTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTC ATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAAC CGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTT CCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAA ATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAAC TCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCC ATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTGGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCC CTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGA TTTTGAGACACAACGTGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGGGTAACTCCCGCTGTAGAC GATTTCGCGATTTAAATAAATAATAAAAAAGCCGGATTAATAATCTGGCTTTTTATATTCTCTCTCTAGTATATAAACGC AGAAAGGCCCACCCGAAGGTGAGCCAGTGTGAAGAGGTTCAGCCCGTAGTCCTTATTATTGCAGAAAGCCATCCCGTC CCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGTGGTTACTGTCGCTGAATCCAC

	AGCGATAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGTCGCAGGCGGTTCA GGTATCGCTGAGGCGTCAGTCCCGTTTGCTGCTTAAGCTGCCGATGTAGCGTACGCAGTGAAAGAGAAAATTGATCCG CCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCAGCCAGGCCAGAAGCAAGTTGAGACGTGATGCGCTGT TTTCCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAGTAATTGCATAAACAAGATCTCGCGACTGGCGGTCGA GGGTAAATCATTTTCCCCTTCCTGCTGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGC GCCAGTGAGACGGATACTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCG GCGAGCGATACAGCACATTGGTCAGACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAAC AGACCGTGCCACCGGTGATGGTATAGGGCTGCCCATTAAACACATGAATACCCGTGCCATGTTCGACAATCACAATTTC ATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGGGAGCCGGGGTTCTATCGCCACGGACGCGTTACCAGACG GAAAAAAATCCACACTATGTAATACGGTCATAGTAGCCCCCCCTCTGCTTTGGCCCTGGAGCCGAATACCTGCCGTGCT AGCATAATACCTAGGACTGAGCTAGCTGTAAAAGACCTTTACGCCGCTGGAGACTAGTCACAGCTAACACCACGTCGT CCCTATCTGCTGCCCTAGGTCTATGAGTGGTTGCTGGATAACGAATTCGCGGCCGCTTCTAGAGAAAGAGGAGAAATA CTAGATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCAC AAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAA AACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCGAGATACCCAGATCATATG AAACAGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGG AACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTA AAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAAC AAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATC AACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAA GATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAA CTATACAAATAATAATACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATC TGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGG

Table 10.1 Full sequences for the backbone plasmids used in this research.

All 3 plasmids were fully sequenced to check for mutations plasmid to plasmid. The individual modules are shown and annotated in table 6.3.

10.2 Appendix B: Alternate view of the dilution optimisation 3D plot

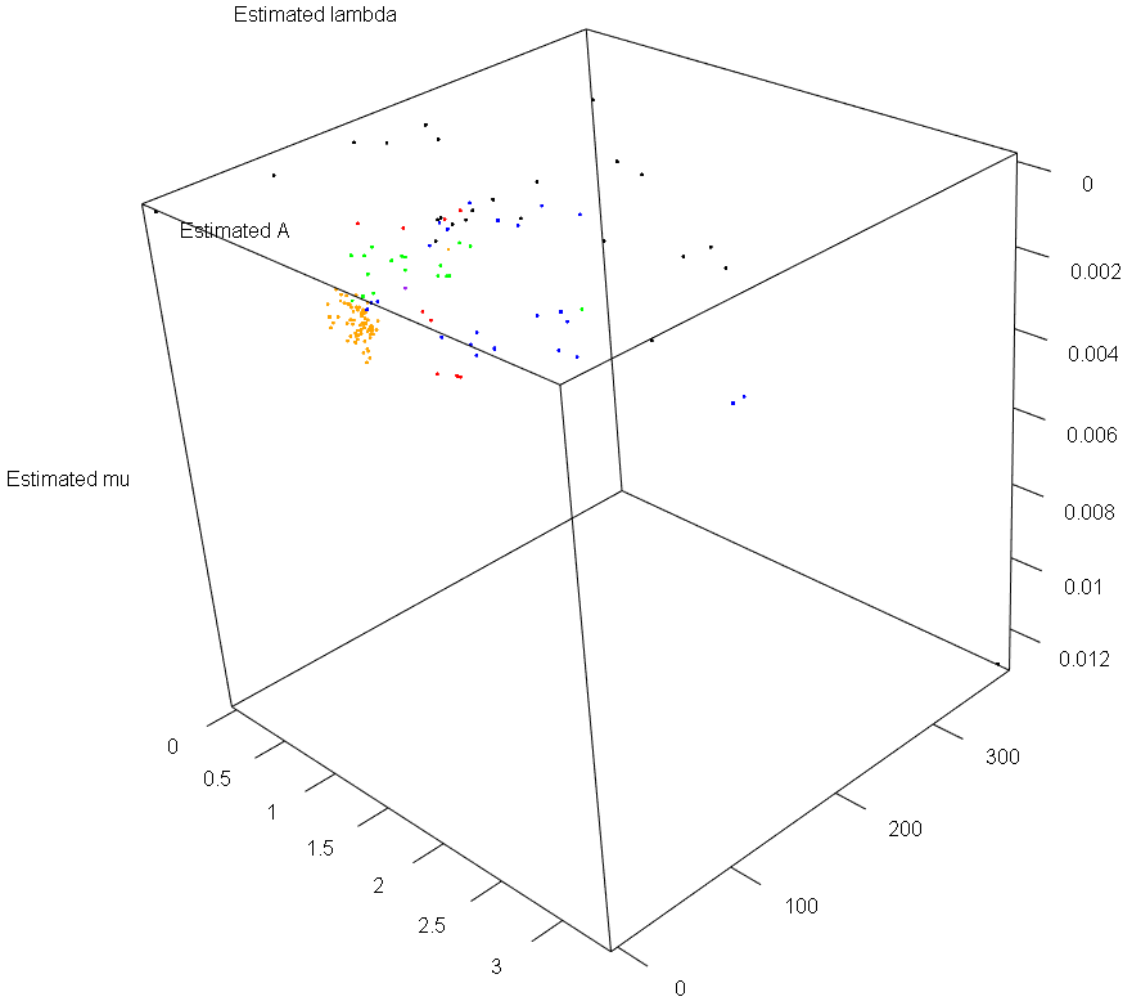


Figure 10.1 Alternative view of the 3D plot from figure 6.13

The growth of the SVc containing cells in the dilution optimisation experiment was tracked and analysed by the scripts of C. Ainsworth which fit these results to a model of *E. coli* growth with parameters of lag time – lambda, growth rate - mu and carrying capacity – A. These parameters were plotted for various experiments in chronological order of Black, Red, Blue, Green and finally orange where optimisation had been completed and the samples are tightly clustered for growth characteristics (n > 6).