Development of a synthetic biology approach to targeted directed evolution of proteins *in vivo*

Thomas Adie

Imperial College London
Department of Life Sciences
Centre for Synthetic Biology and Innovation
Submitted for the Degree of Doctor of Philosophy

October 2013

Supervisors: Dr. Geoff Baldwin (Principal)
Prof. Paul Freemont
Prof. Richard Kitney

Abstract

This thesis describes the development of a novel platform for targeted directed evolution, designed to operate entirely *in vivo*. The system comprises a fusion of T7 RNA polymerase and activation-induced deaminase (AID); targeting is achieved by the placement of the sequence of interest under the control of a T7 promoter, whereby transcription by the polymerase exposes the DNA to mutation by the AID moiety of the fusion. The localisation effect serves to target mutation to the area downstream of the promoter, and increase the rate of that mutation compared to non-directed background activity. The system, and appropriate controls and targets, are constructed and tested in a plasmid-based experimental work-flow. The targets are, further, integrated into the genome of *Escherichia coli* to allow high-throughput analysis of the mutation rate of a single copy target. Nucleotide sequencing is used to confirm both enhanced mutagenesis of the system, and a high degree of targeting. The system is then applied to a test case, diversification of a transcription factor (LasR from *Pseudomonas aeruginosa*, encoded by the *lasR* gene), with an eye to producing an orthogonal signal/response pair with promoter P_{Las}. A logic gate-based filter is designed and constructed to allow tight moderation of a feedback loop to control the mutator, allowing it to be 'shut off' once desired function is exhibited by the target protein.

Declaration

I hereby certify that this text represents my own work and that all else is appropriately referenced.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgments

I would like to acknowledge first my supervisor, Dr. Geoff Baldwin and the members of his laboratory at Imperial College for their guidance, and insight – J.S. particularly is an outstanding human being.

This thesis would not exist without the encouragement and support of my family and friends, to whom I owe everything. Thank you for listening to me when I was excited about good ideas, shooting down bad ones, and pointing out when I should be doing something I wasn't. I am particularly grateful to K.Y.W in that regard.

Thanks finally to Rui Rodrigues, Songyuan Li and the many other enthusiastic students I had the pleasure of supervising, for their input and discussions.

Table of Contents

1. Introduction 9
1.1. Directed evolution
1.1.1. *Pure in vitro methods* 14
1.1.2. *Mixed in vitro/in vivo methods* 15
1.1.3. *Pure in vivo methods* 17
1.2. Developing a system for in vivo directed evolution
1.2.1. *Activation-induced deaminase (AID)* 23
1.2.2. *T7 RNA polymerase (T7 RNAP)* 25
1.2.3. *Coupling DNA deamination to specific transcriptional activity* 27
1.2.4. *Targeted deamination in context* 29
1.2.5. *Aims of the project* 32
2. Constructing a system for in vivo directed evolution 33
2.1.1. *Aims* 33
2.2. Experimental strategy
2.2.1. *Rifampicin reversion* 34
2.2.2. *Controls required for assessing targeted mutation* 35
2.2.3. *Masking as a consideration in gain of function assays* 38
2.2.4. *Fluctuation analysis* 42
2.2.5. *Culture conditions* 45
2.3. Results
2.3.1. *Construction of mutator protein* 48
2.3.2. *Construction of controls* 49
2.3.3. *Construction of GFP and XFP targets* 51
2.3.4. *Confirmation of activity of devices* 53
2.4. Discussion
2.4.1. *Summary of strategy* 55
2.4.2. *Construction* 55
2.4.3. *AID retains functionality in the fusion* 56
3. Multi copy target analysis 57
3.1.1. *Aims* 57
3.1.2. *Plasmid-based systems require segregation of targets* 57
3.2. Results
3.2.1. *Automation* 59
3.3. Discussion
3.3.1. *Adaptation to automation* 64

3.3.2. *Data quality and limitations* ... 65
4. Single copy target analysis ... 70
4.1.1. *Aims* ... 70
4.2. Results
4.2.1. *Genome integration of targets for single-copy analysis* ... 71
4.2.2. *Analysis of mutation of single-copy targets by flow cytometry* ... 75
4.2.3. *Sequencing of genome-integrated GFP target* ... 80
4.3. Discussion
4.3.1. *Production of strains with genome-integrated targets* ... 84
4.3.2. *Flow cytometry* ... 85
4.3.3. *Sequencing* ... 87
5. Implementation of the system toward an orthogonal signal/response pair 90
5.1.1. *Aims* ... 90
5.2. System overview
5.2.1. *The mutation target: a quorum-sensing system* ... 91
5.2.2. *The control circuit* ... 96
5.2.3. *Leakiness* ... 99
5.2.4. *Bacterial logic gates, circuit design* ... 102
5.3. Results
5.3.1. *Construction* ... 108
5.3.2. *Testing* ... 110
6. Conclusion ... 119
7. Methods ... 122
7.1. Tables
7.1.1. *Strains* ... 122
7.1.2. *Plasmids* ... 123
7.1.3. *Primers* ... 124
7.2. DNA construction and manipulation
7.2.1. *Plasmid isolation* ... 126
7.2.2. *PCR* ... 127
7.2.3. *Colony PCR* ... 128
7.2.4. *Site-directed mutagenesis* ... 128
7.2.5. *In-Fusion assembly* ... 128
7.2.6. *Gibson isothermal assembly* ... 129
7.2.7. *CPEC* ... 129
7.2.8. *BioBrick cloning* ... 129
7.2.9. *Digestion* ... 130

7.2.10. *Dephosphorylation* 130
7.2.11. *Ligation* 130
7.2.12. *CRIM integration* 130
7.3. Quantification
7.3.1. *DNA concentration & sequence verification* 131
7.3.2. *Gel electrophoresis/extraction* 131
7.3.3. *SDS-PAGE* 132
7.4. Cultures & propagation
7.4.1. *Cultures* 133
7.4.2. *Chemically-competent cells/chemical transformation* 134
7.4.3. *Electro-competent cells/electroporation* 135
7.4.4. *Glycerol stocks* 136
7.5. Automation
7.5.1. *Preparation* 137
7.5.2. *Plasmid purification* 137
7.5.3. *Transformation* 138
7.5.4. *Semi-automated continuous culture* 138
7.6. Analysis
7.6.1. *Rifampicin reversion assays* 139
7.6.2. *Fluorimetry* 140
7.6.3. *Flow cytometry* 141
8. Bibliography 142
9. Appendices 154
9.1. Plate counts for automated plasmid segregation
9.1.1. *Automated run* 154
9.1.2. *Semi-automated run* 155
9.2. Automation script
9.2.1. *Miniprep & transformation* 156
9.3. Details of mutations detected by sequencing (4.3.3)
9.3.1. *Mutation information* 157
9.3.2. *Trace files* 158
9.4. DNA sequences of construct parts
9.4.1. *Sequence information* 160

Index of Figures

Figure 1: The directed evolution cycle 13
Figure 2: Directed evolution explores a 'sequence space' 13
Figure 3: Schematic representation of the system 28
Figure 4: Masking effect of gene knockout, given one reversion and one knockout base 39
Figure 5: Effect of number of knockout bases on reversion frequency 40
Figure 6: The device plasmids; the fusion protein and two controls, on a high-copy backbone 50
Figure 7: Agarose gel showing the products of a colony PCR confirming successful construction of device plasmids 50
Figure 8: Agarose gel showing the products of a colony PCR confirming successful construction of GFP target plasmids 52
Figure 9: Agarose gel showing the products of an analytical digest confirming successful production of XFP by SDM 52
Figure 10: The target plasmids 52
Figure 11: Rifampicin reversion assay 54
Figure 12: Colony counts from automated plasmid isolation and transformation 61
Figure 13: Colony counts from semi-automated plasmid isolation and transformation 63
Figure 14: GFP target cloned into pAH153 for integration onto the E. coli genome . . 73
Figure 15: Agarose gel showing the products of an analytical PCR demonstrating successful integration of GFP/XFP 73
Figure 16: Time-course flow cytometry data showing loss of function profiles for the device plasmids 78
Figure 17: Time-course flow cytometry data showing loss of function profile over 32 hours for the unfused device plasmid 79
Figure 18: Mutations in the GFPmut3b sequence induced by exposure to the mutator protein 82
Figure 19: Map of observed mutation events along the GFPmut3b target gene 83
Figure 20: Response profile of the PLux system to a range of AHL molecules 95
Figure 21: The HrpR/S-based AND gate 107
Figure 22: Schematic of AND gate and control circuitry 107
Figure 23: The PLas diversification and reporting devices, with and without AND gate 109
Figure 24: Induction of the LasR/PLas system by AHL failed 112

Figure 25: Growth curve for LasR testing constructs 112
Figure 26: Fluorescence detection is validated with a positive control 112
Figure 27: The original LasR system responds to induction with 1 µM AHL 114
Figure 28: PLac-T7RNAP induced with IPTG shows strong expression 114
Figure 29: SDS-PAGE gel confirming expression of LasR 117
Figure 30: Flow cytometry data showing function of the LasR circuit, and response to induction 118
Figure 31: Flow cytometry showing the LasR circuit and AND gate circuit responding to AHL induction 121
Figure 32: Colony counts of the semi-automated multi-copy mutation assays 154
Figure 33: Colony counts of the automated multi-copy mutation assays 155
Figure 34: The automation script for combined plasmid purification and transformation 156
Figure 35: Mutation trace files 159

Index of Tables

Table 1: Requirements for an accurate determination of m by fluctuation analysis, with actions taken to attempt to satisfy them in the design of experiments 47
Table 2: Experimental strategy; questions and the experiments designed to answer them 55
Table 3: Key strains used in this work 122
Table 4: Key plasmids used in this work 123
Table 5: Key primers used in this work 125
Table 6: Further information on mutation events 157
Table 7: DNA sequences of parts and plasmids used in this work 162

Index of Abbreviations

AHL – Acyl homoserine lactone
AID – Activation induced deaminase
BER – Base-excision repair
CFU – Colony-forming units
CPEC – Circular polymerase extension cloning
CRIM – Conditional replication integration and modular
ddH_2O – double-distilled water
FACS – Fluorescence-activated cell sorting
GFP – Green fluorescent protein
MAGE – Multiplex automated genome engineering
PACE – Phage-assisted continuous evolution
PCR – Polymerase chain reaction
RBS – Ribosome binding site
SDM – Site-directed mutagenesis
ss/dsDNA – Single-stranded/double-stranded DNA
T7 RNAP – Type 7 bacteriophage RNA polymerase

1. Introduction

Life on Earth, with all its intricacies of function, with all its divergences and convergences, has been produced by a fundamental process: Evolution. Variation introduced by the constant, incremental mutation of the genetic code at the heart of all organisms allows new functions to develop from old ones or entirely parallel ones to appear. The elegant simplicity and power of the process make it only natural that we should attempt to control and direct it.

This thesis presents a new approach to harnessing the potential power of evolutionary processes. The introduction will first define directed evolution and provide an overview of current techniques and approaches that are employed in that field, then progress to a review of studies on Activation Induced Deaminase (AID) and T7 RNA polymerase (T7 RNAP) which are enzymes that have relevance to the work presented.

1.1. Directed evolution

The attempt to produce new function, or an improvement in current function, in a gene or protein within the laboratory via mimicry of the process of natural selection is termed directed evolution. Applying the principles of the natural process, the genetic sequence encoding the target is altered (diversified) either by direct mutation of the DNA or some other method of introducing variability (for instance, gene shuffling).

Screening of the sequences thus produced, with selection of candidates based on proximity to the target phenotype, allows promising candidates to be subjected to further rounds of selection. These cycles of mutation/variation, screening and selection allow the experimenter to actively "evolve" the genes in question toward the desired phenotype.

Directed evolution-based strategies are most commonly employed in protein engineering projects, when the aim is to achieve to enhance efficiency of a desired process; proteins may be engineered for increased activity, improved stability, greater specificity or a relaxation of same, or some other desirable attribute (Romero 2009). In the ideal case, a protein would be rationally designed and constructed to incorporate the desired improvements – this requires comprehensive data about the structure, folding and molecular interactions within the target protein, however, which are rarely available. The field of computational protein design (reviewed in Hilvert 2013; Kiss 2013) aims to rationally engineer proteins, but although great advances have been made in optimising existing proteins and creating new functionality (Tinberg 2013) the process is far from solved (Schreier 2009). Though protein design projects are often complemented with directed evolution approaches (Röthlisberger 2008), they fall outside the scope of this project.

Directed evolution may also be used to 'tune' genetic circuits, or adjust regulatory elements within a stretch of DNA incorporating one or more genes (Cobb 2013). Even a circuit designed and built using well-characterised elements, about which many of the defining parameters are known, is subject to unpredictable behaviour as a result of host compatibility, noise, and changes in cellular environment (among other confounding factors). This is a problem particularly in the field of Synthetic Biology, which – via standardisation of techniques and characterisation of parts and devices – aims to make biology easier to engineer; more predictable behaviours leading to more reproducible results. Emergent effects arising from the combining of a number of parts, even well-known parts, conspire to reduce predictability of a system as the complexity of that system increases (Canton 2008). Directed evolution of circuits, while divergent from the 'standardisation' aim of the field, could help tune systems toward their predicted and desired operation post-construction – for instance, to maximise the output of a reporter gene within a complex biosensor. Variants or libraries of parts can also be produced, allowing for more flexibility of design; either by offering alternatives in e.g. promoter strength, or by creation of orthogonal signalling systems, for example, to eliminate 'cross-talk' (where two such systems interfere with each others' operation) and allow repeated use of a useful control element (Morey 2012).

Current methods for directed evolution generally follow an archetypal cycle, shown in figure 1. Typically, rounds of mutation are carried out *in vitro* prior to insertion of DNA into cells for screening and selection. There is an inherent inefficiency to this process, however, as the number of DNA sequences successfully incorporated into cells is a reduction of several orders of magnitude from the number produced in the mutation step. Thus, only a few thousand sequences will reach the screening step from an original pool of millions, or even billions of genetic variants. Further, those candidates which most closely approach the phenotype being screened for each round must be favoured in the selection process, meaning that escaping 'local peaks' and crossing fitness valleys to a more effective variant is difficult (Romero 2009); see figure 2 for an illustration of such a case. Extended diversification times are required to escape a local peak, but the efficiency limitation introduced by transformation means that successful incorporation of a superior variant of a gene into a cell for selection is unlikely.

Current methods for directed evolution usually involve one or more *in vitro* steps. A review of both those methods that take place exclusively *in vitro* and those that utilise a mixture of *in vitro*/*in vivo* steps will be presented here, followed by an examination of the more restricted field of purely *in vivo* systems for directed evolution.

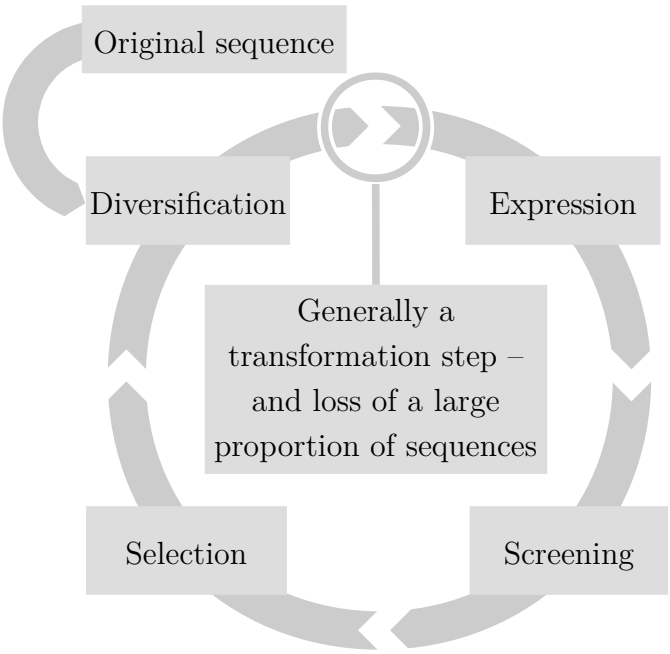


Figure 1: The directed evolution cycle

DNA is first diversified to produce a mutant library. The protein products of that library must then be expressed and each linked with the DNA sequence encoding it, such that screening based on the protein products allows the subsequent selection of the corresponding DNA. The selected sequences may then be further diversified, and the cycle repeated.

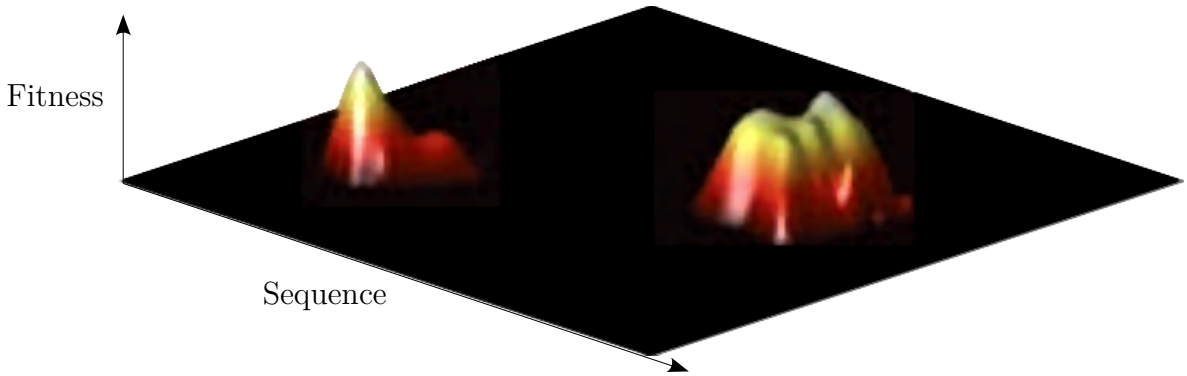


Figure 2: Directed evolution explores a 'sequence space'

A protein library from a pool of diversified DNA may be visualised as occupying a 'fitness landscape'. Fitness describes the ability of the protein to perform a desired function, with a higher peak (lighter colour in this figure) indicating a more fit protein variant. Most sequences will not function (black area), and functional sequences (coloured) will cluster with other functional sequences – i.e. fully-functional sequences may arise via less functional intermediates. Transitioning from one peak to another may result in a more fit protein but requires movement through a 'fitness valley'. Directed evolution methods which only search for iterative improvements to an original sequence (seeding a second cycle with successful variants from the first, for example) may never identify a discrete peak which may produce a more fit protein.

Figure & legend adapted from Romero 2009.

1.1.1. Pure in vitro methods

There are a small number of methods for directed evolution which do not involve cells at all – namely, mRNA (Xu 2002) and ribosome (Hanes 1997) display. In these methods a set of nucleic acids, diversified so that the resulting proteins expressed from them display a wide range of sequence variation, are scanned. They are then screened to select for binding affinity to a desired target ligand. The central problem with a purely *in vitro* system is the linking of phenotype – which is selected for – and the corresponding genotype. In *in vivo* systems the DNA encoding a successful protein is contained within the cell with it, establishing a link between a successful mutant and its sequence. The *in vitro* display methods rely on fusion of the protein/mRNA complex by puromycin-induced translation abortion (in the case of mRNA display), or the isolation of the protein/ribosome/mRNA complex (ribosome display), to create a physical link between protein variants and their encoding nucleic acid sequence. This allows the linking of successful phenotypes to their associated genotypes, and retains the genetic code responsible for producing the protein for purification.

These 'display' methods allow optimisation of single proteins with high efficiency, given that selection may be performed *in vitro*. This allows the screening of a significantly larger number of sequences than those methods which incorporate an *in vivo* selection step, as screening and selection are applied directly at the molecular level without the

need to culture and examine cells. Display methods are particularly attractive for the evolution of proteins which are deleterious or deadly to cells (Spirin 2004), but may not be used for pathway evolution, for example, as the linking of phenotype and genotype cannot work beyond a single gene. Additionally, proteins which rely on an intra-cellular environment to fold or function properly are not amenable to optimisation via *in vitro* methods. These methods sacrifice flexibility for great gains in efficiency within their narrow scope.

1.1.2. Mixed in vitro/in vivo methods

An abundance of approaches exist that utilise a mixture of both *in vivo* and *in vitro* steps (Lutz 2004), and these mixed methods are certainly the most common approach to directed evolution and protein engineering. Typically, the diversification of DNA occurs *in vitro*, and the subsequent steps of screening and selection take place within cells. The expression of proteins in the cellular context means they may be selected based on how they act in the native environment; this eliminates the possibility of a promising candidate selected *in vitro* not translating to a successful *in vivo* expression. Common approaches to the diversification of nucleic acid sequences in mixed directed evolution methods include error-prone PCR, recombination *in vitro* and oligonucleotide-directed randomisation. The resulting pools of diversified DNA are then transformed into cells, for screening and selection.

A notable strategy that diverges from these standard methods is phage display (reviewed in Fernandez-Gacio 2003). Here, the genotype and phenotype are linked via the integration of the diversified library into a gene encoding a coat protein of a phage. Upon assembly of the coat, a variant of the library of proteins is displayed upon the coat. A desired target for interaction is then immobilised on a matrix. Phage which display a protein variant which can bind to the target will remain after washing the rest of the pool away, and may subsequently be cleaved and eluted. Enzymes may be selected for by the use of substrates which break down into strong inhibitors, thus trapping any enzymes which successfully cleave the target; in both cases, the transformation of successful phage particles into *Escherichia coli* allows the propagation of relevant genes for sequencing. The library may be panned in cycles to determine the strongest interactors, and those subjected to more rounds of diversification – the efficiency of screening is limited by the phage titre achievable, as well as the transformation efficiency.

These mixed methods are susceptible to a loss of efficiency during the transformation step, as mentioned earlier in this text. The transfer of the sequence pool generated by, for example, error-prone PCR into the host cells for screening and selection is a relatively inefficient process. While billions of diverse sequences may be produced in the PCR reaction, transformation and screening will limit the number of sequences for

selection to at most several hundred thousand sequences – even in a scenario involving high-throughput screening by fluorescence-activated cell sorting (FACS), for example. An additional layer of inefficiency is introduced by the time-consuming nature of the protocol – transferring back and forth between cells and *in vitro* environments is time- and often labour-intensive.

1.1.3. Pure in vivo methods

The facilitation of directed evolution within living cells is most simply accomplished by the incorporation of the target DNA into a strain possessing a mutagenic phenotype. One example is *E. coli* XL-1 Red (Greener 1994, 1996), which derives its mutagenic phenotype from the mutations in the mismatch repair pathway, the oxo-dGTP repair pathway, and the 3'– 5' exonuclease subunit of DNA polymerase III. The strain therefore does not strictly increase the incidence of mutation events, but reduces the number of those which are repaired to wild-type. Transfer of a target plasmid into such a strain via transformation causes accumulation of mutations over many generations with a low frequency: It is reported by Greener et al that the rate is approximately 1 mutation every 30 generations for a 2,000 bp plasmid within the XL-1 Red host.

While this can result in optimisation of a target protein (for example, Bornscheuer 1998), the small sequence space explored restricts the strain to minor optimisations

rather than more major adaptation. More mutagenic strains encounter a trade-off between accelerated evolution of a target, and faster decline in health of the cells in the population as mutations accumulate within their genomes. One extension of this idea utilises repression of the repair genes via induction of antisense RNAs, based on a plasmid, which silence the genes in question (Nakashima 2009). This allows mutagenesis to occur in response to an inducer, rather than continuously, and was shown to increase mutation frequency by 2,000 fold over background. The lack of a targeting method, meaning the mutagenic effect does not discriminate between the intended target and the cell's own DNA, will still result in a decline in fitness of the host over time. This necessitates re-transformation of the plasmids into a new set of cells, bringing along with it the very inefficiency *in vivo* methods seek to avoid.

One promising method which is reminiscent of the *in vitro* display methods mentioned above, is phage-assisted continuous evolution, or PACE (Esvelt 2009). This method utilises a global mutator strain of *E. coli* in continuous culture to perform the diversification step of the directed evolution cycle – screening and selection of members of the target library, however, are carried out by M13 bacteriophage. Target genes are initially diversified, and the members of this library are packaged onto the genomes of a phage population deficient in protein III. Phage lacking pIII suffer a 10^8-fold reduction in infectivity; protein III is introduced by transformation into the mutator strain on an

accessory plasmid. The production of protein III from the plasmid is then linked to the degree of function of the target gene. Upon infection of the mutator cells, phage harbouring successful variants will therefore replicate more efficiently; as the phage continuously infects mutator cells, reproduces, repackages (possibly after incorporating further variation) and exits the host, library members will proliferate more successfully in line with their functionality. This method may theoretically be used to diversify and screen any sequence which may be linked to the transcription of protein III, or the replication of M13 phage in general, and its main advantage is its continuous nature. This allows exploration of a sequence space with no upper bound, beyond the time limit set on the culture procedure – in the test case, T7 RNA polymerase was diversified and 200 rounds of protein evolution were performed over eight days, resulting in a number of interesting variants of the target.

An *in vivo* method which involves direct targeting of mutation is described by Camps et al (2003). It utilises a variant DNA polymerase I (PolI), modified to be more prone to error than its native counterpart, to introduce mutation during the replication step of DNA. In *E. coli* the first ~500 nt of DNA synthesised during replication are done so by PolI, with the remainder completed by DNA polymerase III: The authors took advantage of this 'specificity' by co-transforming a cell with one plasmid containing

their variant PolI and another – which relies on PolI for replication – containing a target sequence immediately downstream of the origin of replication.

The authors expected to see increased mutation in the stretch of DNA replicated by PolI; that is, the first ~500 bp following the site of replication initiation. In fact, they saw greatly increased mutation evenly along the first ~700 bp, and an increase in mutation over background up to 3.7 kb into the plasmid sequence. They report an average of 8.1×10^{-4} mutations per base pair for that stretch of sequence, or an increase of 80,000 times when compared to the native PolI. The method is therefore powerful for achieving targeted mutation of sequences based on plasmids.

While altering the mutagenic activity of the chassis (herein defined as the host environment, usually a prokaryotic cell but not necessarily) is one approach, others involve more active engagement in the process of diversification. Perhaps the most prominent example of such a method is multiplex automated genome engineering – or 'MAGE' (Wang 2009). MAGE uses a specific chassis strain (*E. coli* EcNR2) which expresses a single-stranded DNA (ssDNA) binding protein from bacteriophage. This allows incorporation of oligonucleotides introduced into the cell into homologous areas of the genome while it undergoes replication; diversification of a pool of oligonucleotides bearing homology to the target sequence beforehand means that variation may be introduced into the genome in this manner. By carrying out several cycles of

introducing pools of diversified oligonucleotides and allowing DNA replication and cell division, Wang et al caused one third of cells in the population to take up some variant of the target sequence.

By adjusting one or both of the level of diversity represented in the oligonucleotide pool and the cycle count, the scope and efficiency of the generation of diversity within the target or targets may be tuned. As a test case, Wang et al demonstrated a 5-fold increase in the production of a target molecule – lycopene – after 3 days of continuous cycling. In this case, 24 genes in the lycopene production pathway were targeted simultaneously for diversification, toward the overall optimisation of the whole pathway itself. Each target gene was found to be different from wild type in the successful strains, with four genes targeted for knockout showing reduced or absent activity. This multi-gene targeting approach allows pathways to be optimised, with a large sequence space able to be explored; up to 15×10^9 variants are able to be generated over 35 cycles.

Outside the realm of prokaryotic chassis and systems, a number of eukaryotic methods for *in vivo* directed evolution are being developed – or have already been used to some degree of success. Of particular relevance are the development of mutagenic strains based on the cytidine deamination activity of AID. By over-expressing AID and simultaneously suppressing repair pathways, DNA near AID target sites undergoes

increased mutation. By integrating targets near these sites, mutation can be promoted with a degree of targeting. Somatic hypermutation caused by such expression of AID in a chicken cell-line modified to impede repair has been used to diversify a GFP gene (Arakawa 2008). The GFP was integrated proximal to an AID target site, with the analysis of the resulting mutations being used to engineer GFP to produce variants with increased brightness – up to threefold.

1.2. Developing a system for *in vivo* directed evolution

An extremely powerful variant of directed evolution, therefore, is that subset of techniques that takes place entirely *in vivo*. A robust *in vivo* method should allow efficiency to be maintained between stages or cycles of a directed evolution protocol, allowing the use of long diversification times. Coupling this with selection in real-time, i.e. in parallel with diversification, would allow the exploration of a much broader sequence space, potentially meaning more non-local peaks could be discovered. A common drawback of pure *in vivo* approaches to directed evolution which currently exist is the inherent 'background' damage that the host genome undergoes when cells are exposed for extended periods of time to whichever mutagenic approach is used in the technique. This research project examines the feasibility of using a protein-based device to effect mutation *in vivo* in a targeted manner – i.e., the restricting of mutagenesis to the target sequence and minimising its contact with, and impact on, the

host genome. Before introducing the design of the system, the individual components will be introduced to provide contextual information as to why they were chosen.

1.2.1. Activation-induced deaminase (AID)

AID is a protein which plays a role in the immune system of higher organisms – specifically, in somatic hypermutation or the diversification of antibodies. It thus already operates as a mutator in nature and is a good candidate for the mutagenic part of the fusion protein.

AID was discovered in 1999 by Muramatsu et al, and was thought at first to target RNA. Later studies in *E. coli* determined that it actually acted on DNA (Petersen-Mahrt 2002) via deamination of cytosine residues to uracil (Sohail 2003) – these errors may then be corrected by the base-excision repair (BER) pathway. In the case of AID mutation, uracil DNA glycosylase (encoded by *ung*) produces an abasic site by excising the uracil from the backbone. An AP lyase may then cleave at the abasic site, allowing it to be repaired. If not excised by uracil DNA glycosylase, the mismatch will result in the incorporation of a deoxythymidine at the site during replication. As a result of this, AID commonly induces point mutations of C to T in DNA, but may cause more varied mutation rates at reduced frequencies.

AID can only bind to – and act on – ssDNA (Dickerson 2003; Pham 2003). Double-stranded DNA (dsDNA) undergoing transcription exposes each strand, however, so AID may act on dsDNA that is actively being transcribed. Sohail et al (2003) and Ramiro et al (2003) found that AID acts with greater efficiency on targets which are highly transcribed. These studies demonstrated that a kanamycin resistance gene engineered to knock out function reverted to function 20- (Ramiro 2003) to 50-fold (Sohail 2003) more often when it was actively transcribed in a cell expressing AID – and that if expressed *in vitro* under the control of a T7 promoter, that frequency reached 300-fold (Sohail 2003).

AID has been identified as exhibiting preferential binding to transcription bubbles and similar structures (e.g., hairpin loops) as opposed to linear ssDNA (Larijani 2007). After binding to DNA it acts processively, and may thus cause several mutation events before disengaging (Brar 2008; Pham 2003). A 15-fold preference for the non-transcribed strand was found to be exhibited by AID in a 2003 *in vitro* study by Pham et al., who showed that the transcribed strand is 'protected' to a degree from action by AID due to its hybridisation with RNA. The same study also uncovered a preference for action at particular DNA motifs; that is, that AID shows some degree of sequence-dependency. This is consistent with its activity in somatic hypermutation, where it is

more likely to act at hotspots, although this is maintained to be an artefact of reaction conditions by other investigators (Sohail 2003).

AID is a close relative of proteins in the APOBEC family (Bransteitter 2009) of cytidine deaminases. A number of APOBEC proteins are active as dimers or other higher-order oligomers (Prochnow 2007), and tetrameric AID has been purified and shown to cause deamination *in vitro* consistent with its activity (Dickerson 2003). Mutations posited to interfere with oligomerisation, as informed by comparison to solved structures of APOBEC-2, a close relation, caused a decrease in the ability of AID to cause deamination (Prochnow 2007).

Whether oligomerisation is a requirement for the maintenance of functionality (in full or in part) has as of yet not been elucidated, and is a contentious area within the field (see Sacho 2008). A study by Brar et al (2008) used atomic force microscopy to demonstrate that AID has catalytic activity as a monomer, and does not aggregate within transcription bubbles – other researchers contest these data, however, and the full truth is yet to be determined.

1.2.2. T7 RNA polymerase (T7 RNAP)

First isolated almost half a century ago from Type 7 (T7) bacteriophage (Chamberlin 1970), T7 RNAP has since become a workhorse in the fields of molecular biology and

biotechnology. It exhibits high specificity and affinity for its promoter (P_{T7}), only initiating transcription where this promoter is found and not from other sequences – likewise, P_{T7} does not bind native RNA polymerases so can only allow transcription in the presence of T7 RNAP. Transcription is highly efficient, progressing at a rate 5 times that of *E. coli*'s native RNA polymerase (Chamberlin 1973; Golomb 1974).

This increased activity over native polymerases can result in a large number of RNA transcripts, so that even a small number of active T7 RNAPs may cause saturation of the cell's translation machinery – this was demonstrated in cells transformed with a T7-driven protein expression cassette on a multicopy plasmid by Studier (1986). Termination efficiency at sequences which effectively terminate transcription by other polymerases is often markedly lower for T7 RNAP (Jeng 1990), and variants of T7 RNAP have been characterised that read through an even larger set of terminator sequences (Lyakhov 1997). T7 RNAP will transcribe continuously in a stable manner if not terminated, and may run full-circle around an expression plasmid many times over (Studier 1986).

The polymerase itself consists of only a single subunit, and does not require additional cofactors to initiate and complete successful transcription (Sousa 1993). The combination of high efficiency and structural simplicity, along with the relative

orthogonality granted by the promoter/polymerase specificity pairing, make it an excellent candidate for the abundant expression of a target gene.

1.2.3. Coupling DNA deamination to specific transcriptional activity

The central aim that is presented in this thesis is to provide spatial and temporal coupling of AID activity to a transcription bubble in a specific manner. This has been approached by creating a fusion protein of AID and T7 RNA polymerase using a modular, extendable linker. Figure 3 shows a schematic representation of the system.

Expression of the mutator protein in a cell with a target gene downstream of a T7 promoter should cause a high rate of transcription of that target by the fusion protein, during which single-stranded DNA in the transcription bubble will be exposed to the mutagenic AID moiety. The localisation of the AID molecule, along with the high transcription rate of the target, should result in a greatly elevated mutation rate above the background rate of the cell. As an added benefit, the constant strong expression of the target ensures that the cell's phenotype is kept 'up-to-date' with its changing genotype, which is key for effective selection strategies – successful candidates may be selected more accurately and those which mutate 'past' a desired peak will not be included (or may be selected prior to this occurring, if selection is in real-time).

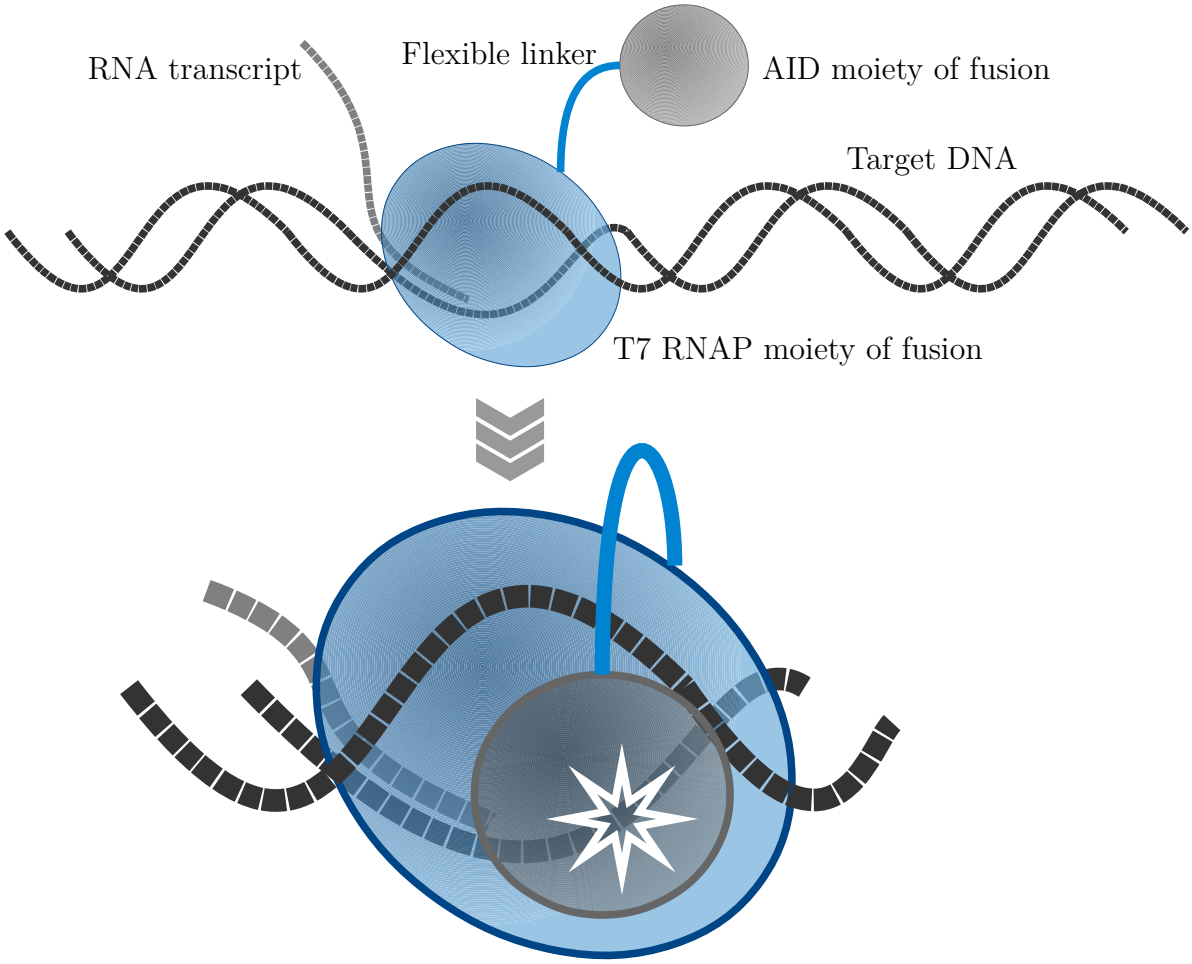


Figure 3: Schematic representation of the system

Top: The T7 RNA polymerase is fused to the AID moiety by a flexible polypeptide linker. The polymerase will transcribe target DNA downstream of a T7 promoter.

Bottom: AID may interact with the ssDNA in the transcription bubble, causing deamination of the target DNA.

1.2.4. Targeted deamination in context

The T7-AID fusion should demonstrate several advantages over the *in vivo* methods described in the literature.

Error-prone Polymerase I can only cause mutations within a short stretch of DNA; furthermore, mutation frequency along that length is biased/skewed toward the start and tails off as the distance from the initiation site increases (Camps 2003). This means that this method may not be used to efficiently and reliably mutate larger genes, or operons, with ~3.7 kb being the apparent upper limit before the mutation frequency is not significantly higher than background. In contrast, the T7-AID fusion system should be able to cause mutations along great lengths of DNA. T7 RNA polymerase has been shown to transcribe an entire expression plasmid several times over if left unterminated (Studier 1986). The addition of further T7 promoters evenly spaced along a sequence, or a reverse promoter at the end of the target sequence, would be a trivial way to increase consistency of mutation along an arbitrary length of DNA. Indeed, bacterial terminators may be engineered to allow T7 RNA polymerase to pass through unhindered (Jeng 1990) while still retaining good termination efficiency, meaning that stretches of DNA including transcriptional terminators would not necessarily hinder the fusion.

Another significant disadvantage of the modified PolI approach is that it may only be used in a plasmid-based system. While a plasmid-based system presents more copies of a target for diversification, it also serves to obfuscate the linkage of genotype and phenotype that is the advantage of *in vivo* screening and selection. Mutations leading to a desirable phenotype will originate in only a subpopulation of the plasmids in the cell; this may be masked by other plasmids, or may be 'outcompeted' and the successful candidate lost. While this is not an issue for strict gain of function mutations, it prevents the approach from being used for diversification of function. In contrast, the T7-AID fusion system is amenable to the evolution of targets both in plasmids and integrated in single-copy within the genome of a chassis. This allows for much more flexibility in the type of mutations that may be screened for.

MAGE may be used to target genomic DNA – indeed, it requires integration of targets – and may be used to diversify several targets at once. Those targets need not be in series, and the targeting is such that there is very little risk of diversification in non-target sequence. The primary disadvantage of MAGE is in the initial construction of libraries of variants of each target; these must be large, and are generated *in vitro*. An analysis by the authors of the cost estimated that the price of oligonucleotides required to allow fully-degenerate diversification of 27 base pairs at one target site was only $30 at the time of writing (Wang 2009), but the time factor is not included in that analysis.

Longer sequences or pathways, while generally not needing to allow full degeneracy, increase costs further.

Both PolI and MAGE are not applicable in a eukaryotic context. *In vivo* directed evolution of genes in eukaryotes which may rely on folding or post-translational modification impossible in prokaryotic chassis can not be approached by those techniques. In contrast, the T7-AID fusion may be amenable to use within a eukaryotic context. As described earlier, methods exist which utilise the activity of AID in eukaryotic cells to achieve directed evolution, and T7 RNA polymerase – along with associated promoters and terminators – retains function and specificity within eukaryotic cells (Lieber 1989).

The fusion system offers an approach that may be used to optimise a target on a plasmid or on the genome, and within both pro- and eukaryotic hosts. Additionally it does not involve the strenuous set up associated with a MAGE run or the length-dependence of error-prone PolI. One disadvantage is its narrow mutation range (C to T or G to A), but since AID-based methods have been used with success this may be considered a secondary issue to be approached and optimised when the system is developed.

1.2.5. Aims of the project

Assessing the viability of the proposed system is the primary objective of this research, and this effort may be broken up into several distinct stages:

- Define a strategy for measuring mutation and use this as the basis of experimental design.
- Show that the fusion protein retains the ability to a) effect transcription from T7 promoters and b) cause mutation in a targeted manner, without imposing such an untenable burden on the chassis – either mutagenic or metabolic – as to destroy it
- Determine the effectiveness of the fusion at targeted mutation – to ensure it incorporates a) a high enough rate of mutation that a useful sequence space may be explored without being high enough to consistently degenerate the target sequence faster than selection may be applied and b) an improvement in targeting over expression of AID and T7 in an unfused conformation
- Demonstrate a practical application of the system – to mutate a chosen target in a useful way, by diversification of a signalling molecule.

The genetic circuits used in the project are to be constructed using Synthetic Biology techniques, with parts and devices kept compatible with Synthetic Biology standards.

2. Constructing a system for *in vivo* directed evolution

2.1.1. Aims

Construction and testing of the device with an eye toward quantitatively estimating the mutation rate of the system may be subject to experimental considerations or constraints. Prior to beginning any experimental attempts to characterise the mutational efficiency of the device, a strategy must be defined for that characterisation. From there, the experimental design constraints which must be adhered to can be specified.

The next fundamental aim is the construction of the device and any variants thereof, which may be required to determine and rule out obfuscating factors contributing to an observed mutation rate, such as background. Along with these device plasmids, target sequences must be selected and placed behind P_{T7} in an appropriate context. The devices and targets required for construction will depend on the experimental strategy informed by the method to be used for measuring mutation rate.

These constructs must be tested to ensure basic functionality – in the case of the fusion protein, to ensure that both moieties retain function when fused.

2.2. Experimental strategy

Assessing mutation rate may be done in a number of ways which fall broadly into two classes: quantitative and qualitative/comparative methods. Qualitative methods examine an increase above background, for instance, or compare two variants and rank them without determining the defined mutation rate. Calculating the rate itself quantitatively requires methods which are necessarily more complex and must adhere to strict parameters.

2.2.1. Rifampicin reversion

One widely-used method for comparative mutation rate experiments is rifampicin reversion. Rifampicin is an antibiotic which interferes with native bacterial RNA polymerases; it binds to the beta subunit of the RNAP and blocks transcription initiation, preventing transcription of proteins and leading to the death of the cell (Wehrli 1971). There are several possible point mutations that can occur in the sequence encoding the RNAP which can result in amino acid changes but will not alter the function of the polymerase (Jin 1988). As such, they confer no reduction in fitness, but they do prevent the binding of rifampicin and render the cell resistant to its effects. These mutations may occur spontaneously (thus rifampicin is used therapeutically only in concert with other antibiotics (Tupin 2010)) but do not represent the wild type.

Culturing cells in the presence or absence of a putative mutator, and then exposing the cultures to rifampicin (by plating on antibiotic-supplemented media for example) and determining resistance reversion rates, allows mutagenic activity to be benchmarked. Increased mutagenic activity will result in increased mutations at one or more of the points on the RNAP gene, resulting in increased frequency of reversion to resistance. Importantly, excessive mutation of the RNAP gene will result in cell death, eliminating the problem of masking of successful reversion mutations by deleterious knockout ones. This will be examined in more detail in section 2.2.3.

Rifampicin reversion will therefore be used to assess whether AID retains activity in the fusion. Comparing the reversion rate of a strain transformed with the T7-AID fusion protein to that of a negative and positive control (i.e. no AID and free AID respectively), will elucidate to what extent – if any – AID activity is inhibited by its fusion to T7 RNAP. It is worth noting here that rifampicin does not interact with the T7 RNA polymerase, as it lacks the structural motifs required (Kuderova 1999).

2.2.2. Controls required for assessing targeted mutation

For the mutator protein, however, the general increase in the mutagenic phenotype of the cell is only part of the story. The rate of mutation of the target should exceed the background increase, and it is this comparison that is important in determining the

efficiency of the system as a whole. The conventional approach to the assessment of the level of mutagenesis displayed by targeted directed evolution systems relies simply on the sequencing of a large number of target DNA molecules post-mutation, and comparison of the rate of mutation in those sequences to the native background. This approach must be adapted for the assessment of the fusion protein, however, as it may not be a strictly targeted system.

Unlike, for example, the modified PolI method developed by Camps et al (2003) in which mutations must fall within a specific range after the origin of replication on a plasmid; or MAGE where all mutation is confined to those areas targeted by diversified oligonucleotides, the T7-AID fusion's mutagenic activity may not be strictly constrained to areas under the control of P_{T7}. Since each mutator molecule will not be bound to and transcribing DNA at all times, they will cause mutation at non-target sites – the fusion effectively increases the background mutation rate of the cell. An estimate of the rate of mutation which is only based on examining the sequence of the target DNA 'before and after' exposure to the fusion protein will therefore not be able to be attributed to a targeted mutation effect without first ruling out an enhanced background mutation rate as being the cause.

Should AID be confirmed to have retained function in the fusion, the next step toward assessing its viability as a directed evolution platform will be to culture it with a target

under P_{T7}. In this case a comparison will still need to be made against free AID and background mutation of the target, with an important additional consideration – transcription-induced mutation.

Double-stranded DNA (dsDNA) undergoes spontaneous deamination at a rate up to 140 fold lower than single-stranded DNA (ssDNA), and the rate of such deamination of a stretch of DNA increases proportionally with the amount of time that DNA is in a single-stranded state (Lindahl 1974; Frederico 1990; Beletskii 2000). When DNA undergoes transcription, the separation of the strands exposes them to increased mutation, and a higher rate of transcription means the DNA is separated more often. Genes behind P_{T7}, as a result of T7 RNAP's high activity, show a rate of spontaneous deamination (C to U) which is up to 5 times higher than background levels (Beletskii 2000).

This effect must be controlled for – the negative and free AID controls must include a T7 RNAP gene to ensure that any increase in targeted mutation over background is not due simply to increased transcription by T7 RNAP. Thus the controls required – alongside the fusion protein – were determined to be T7 RNAP alone and T7 RNAP with AID unfused, both on otherwise-identical plasmids to the fusion protein.

2.2.3. Masking as a consideration in gain of function assays

In the ideal case a gain of function assay would be used when determining the rate of mutation at a target. This would typically be a reversion to resistance assay, as they allow screening by plating on the appropriate antibiotic; this automatically eliminates failures and leaves only successful revertants for scoring. Reversion to kanamycin resistance using a kan^S target is the standard gain of function assay, used commonly in the literature (Sohail 2003; Ramiro 2000). When a gain of function approach is applied to the analysis of a mutator with a mutation bias – i.e. a targeted system – however, one must account for a parameter that is otherwise safely ignored; masking of successful revertants by knockout. If a given DNA sequence contains one base pair that has been deliberately mutated to disrupt function, reversion of which will restore that function, it may also include one or more base pairs (or a combination of several) which will eliminate that function – even given successful reversion of the original mutation.

Figure 4 shows the effect of masking given one base that will restore function if mutated and one base which will destroy it in a putative target; additional deleterious mutation sites serve to shift the curve as shown in figure 5.

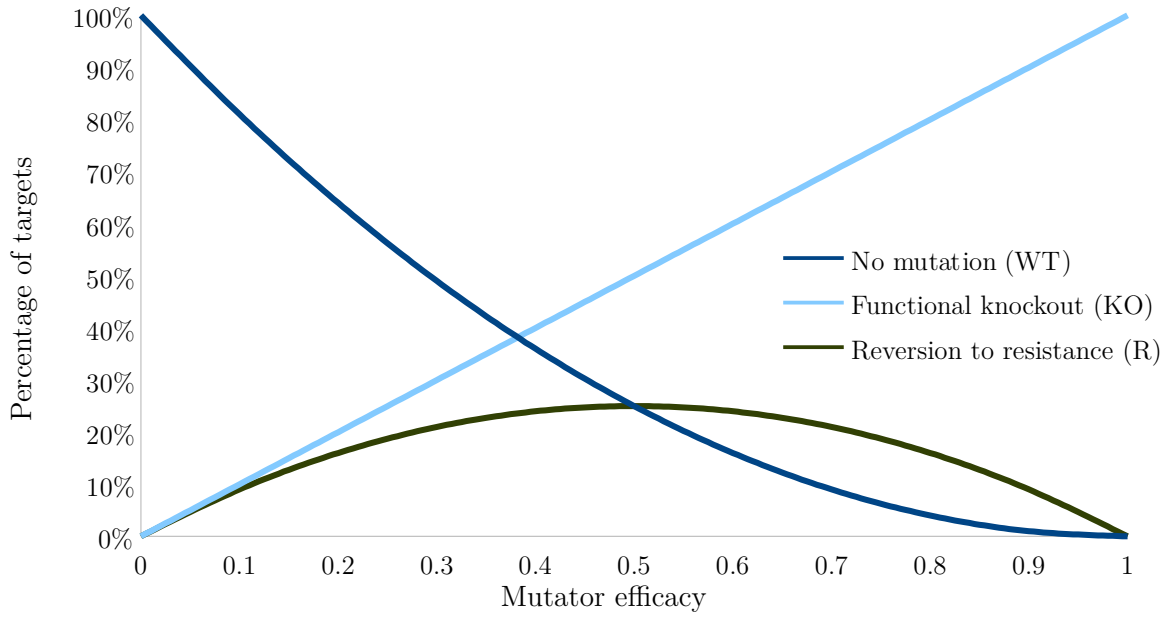


Figure 4: Masking effect of gene knockout, given one reversion and one knockout base

Demonstration of the masking effect introduced when a gain of function mutation may be 'overridden' by a separate knockout mutation. In this graph, a synthetic sequence is assumed to have one site each where a mutation event will restore function and knock it out, with knock out 'trumping' reversion.

Determination of the number of sequences avoiding mutation at either site WT is the product of the chance the mutator with efficacy e (ranging from 0, never mutates a target, to 1, mutates every target) will miss both the knockout site (number of knockout sites k) and the reversion site (number of reversion sites r) so

$$\mathrm{WT} = (1 - \mathrm{e})^{\mathrm{k}} * (1 - \mathrm{e})^{\mathrm{r}}$$

$$or \ = (1 - \mathrm{e})^{(\mathrm{k} + \mathrm{r})}$$

Determination of the number of knockouts K is by subtraction from the total, 1, of the cases where all k knockout points are missed so

$$\mathrm{K} = 1 - (1 - \mathrm{e})^{\mathrm{k}}$$

The number of sequences successfully reverting to function R can be determined by subtraction of the above cases from the total, 1, or may be derived from the chance for the mutator efficacy e to miss k knockout points and hit one of r reversion points so

$$\mathrm{R} = (1 - \mathrm{e})^{\mathrm{k}} * (1 - (1 - \mathrm{e})^{\mathrm{r}})$$

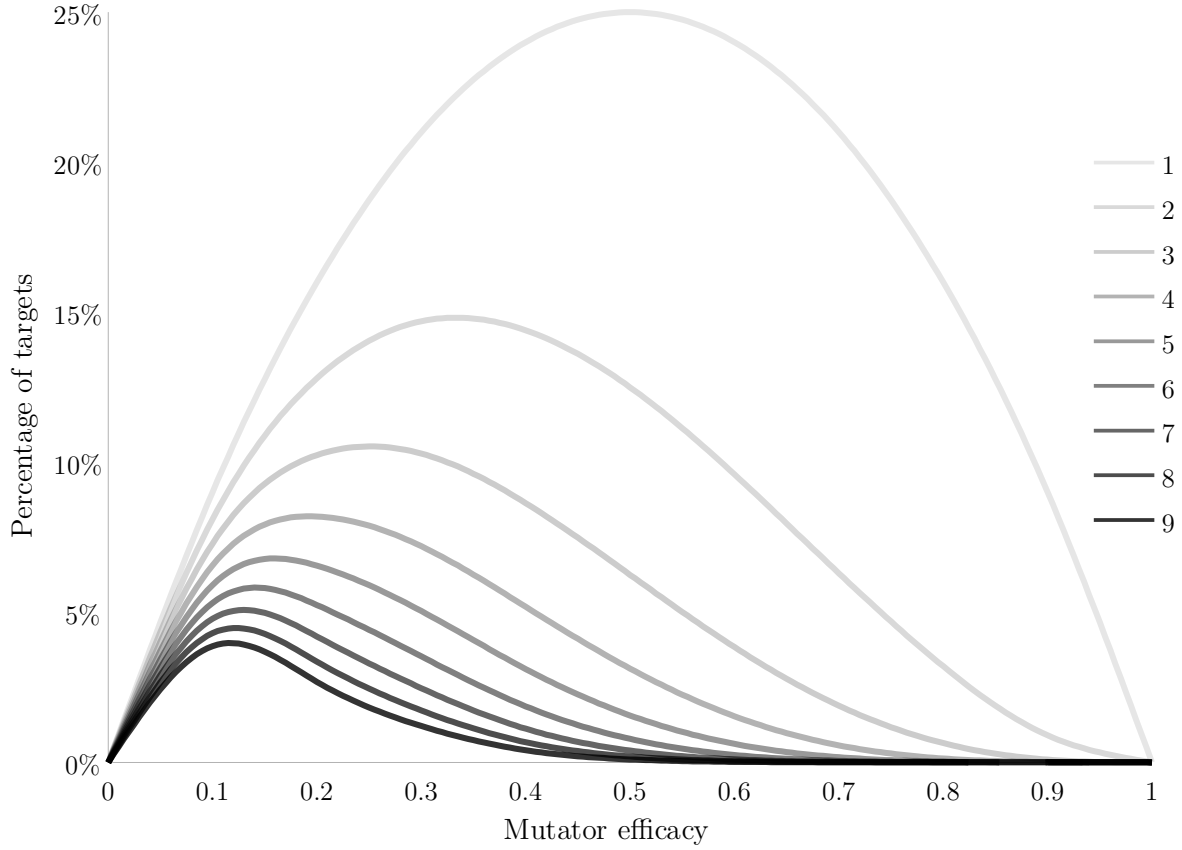


Figure 5: Effect of number of knockout bases on reversion frequency

Demonstration of the masking effect introduced when a gain of function mutation may be 'over-ridden' by a separate knockout mutation. In this graph, a synthetic sequence is assumed to have one site where a mutation event will restore function – plots for the frequency distribution for different efficacies of mutator are given for a range of possible numbers of knockout points (knock out 'trumps' reversion).

So we can see that while with one possible knockout point, masking only becomes a problem when mutator efficacy passes say 0.4, more knockout points causes masking to come into play earlier and with more drastic effect. In a realistic scenario, the number of knockout points in an assay is not only likely to be greater than the number of reversion points, but is likely to be an unknown parameter and therefore not able to be corrected for.

Masking may be safely left out of analysis only at very low levels of mutator efficacy; this is the case in almost all scenarios in nature, as such an effective mutator would be toxic to a cell, but targeted mutation rate may outstrip background to such a degree.

With imperfect knowledge of both the efficacy of mutation and targeting of the fusion, as well as of the number of bases in the target which could eliminate function if mutated, a reliable assessment of mutation rate cannot be derived purely from a reversion to function assay.

For non-targeted systems, this is not a consideration. The reasoning behind this assumption is that any global mutator with a high enough rate to fall prey to disruption by the masking effect would be so effective as to be lethal to the cell. When considering a targeted system, however, the rate of background mutation may be low enough to endure while the rate of mutation at the target is high enough to skew data.

It is therefore necessary to determine what proportion of putative revertants would also incur a knockout mutation. This may be done by performing a loss of function assay; if, post culture and screening, half the targets have lost function (for example), then in a gain of function test the true number would be double the number indicated by scoring revertants.

While gain of function assays are best performed using antibiotic-based targets and selection procedures, the same is not true for loss of function assays. Thus, while kan^S is generally used in this context, there is no corresponding loss of function assay to allow the analysis of masking. Instead, a pair of targets were designed – the GFP/XFP pair,

a functional and knockout version of a fluorescent protein which differ by one base (a T to C change from GFP to XFP – meaning AID can revert function by inflicting C to U deamination at that base). The data produced from these assays can then be analysed to determine the efficacy of targeted mutation by T7-AID.

2.2.4. Fluctuation analysis

For this project, fluctuation analysis was considered the most suitable method for determining mutation rate (Foster 2006; Rosche 2000). In brief, the method consists of calculating the number of mutation events, referred to as m, which most accurately describes the distribution of the number of mutants observed across parallel cultures. Counting the number of mutants alone will not give an accurate measurement of the number of mutation events in a bacterial host, as a single event will be repeatedly duplicated during the process of cell replication. The earlier the mutation event occurs, the greater the effect will be on the final number of mutants observed. Therefore, comparing the distribution pattern of different cultures is a more accurate method to determine m and remove the influence of this effect.

After the seminal work by Luria and Delbruck (1943) first described the distribution of mutants in a population given that new mutants arise both spontaneously and by division of pre-existing mutants, Lea and Coulson (1949) created the first model to

represent this distribution mathematically. Since then, a number of different methods have emerged for calculating m from the observed distribution, such as the P_0 method, Lea-Coulson's method of the median, or the maximal likelihood method. A thorough review of the advantages and disadvantages of each available method can be found in Rosche and Foster (2000).

The key point to remember when choosing which one to use for fluctuation analysis is that each of these methods is suitable only for a certain range of m. Therefore, it is necessary to have a rough estimate of the m value before performing more detailed analysis. This is a common mistake when fluctuation analysis is carried out; for example, in Sohail et al (2003) the researchers aimed to determine m of a free AID protein using a kanamycin reversion to resistance assay. They chose Lea and Coulson's method of the median to analyse their data, which is valid when m is in the range of 4 to 15 (corresponding to a median number of mutants between 10 and 60) (Rosche 2000). However, the lowest m value calculated in their research was 21 for a control sample while the highest value when AID was expressed was multiple orders of magnitude above the upper limit of 15. The authors estimated a 50-fold increase in mutation over background, but a comparable study by Ramiro et al (2003) produced a more conservative estimate of 20-fold, demonstrating the variability that may be introduced by the selection of the mutation rate estimation technique.

Finally, an additional consideration when using these analysis methods is that each one makes specific assumptions about the collection of data for analysis, which must be replicated as closely as possible in the experimental conditions. One of the common assumptions is that every mutant is observed. This makes reversion to resistance assays particularly suitable for fluctuation analysis, as every revertant can be counted.

When m cannot be estimated beforehand (or when it is very high, e.g. greater than 15), there are two methods that are considered most suitable to use. The first is Ma-Sandri-Sarkar (MSS) maximum likelihood (Sarkar 1992), a recursive method which generates distributions for a range of different m values and compares this to the distribution of the observed values until the m which produces the best fit is found. An advantage of maximal likelihood methods is that all the available experimental data are included, as opposed to methods such as the median. The second method which is suitable for high or unknown m values is a Bayesian approach (Asteris 1996). The first step is to express what is previously known about m as a predetermined probability density function (referred to as the prior pdf). Following Bayes' theorem, the experimental data (number of mutants in parallel cultures) can then be used to transform the prior pdf into a posterior pdf. This newly-generated pdf will therefore encompass information from both prior knowledge and the most recent experimental results.

In the original work describing the Bayesian approach, Asteris and Sarkar (1996) derived m from experimental data using both Bayesian methods and MSS maximal likelihood. They demonstrate that the Bayesian approach is the most efficient at estimating m compared to previously described methods, including MSS. It is even possible to incorporate both methods, by using MSS maximal likelihood first to generate a prior pdf before carrying out the transformation. The main benefit of the Bayesian approach is that it can be used to combine the results of multiple experiments, by using the results from the analysis of one experiment as the prior pdf for subsequent runs.

Either a Bayesian or MSS maximal likelihood approach was determined to be the most suitable for analysis of the data in this project, due to their advantage of being able to calculate m without the need for an accurate starting estimate of its value.

2.2.5. Culture conditions

Regardless of the method used to perform fluctuation analysis, it is important to keep in mind that the system being studied must comply with a number of assumptions as described in the original Lea and Coulson model (1949) in order to provide meaningful results. These constraints can reasonably be met using a well-designed protocol; the

exact experimental conditions chosen in this project to ensure the assumptions are fulfilled are described in table 1.

To ensure that the cell's repair machinery did not revert any mutation events arising from action by AID – in line with the assumption of fluctuation analysis – it was decided that all experiments would take place in a strain deficient in uracil DNA glycosylase, the enzyme responsible for the first step in the base excision repair pathway responsible for repairing C to T deamination. A strain with the gene (*ung*) encoding uracil DNA glycosylase knocked out was available in the laboratory, termed BH156 (Lutsenko 1999). Full genotype information is available in section 7.1.1, table 3.

In order to be able to determine m with reasonable precision – within 20% – at least 20 distinct cultures are required (Rosche 2000). It was decided that each run would consist of 24 parallel cultures to allow for possible failures, and each would be repeated a minimum of 3 times. A co-transformation approach was taken with the target and device on separate, compatible plasmids.

Requirement	Adjustment
Number of seed cells negligible compared to final cell count	Cultures to be grown to saturation, and seeded with ~150 CFU.
Successful mutation does not impair fitness	Target gene non-native and not vital
Negligible reversion of successful mutations	AID causes C to T mutations, cannot revert T to C. *ung*$^-$ background favours retention
Every reversion event is detected	Lethal selection is required, not actioned. Sampling error must be taken into account
Post selection, no new revertants appear	No selection pressure toward revertants
Each mutation event is independent	Not actionable

Table 1: Requirements for an accurate determination of m by fluctuation analysis, with actions taken to attempt to satisfy them in the design of experiments

2.3. Results

2.3.1. Construction of mutator protein

The mutator protein, comprising the AID-T7 fusion between a lac-inducible promoter/ribosome binding site (RBS) pair and a terminator flanked by BioBrick prefix and suffix sequences (Shetty 2008), was constructed prior to the start of this project. A brief outline of the original construction is given here.

Mammalian AID was codon-optimised, and the sequence sent for *de novo* gene synthesis (by Geneart AG) along with an upstream BioBrick prefix sequence, lac-inducible promoter P_{Lac} (part BBa_R0011 from the Registry of Standard Biological Parts, parts.igem.org) & RBS32 (Registry part BBa_B0032) and a trailing sequence comprising the first half of the linker sequence used to fuse AID to T7 RNAP. The T7 RNAP gene, minus its start codon, was amplified from the genome of *E. coli* BL21(DE3), using primers designed to incorporate a restriction site (EcoRI) and the latter half of the fusion linker sequence on the 5' end of the gene, and a double stop codon followed by a BioBrick suffix on the 3' end. The T7 RNAP sequence was PCR purified and cloned into a BioBrick vector, pSB1AK3 (henceforth abbreviated pA3) containing the transcriptional terminator BBa_B0015 (Registry) via insertion before that part. Following ligation and purification, the P_{Lac}-AID and T7 RNAP-terminator

sections could thus be cloned together via restriction cloning at the shared SphI site in the linker, creating the fusion (pA3:TA). The linker region used here for cloning contains AgeI, SphI and NgoMIV sites. These allow modular extension of the linker by restriction and ligation, in the event the fusion is discovered to lack AID activity; adjusting the length of the linker may restore it in that case.

2.3.2. Construction of controls

The control devices, T7 RNAP alone and T7 RNAP with AID expressed separately, were constructed by PCR. Primers T7F/R were designed to linearise the fusion plasmid minus the AID moiety, with the missing start codon being added to T7 RNAP via addition to the forward primer. The two primers also included 15 bp homology with each other on the 5' ends, allowing the InFusion assembly method to be used to re-circularise and ligate the plasmid (pA3:T).

This technique was repeated to create an AID-alone variant of the plasmid using analogous primers AIDF/R (adding a stop codon), and the two genes were then cloned together using BioBrick methods to create the unfused control (pA3:T+A).

The control plasmids were sized following colony PCR with primers VF2 and VR, and the correct sequence was confirmed by DNA sequencing.

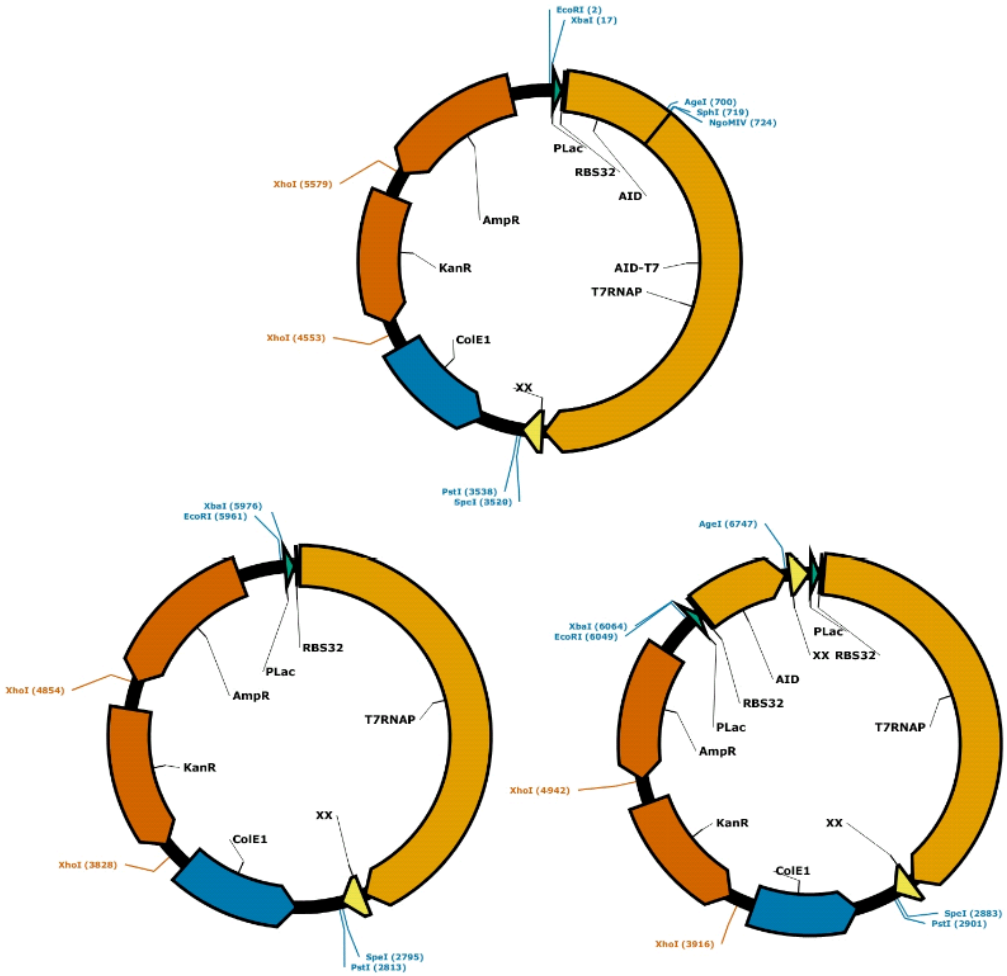


Figure 6: The device plasmids; the fusion protein and two controls, on a high-copy backbone

Top: pA3:TA, the T7-AID fusion protein in pSB1AK3. The T7-AID insert is behind a lac-inducible promoter (P_{Lac}) with a strong RBS (RBS32) and a pair of transcriptional terminators (XX) directly flanking it. The insert is appended with the BioBrick prefix and suffix, containing EcoRI/XbaI and SpeI/PstI restriction sites respectively. The backbone possesses kanamycin and ampicillin resistance (KanR and AmpR respectively) and utilises a high-copy ColE1 origin of replication.

Bottom, left: pA3:T, T7 RNAP without AID in pSB1AK3. Bottom, right: pA3:T+A, T7RNAP & AID unfused, each behind their own P_{Lac}.

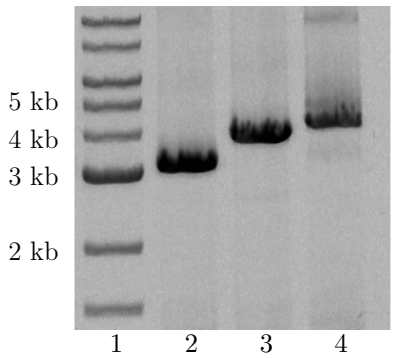


Figure 7: Agarose gel showing the products of a colony PCR confirming successful construction of device plasmids

Lane 1, 1 kb DNA ladder; Lanes 2-4, T7RNAP, T7-AID and T7&AID PCRs with VF2 and VR showing expected sizes.

2.3.3. Construction of GFP and XFP targets

GFPmut3b, a mutant of the green fluorescence protein (GFP) from *Aequorea victoria* engineered to allow faster folding and increased quantum yield, as well as a tighter excitation/emission spectrum (501/511 nm respectively), was used as the target sequence (Cormack 1996, Andersen 1998). GFPmut3b behind P_{T7} was cloned from the Registry plasmid pSB1A2 into pACYC-Duet (abbreviated pCD) and pSB4C5 (pC5), using Biobrick methods (section 7.2.8). Briefly, genes were excised from their backbones by either EcoRI/SpeI or XbaI/PstI double digest and inserted into vectors cut EcoRI/XbaI or SpeI/PstI respectively. This concatenated the sequences, re-forming the EcoRI/PstI site and forming a scar between the parts where XbaI/SpeI were ligated. This produced plasmids pCD:P_{T7}-GFP and pC5:P_{T7}-GFP.

Site-directed mutagenesis was performed on the resultant plasmids, introducing a T to C point mutation at base pair 196 (amino acid 66) of the mut3b gene; the resultant amino acid change from Tyrosine to Histidine (Y66H) shifts emission from green toward the blue end of the spectrum in wild type GFP (Heim 1994) but knocks out fluorescence entirely in mut3b. This non-fluorescent variant was termed XFP, and both GFP and XFP were tested for expected fluorescence profiles by transformation into BL21(DE3) after confirmation of successful construction by sequencing.

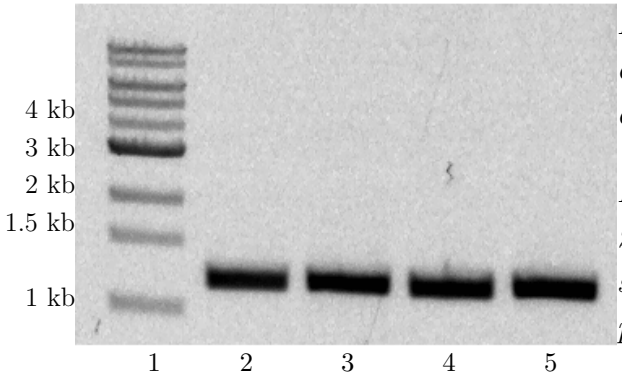


Figure 8: Agarose gel showing the products of a colony PCR confirming successful construction of GFP target plasmids

Lane 1, 1 kb DNA ladder (NEB); Lanes 2-5, colony PCRs with VF2 and VR showing expected size of GFPmut3b in pSB4C5.

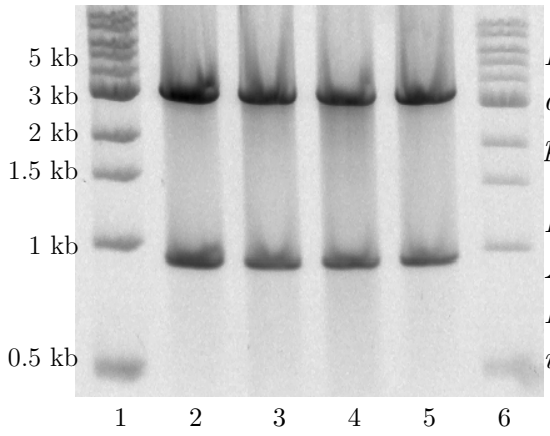


Figure 9: Agarose gel showing the products of an analytical digest confirming successful production of XFP by SDM

Lanes 1 & 6, 1 kb DNA ladder; Lanes 2-5, XFP in pSB4C5 double-digested with EcoRI & PstI showing expected sizes of both vector and insert.

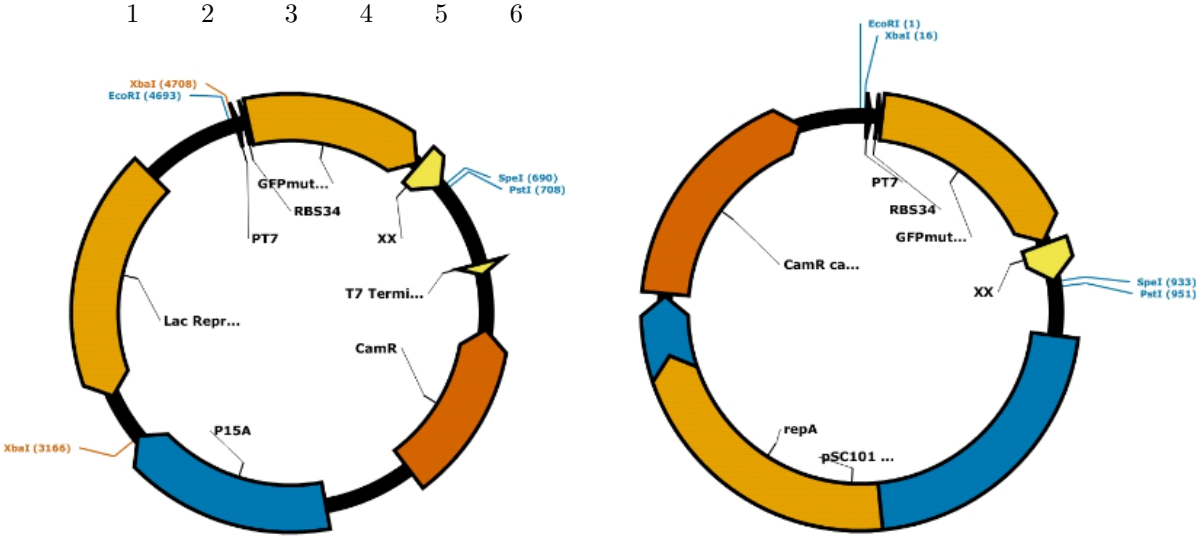


Figure 10: The target plasmids

Left: pCD:P_{T7}-GFP, the P_{T7}-GFP target in pACYC-Duet. The backbone possesses chloramphenicol resistance (CamR) and utilises a medium-copy P15A origin of replication. Right: pC5:P_{T7}-GFP, the P_{T7}-GFP target in low-copy pSB4C5 which also contains chloramphenicol resistance but utilises a low-copy pSC101 origin of replication. XFP targets are identical, with XFP replacing the GFP target producing pCD:P_{T7}-XFP and pC5:P_{T7}-XFP.

2.3.4. Confirmation of activity of devices

Activity of both the T7 RNA polymerase and the AID module within the chassis needed to be confirmed before any attempt to utilise them for more complex experiments. To this end, pCD:P_{T7}-GFP was co-transformed into a non-DE3 strain (GM31) along with pA3:T, pA3:TA and pA3:T+A device plasmids. These were then plated and all exhibited glowing colonies, confirming T7 RNAP activity was maintained post cloning and fusion. AID activity was assessed by using the rifampicin reversion assay described earlier and detailed in section 7.6.1. Twenty platings were performed over two separate runs of ten independent cultures each, and the results tabulated and plotted as shown in figure 11.

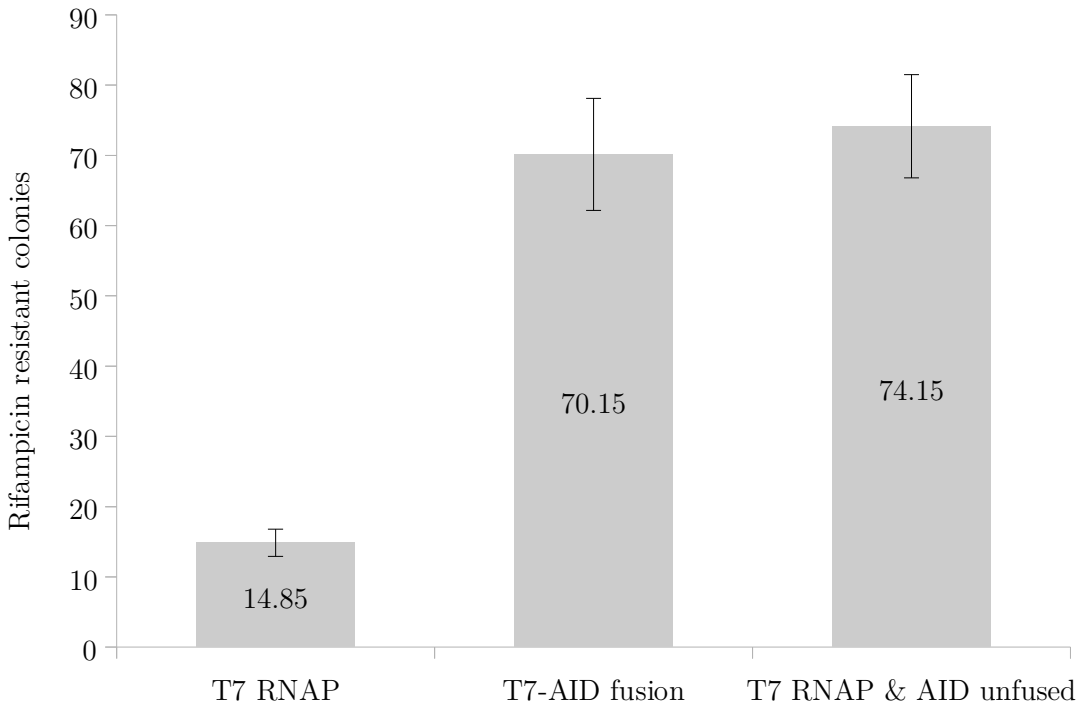


Figure 11: Rifampicin reversion assay

Device plasmids were tested for a mutagenic phenotype. 10 independent cultures were assayed with 2 repeats. Standard error across all samples is shown.

2.4. Discussion

2.4.1. Summary of strategy

Table 2 presents a summary of the experimental strategy that was devised based on the considerations discussed above. This strategy informed the design and construction of the devices, the choice of targets, and the experimental work-flow.

Question	Experiment
Does AID retain function in the fusion?	Rifampicin reversion assay vs. background
Is that function inhibited in any way?	Comparison to unfused AID
Does the fusion cause mutation of a target gene?	Co-transformation of the fusion with a target gene under T7
Can this be accounted for by transcription-induced mutation?	Co-transformation of T7 RNAP without AID with a target under T7
Can it be accounted for by non-targeted AID activity?	Co-transformation of T7 RNAP and AID unfused with a target under T7
Can repair be eliminated?	All experiments to be performed in a repair-deficient (*ung-*) chassis
Can m be assessed quantitatively?	'XFP' reversion assay will allow this

Table 2: Experimental strategy; questions and the experiments designed to answer them

2.4.2. Construction

The requisite control devices (figure 6) and target constructs (figure 10) were successfully cloned into appropriate backbones, with all construction confirmed first by either colony PCR (figures 7, 8) or analytical digest (figure 9), then by gene sequencing. The placement of the devices on high-copy plasmids should result in saturation of transcription at P_{T7} – the targets being placed on low-copy plasmids

(pC5, ~5 copies per cell and pCD, 10-15 copies). Multiple backbones were used for the target to allow potential optimisation of the mutagenesis protocol; higher copy numbers could improve plasmid yields, with a lower copy number allowing more effective screening.

2.4.3. AID retains functionality in the fusion

The cells containing a plasmid expressing AID (whether fused to T7 RNA polymerase or separate from it) exhibited a higher background mutation rate than the cells containing T7 RNAP alone, demonstrating that AID retained its mutagenic effect post cloning and transformation into the chassis, and further, retained it apparently uninhibited when fused to the T7 RNA polymerase (figure 11). This demonstrates that AID activity is comparable in the fused configuration to that of free AID, so any increase in action on the target may be accounted for by a targeting effect.

3. Multi copy target analysis

3.1.1. Aims

Following the construction of the device and necessary control plasmids, the T7-AID fusion protein is ready to be tested to determine its efficacy at targeted mutation. First, it is necessary to develop an assay for accurate assessment of the targeting effect of mutation. This assay should adhere as far as possible to the constraints required for fluctuation analysis of the data, so that a quantitative measurement of the device's activity can be determined.

3.1.2. Plasmid-based systems require segregation of targets

Gain of function assays may be carried out in plasmid-based systems, as successful mutations will cause a change in phenotype in the cell despite the possibility of it retaining plasmids bearing the non-functional genotype. Loss of function assays, however, require separation of plasmids prior to screening as successful knockout of function in one plasmid will not alter the phenotype – given a stable copy number of plasmids in the cell.

Separation of plasmids post-culture is achieved via purification of the plasmids from lysed cells (miniprep) and re-transformation. This also allows the opportunity for

elimination of the device plasmids, which may cause mutation post-selection, by transforming into a chassis expressing T7 RNA polymerase (i.e. a DE3 strain) and selection only for the target plasmid antibiotic; in this case, chloramphenicol.

Assuming the square of the transformation efficiency to describe the maximum proportion of cells which may incorporate multiple plasmids, the vast majority of successful transformants will begin with a single plasmid and therefore represent a single sample of the entire plasmid population. Any cells which take up a device plasmid also should incur a reduction in fitness, there being no benefit to retaining it in an environment without selection for it and some detriment to the cell (increased burden, and expression of a mutagenic protein). This should assist a reduction in the number of post-selection mutations.

After assessment of manual solutions to mass plasmid purification and transformation techniques, it was determined that the protocol was amenable to adaptation for use on a liquid handling robotic platform; all the stages of the protocol after growth to saturation, from inoculation through culturing to plasmid purification and transformation, were theoretically achievable on a liquid handler.

3.2. Results

3.2.1. Automation

A script was written and optimised to allow the purification and transformation of plasmids from cell culture into fresh cells entirely on the liquid handling platform. The script is included in appendix 9.2. These cells were then plated manually, and colonies scored after overnight growth.

This high-throughput work-flow allowed a full run of 288 cultures, comprising the 4 device plasmids incorporating pA3:TA with the three controls (pA3:T, pA3:T+A and empty) and pCD:P_{T7}-GFP, to be processed over two days. The data from this run is shown in figure 12, with raw data in appendix 9.1.1.

52 cultures did not grow up following inoculation, yielded too little DNA to transform or showed zero colonies post transformation. In addition, one entire repeat of the T7 RNAP control failed to grow up in culture. However, the automated protocol was otherwise successful, with the platform purifying and transforming the DNA over the course of the morning and afternoon. Yields of up to 1 µg were realised after substantial optimisation and adaptation of the protocol. Plating was carried out in the evening. Colonies were observed on most plates, with around 20% showing no growth,

after 16 hours incubation. The plates were moved to a 4°C environment to allow maturation of GFP while minimising overgrowth of the colonies.

The number of colonies was lower than expected (figure 12), which was traced to a greater reduction in efficiency of the miniprep procedure than had been allowed for. Of the 3 repeated runs with a control plasmid and the three device plasmids, rarely did the desired 20 cultures translate to countable colonies, and thus the data could not be used to determine mutation rate (Rosche 2000).

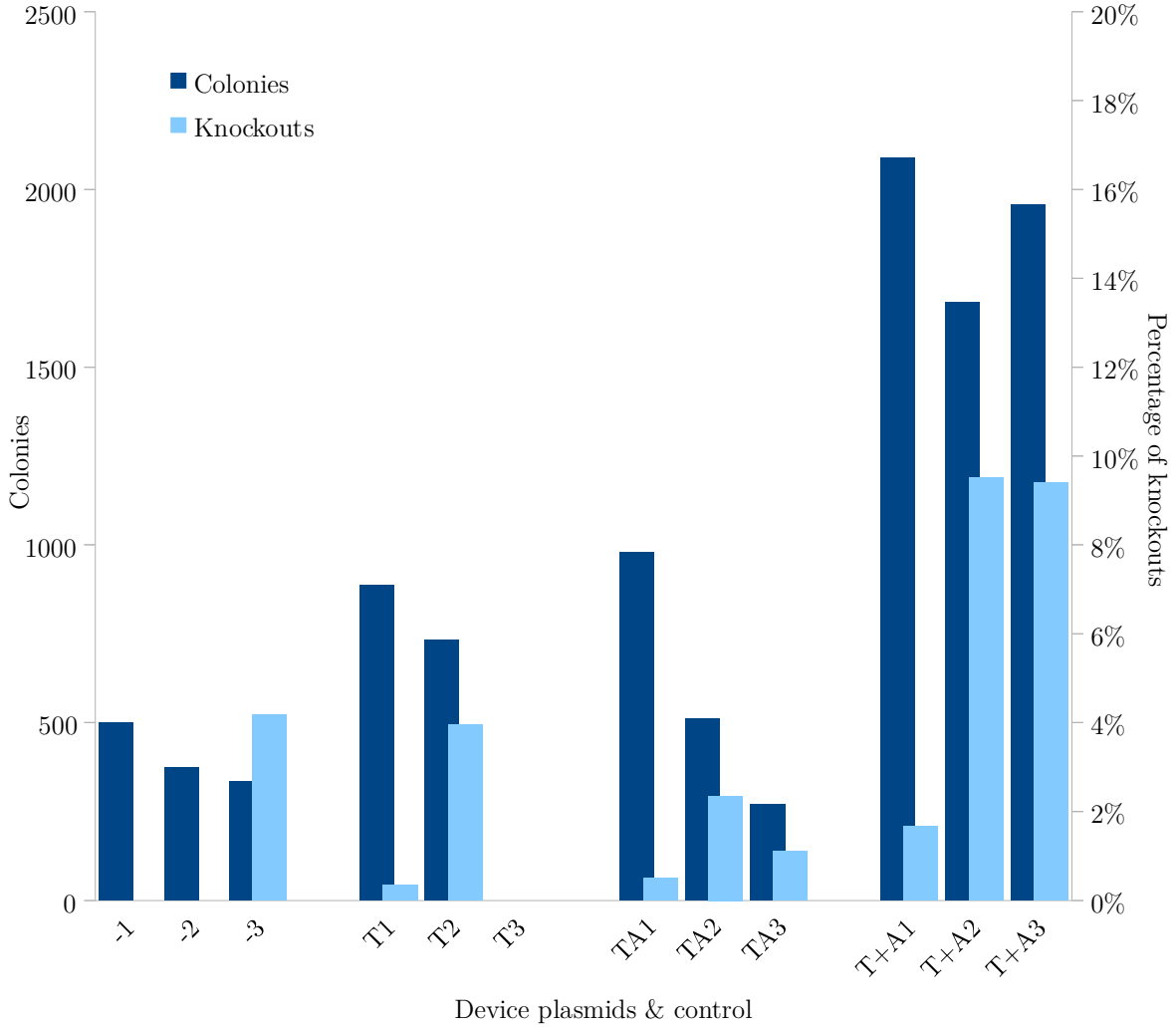


Figure 12: Colony counts from automated plasmid isolation and transformation

Each cluster of three repeats relates to a different experimental condition – left, no device plasmid transformed; second left, T7 RNAP alone transformed; second right, T7-AID fusion and right, T7 & AID unfused. Dark blue bars represent the number of colonies across platings from all cultures (left-hand axis) and light blue bars the percentage of those colonies exhibiting zero fluorescence after transformation into BL21-DE3 (right-hand axis).

The automated protocol was adapted, with the concentration of DNA used to transform being increased in an attempt to improve colony count, but due to a combination of factors including inefficiency of the mini-prep protocol with low-copy pC5 targets, the second run was performed using the alkaline lysis procedure with spin columns (EZNA Silica Column Miniprep Kit, Omega Biotech). The robot was used for pipetting and transformations. This gave acceptable plasmid yields (1.5-4 ug) and colony numbers when transformed into BL21 cells; almost all plates harboured a substantial number of colonies (an average of ~660 per plate), and were moved to a 4°C environment for 48 hours to allow GFP maturation while minimising overgrowth.

Plates were imaged under UV light and colony numbers were scored using colony-counting software. Differentiation of fluorescent vs. non fluorescent colonies was applied via manual level adjustment and curation of each image to ensure accuracy.

The data from these runs for the fused and unfused plasmids are shown in figure 13 (the raw data are tabulated in appendix 9.1.2); the T7 RNAP and cells lacking a device plasmid both showed zero loss of function across all cultures and are not included.

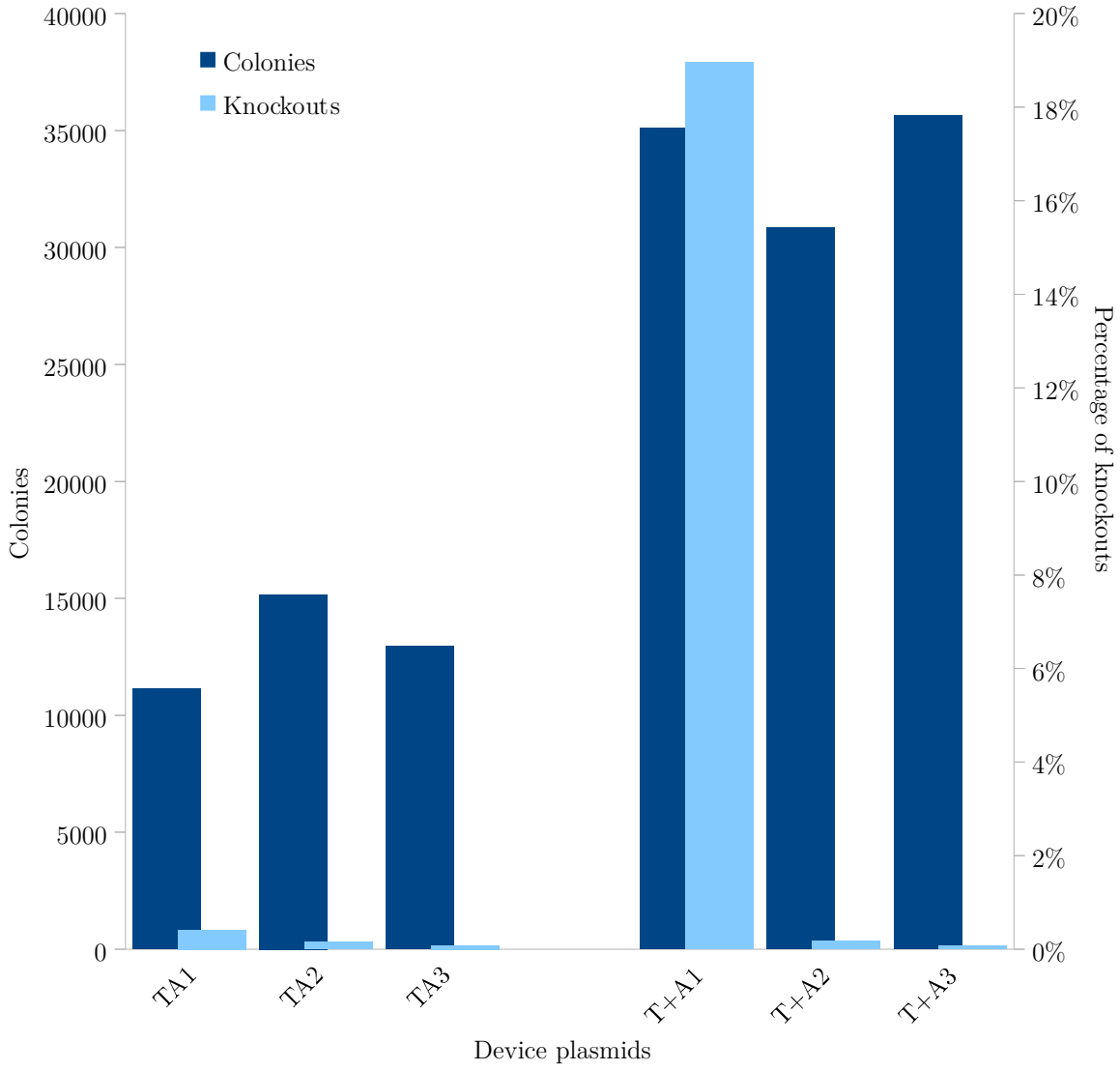


Figure 13: Colony counts from semi-automated plasmid isolation and transformation

Each cluster of three repeats relates to a different experimental condition – left, T7-AID fusion and right, T7 & AID unfused. Dark blue bars represent the number of colonies across platings from all cultures (left-hand axis) and light blue bars the percentage of those colonies exhibiting zero fluorescence after transformation into BL21-DE3 (right-hand axis).

3.3. Discussion

3.3.1. Adaptation to automation

Human error and variability of technique were found to introduce high variability to the yields of DNA acquired during the required segregation of plasmid step. This, as well as the repetitive and time-intensive nature of the process, led to the migration of the work-flow to an automated platform.

Three kits were available for plasmid purification with the liquid handler – two magnetic bead-based (Promega's Wizard MagneSil Plasmid Purification System and Qiagen's MagAttract 96 Miniprep Core Kit) and a column-based (Qiagen's Qiaprep 96 Turbo Miniprep Kit) solution. The column-based kit was ruled out early, as the vacuum system used in place of a centrifuge on the robotic platform suffered from uneven distribution of pressure across the wells in the plates, which would impact the quality and quantity of DNA purified and introduce unwanted variation in yields.

The other two kits were tested manually, to determine expected yields. DNA concentration was assessed by gel electrophoresis and spectroscopy, and confirmed using transformation and plating with colony counting compared to a control. The MagneSil kit yielded very pure DNA, but at lower concentration; the number of colonies post-transformation was below what was required. The MagAttract kit produced acceptable

yields even when reagent volumes were halved, allowing a more cost-efficient protocol. As mentioned, on average ~1 µg yields were reached with the medium-copy plasmid pCD. However, it was less suitable for the low-copy pC5, and thus manual involvement was still required in this instance.

Overall, the use of the liquid handler streamlined an otherwise labour-intensive workflow, which had been shown to impact negatively on the turnover of the procedure to data. It also removed the possibility of mistakes arising from human error from most of the complex, and otherwise standardised (i.e. kits and standard laboratory procedures) portions of the protocol.

3.3.2. Data quality and limitations

As described in Chapter 2, it was decided that to decouple untargeted and targeted mutation by the fusion, an unfused T7-AID device was needed as a control to determine if targeted mutation by the T7-AID fusion was higher than background mutation and a T7 alone device for determination of transcription-induced mutation rates. The data produced by the automated run, shown in figure 12, show that some loss of function colonies were seen in those platings derived from both the AID-containing devices and the negative/T7 RNAP controls. The inconsistencies between the triplicate runs prevents deeper analysis of the data however – while the unfused T7

AID construct demonstrated a substantial percentage of knockout colonies for runs 2 and 3, both above the more consistent level of the fusion, run 1 showed a far lower rate. The T7 alone control, shown in figure 11 to have a negligible mutation rate, showed a loss of function percentage consistent with this in run 1. In run 2, however, it showed a rate above that of all three runs for the fusion – the same pattern was seen in the pure negative control, with run 3 showing some loss and 1 and 2 showing none.

It must be noted that these data cannot lead to firm conclusions due to oversampling as a result of the limited successful transition from culture to colonies (see appendix 9.1.1. for the data tables used to produce figure 12). For this reason, a second run was attempted.

The second run used manual purification of plasmids. While all cultures successfully translated to scorable transformations, there was still a concerning level of inter-repeat variability for the device plasmids containing AID. The negative control and T7 RNAP alone showed zero loss of fluorescence across all plates, in line with expectations and appearing to show that transcription-induced mutation alone was negligible. The devices containing AID showed a consistent level of mutation for 5 out of the 6 combined runs (between 0-0.5% loss of fluorescent colonies). The exception was run 1 of the unfused mutator protein; loss of fluorescence was observed in around 19% of colonies.

This variability in assessment of the rate of knockout of GFP means that an estimate of the masking rate cannot be made with confidence from these data; as discussed in section 2.2.3, this is the required parameter that the loss of function assays aimed to evaluate.

An interesting development observed during scoring of these runs is that, upon close inspection, more of the fused colonies appeared to have lost function when in fact they were weakly fluorescent. The counts in figure 12 are correct, with loss of function/GFP-ve data defined as zero fluorescence, but this suggests that while the fusion is less effective at completely disrupting fluorescence it may be causing knock-down of the output by more sparse disruption of the gene sequence. Responsibility for this pattern lying with some other factor, and the fusion causing no mutation, may be ruled out by the absence of any such pattern in the T7 control set.

An alternative explanation could be co-transformation of knockout plasmids with unmutated plasmids – or unmutated plasmids with device plasmids – causing either an effective reduction in copy number of the wildtype GFP gene or continued mutation post selection and accordingly decreased fluorescence; the assumption that such multiple transformation events were unlikely (co-transformation probability reasonably estimated as the square root of the transformation efficiency) is at the heart of the miniprep/re-transformation procedure. Examining the literature to determine the

validity of this assumption revealed concerning data – Hanahan (1983) shows that co-transformation of compatible plasmids occurred in 70-90% of the transformants in a reaction where cells were saturated with DNA. He posits that rather than all cells in an aliquot having equal amenity to transformation, a smaller subset of those cells are hyper-competent; thus transformation efficiency remains the same but co-transformation is much more likely than previously assumed.

With no selection for the device plasmid, and competition between any distinct target plasmid co-transformations (having the same origin of replication) leading to segregation, it could be expected that by colony growth the cells will have lost the device and only contain one target plasmid. However, the time for this to occur is an unknown variable, and the results may be interpreted differently based on that.

If we assume co-transformation is a common event, with the number of plasmids being taken up by a cell p and the proportion of plasmids with the maskable genotype (in this case, loss of fluorescence) m, we can see that the proportion of colonies which will show that phenotype x can be found by:

$$x = m^p$$

This is an oversimplification, as it ignores post-transformational segregation, and is also the worst-case scenario as it assumes full DNA saturation which may not have been the case as it was in Hanahan's experiments.

Even with those caveats, however, it is clear that because the single plasmid per transformation event assumption is incorrect, the number of colonies counted with a non-fluorescent phenotype will be lower than the number of colonies transformed with a plasmid of the corresponding phenotype on that plate. For this reason, it was decided that further plasmid-based runs were not to be attempted, and the targets were to be adapted to a single-copy assay.

It is worth noting here that the main difficulties – namely efficient scale-up of plasmid segregation without loss of quality, and the need to double the sample size (as a result of the need to examine loss- and gain-of function rates) to account for masking – are not issues which would affect a plasmid-based application of the system toward directed evolution. Once even the approximate mutation efficiency of the device is known, masking can be accounted for or disregarded, and the need to segregate plasmids pre-selection would be eliminated in gain-of function experiments.

4. Single copy target analysis

4.1.1. Aims

The issues encountered during the plasmid-based testing stages, and detailed in chapter 3, forced a necessary change in direction for the project's approach. Further attempts to optimise the multi-copy system for assessing the mutation rate of the system were deemed intractable, as the previous heavy optimisation and automation of much of the process still resulted in a rate of data generation that was not sustainable, with any further improvements providing only diminishing returns.

With the above considerations in mind, the project was refocused toward a single-copy testing strategy. This opened up new avenues for the screening of constructs, namely by the use of flow cytometry to quantitate GFP fluorescence on a single-cell basis. The high-throughput nature of flow cytometry also allows many more distinct cells to be examined than may be reasonably seen by plating and scoring colonies.

The first aim for this chapter, therefore, is to integrate a target onto the host cell genome. This is to be followed by high-throughput analysis using flow cytometry, to examine large cell populations quickly and efficiently. Lastly, target sequences are to be examined directly via sequencing to determine mutation efficacy and any relevant patterns.

4.2. Results

4.2.1. Genome integration of targets for single-copy analysis

Genome integration used the CRIM (for Conditional Replication, Integration and Modular) plasmid system (Haldimann 2001). A variety of CRIM plasmid pairs exist. For the purposes of this project integration plasmid pAH153 and helper plasmid pAH123 were selected, for two reasons; first, they used the Φ-80 integration site, demonstrated to possess a high efficiency of integration by the authors and second, pAH153 relied upon a gentamycin resistance marker on the integration construct. This antibiotic was not used on any of the other plasmids in the mutator system, and would facilitate cloning and screening by allowing elimination of carry-over as it was distinct from that in the original construct (chloramphenicol).

The target genes – GFP and XFP – were separately cloned into the CRIM integration plasmid using CPEC (see Methods section 7.2.7, primers CRIM_Gene_F/R & CRIM_153_F/R) to produce pCRIM:GFP & pCRIM:XFP (figure 14). Biobrick flanking sequences were included, to allow for future cloning by that method should it be required. This also has the benefit of enabling more consistent CPEC/Gibson (see Methods section 7.2.6) cloning, as previously the insertion site was directly flanked by transcriptional terminators.

Upon confirmation of correct construction via sequencing, the plasmids were co-transformed into competent BH156 cells. Successful integration was achieved following the CRIM procedure. This was confirmed (see figure 15) by the 4-way PCR described by Haldimann & Wanner (2001), with some primer sequence alteration to account for the altered flanking sequences (primers CRIM_1/2 & 4 were used as described by Haldimann et al, but CRIM_P3 required modification to account for sequence differences at its site – this primer is CRIM_T3). A positive result, with a single integration, produces two bands; a negative result, one band and a multiple insertion event, three. Figure 15 shows that both XFP and GFP targets were integrated successfully. Successful integration was further confirmed by sequencing of a colony PCR encompassing the entire integration site.

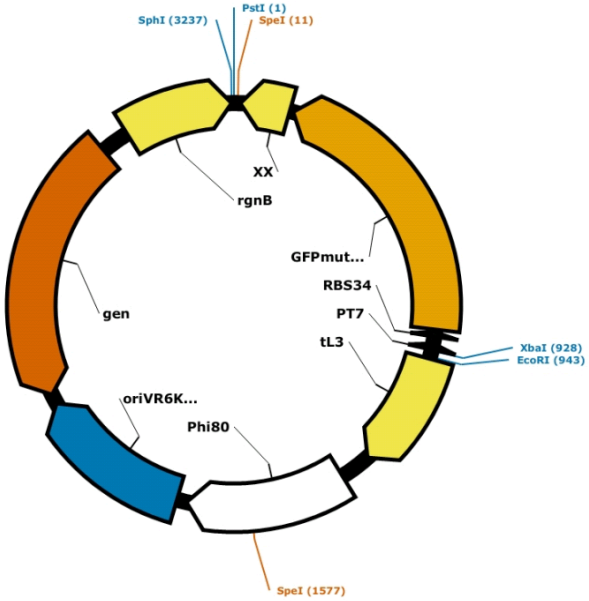


Figure 14: GFP target cloned into pAH153 for integration onto the E. coli genome

Plasmid pCRIM:GFP will allow genome integration of the GFP target via the CRIM integration system. pCRIM:XFP is identical, with GFP replaced by XFP.

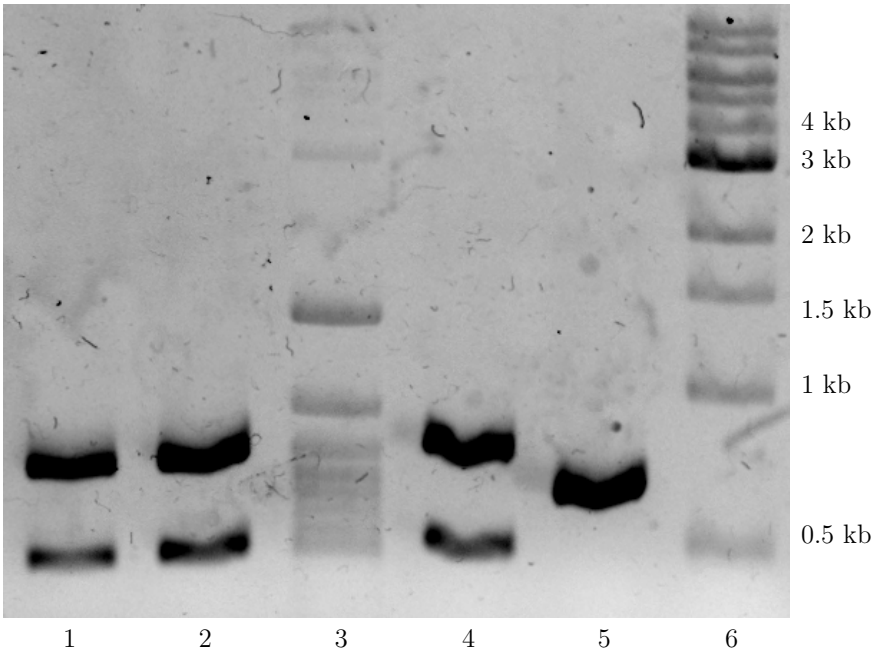


Figure 15: Agarose gel showing the products of an analytical PCR demonstrating successful integration of GFP/XFP

Lane 6, 1 kb DNA ladder; Lanes 1 & 2, colony PCRs of strains with GFP (lane 1) and XFP (lane 2) integrated into the genome by CRIM; Lane 4, positive control; Lane 5, negative control.

Curing of the helper plasmid pAH123 – which carries an ampicillin resistance gene, making it incompatible with the device plasmids – was effected by repeated rounds of re-streaking and positive/negative selection, and was confirmed by colony PCR against helper plasmid DNA compared to controls.

The strains with integrated targets were termed BH156-GFP and BH156-XFP. These were made competent, ready for transformation by the device plasmids; colonies were picked post-transformation, and grown up for glycerol stock as expeditiously as possible to minimise mutation in this period. These glycerol stocks were then used as seed cultures for the mutation runs, which were otherwise carried out following the protocol for the plasmid-based target assay (minus the unnecessary target antibiotic).

These runs produced cultures that were diluted in double-distilled water (ddH_2O) for analysis on a flow cytometer.

4.2.2. Analysis of mutation of single-copy targets by flow cytometry

Early culture runs, following the procedure for the plasmid-based assays, produced several millilitres of culture – since only 20-50 µl are needed per sample for analysis by flow cytometry, it was determined that smaller culture volumes could be used. This allowed optimisation of the protocol, and specifically, allowed adaptation of the culture step to a semi-automated procedure.

Adapting the protocols previously designed for the liquid handler to culture growth was relatively trivial. A second liquid handling platform, of the same make and model, included a plate incubator which allowed temperature and shaking control of 96/384-well micro-titre plates integrated into the platform.

This allowed time-course experiments to take place; the automated platform re-diluting the cultures throughout the day and overnight and saving samples on a cooled position and diluted in ddH_2O to stop further growth.

A negative control, of BH156-GFP cells without a device plasmid – and therefore lacking the T7 RNA polymerase required to transcribe GFP from the genome – was

used to calibrate the machine for cell size and complexity, and to determine a suitable dilution coefficient for the samples which avoided multiple-cell events.

A positive control, pA3:T, was transformed into the BH156-GFP strain. This was used to calibrate the machine for sensitivity; the gain of the detector was increased to allow both fully-fluorescent cells and those with no fluorescence at all to be visible as distinct peaks. Standard settings cluster non-fluorescent cells on the Y-axis, giving misleading data, and an accurate distinction of fluorescent and non-fluorescent cells is a vital requirement of the assay.

With the machine calibrated suitably, the settings were saved and used for every run beyond the first to ensure consistency of data and allow inter-run comparisons.

Flow cytometry results for loss of function assays, with device plasmids transformed into BH156-GFP, are shown in figure 16. The unfused mutator causes drastic loss of fluorescence when compared to the fusion, which shows a less marked effect. Both show reduction in fluorescence relative to the positive control, T7 RNAP alone, which consistently showed no loss of fluorescence over the course of these experiments. Figure 17 shows data from a longer run with the unfused device plasmid; four 8-hour timepoints were assayed. Data from both manual and automated runs were in

agreement as to the relative effects and effectiveness of the device strains, and a mixture of both approaches produced the data shown in figures 16 & 17.

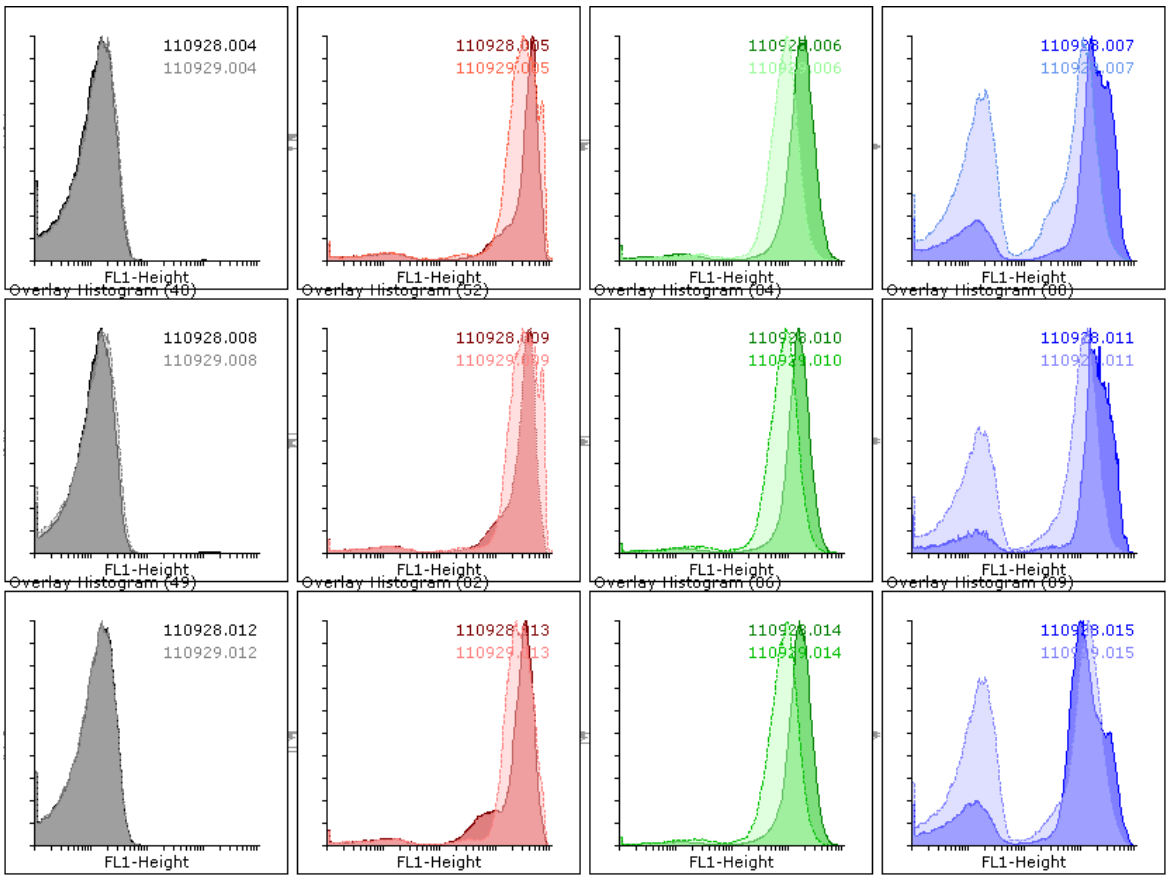


Figure 16: Time-course flow cytometry data showing loss of function profiles for the device plasmids

Each row represents a repeat with an independent culture. From left to right are histograms representing cell counts for a negative control (pUC19) in grey, T7 alone in red, TA fusion in green, and T+A unfused in blue. The darker peak was sampled at 8 h post inoculation; the lighter peak at 24 h after 2 cycles of re-dilution and growth for 8 h in fresh media. Fluorescence is shown increasing logarithmically from left to right on the x-axis.

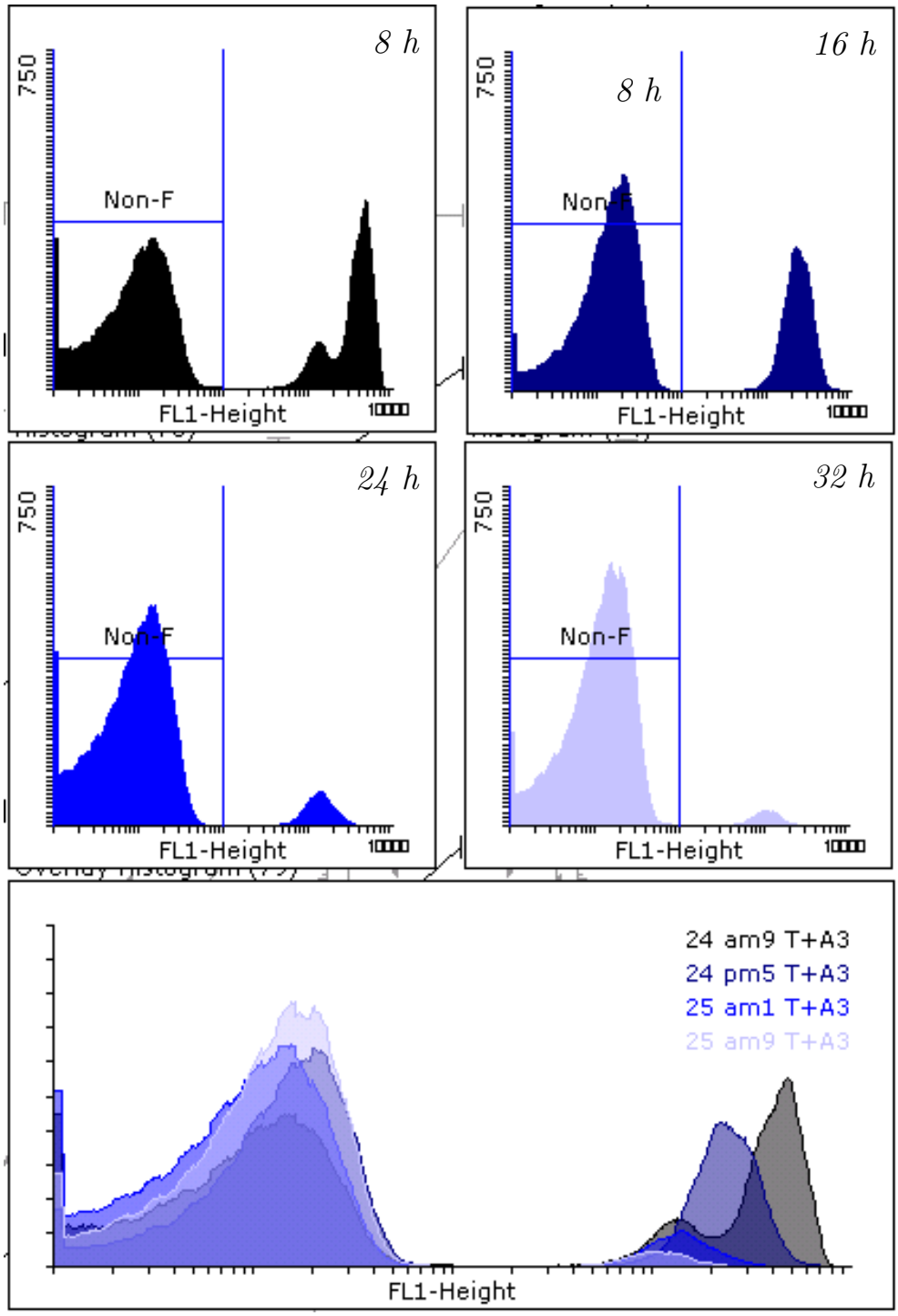


Figure 17: Time-course flow cytometry data showing loss of function profile over 32 hours for the unfused device plasmid

Top left to bottom right: Four 8 hour timepoints were sampled, and the individual graphs combined to highlight the progressive accumulation of loss of function mutants over time.

4.2.3. Sequencing of genome-integrated GFP target

A number of colonies representing each of a stable fluorescent and knock-out phenotype were selected for sequencing from both the fusion and unfused device plates. The T7 RNAP positive control plates did not show any knock-out colonies, so only fluorescent colonies were selected. Elimination of short or low-quality reads left a set size of ten sequences for the control and the fusion fluorescent and non-fluorescent categories, and eight and seven acceptable reads for the unfused fluorescent and nonfluorescent categories, respectively.

These were aligned to the DNA sequence of the GFP construct as verified during construction using the inbuilt alignment feature in ApE (alignment used the Needleman-Wunsch algorithm (Needleman 1970), requiring cropping of sequences and reference to similar lengths, block size 10); basecalls from the sequencing company were initially used to seed the alignment. Post-alignment, each putative mismatch/insertion/deletion mutation event was examined and located within the corresponding sequencing trace file. Erroneous basecalls were corrected where obvious; where significant doubt existed as to the correct call, basecalls were accepted. Appendix 9.3.2 shows sections of the relevant trace files corresponding to each mutation event. All recorded mutation events were observed in sequences amplified from cultures containing the T7-AID fusion device; T7 RNAP alone and T7 RNAP and AID unfused culture

sequences all corresponded to the original GFPmut3b sequence, for both fluorescent and non-fluorescent colonies.

Figure 18 shows the sequence of P_{T7}-RBS34-GFPmut3b, with highlighted letters indicating positions where at least one mutation event was recorded. C to T or G to A mutations make up the vast majority of recorded events, representing 16 of the 19 mutations. Note one position had 2 distinct mutation events, and a second position, 4 – this is represented more clearly in figure 19, which maps the distribution to the amino acid sequence of the GFP target. Full details of the mutation events are in table 6 in appendix 9.3.1.

Figure 18: Mutations in the GFPmut3b sequence induced by exposure to the mutator protein

Uppercase; open reading frame, start codon in green. Hollow letters indicate mutations not resulting in an amino acid change; black, those that did. Underlining indicates C to T or G to A mutations, i.e. those that may have been cause by AID action. Italics are used where more than one event was observed at the same point across multiple samples.

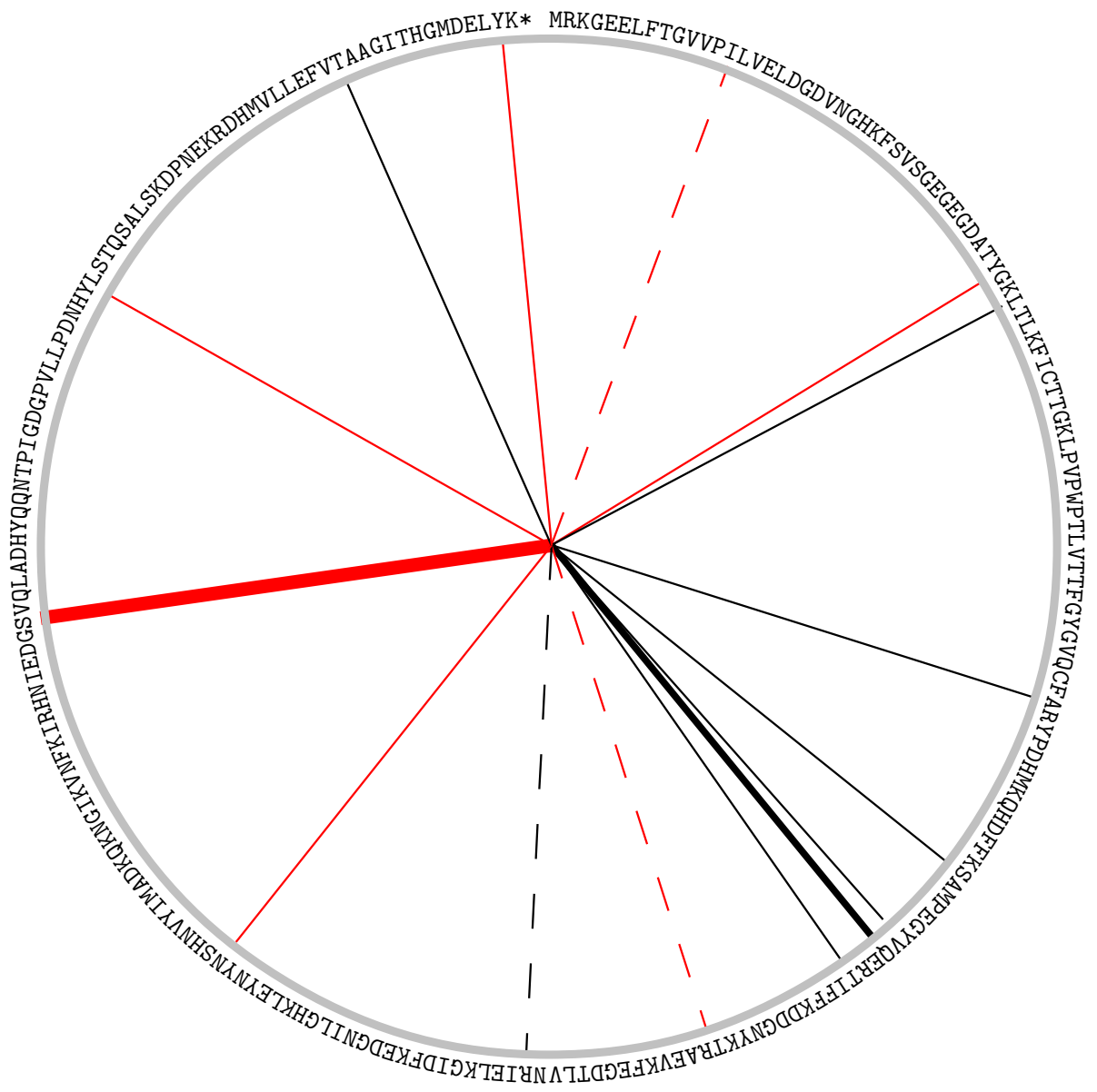


Figure 19: Map of observed mutation events along the GFPmut3b target gene

The amino acid sequence of GFPmut3b runs clockwise from the top; intersections with lines indicate a mutation event was detected in at least one sample at that point. Red lines indicate mutations not resulting in an amino acid change; black, those that did. Solid lines indicate C to T or G to A mutations, dashed indicate other events. Thicker lines are used where more than one event was observed at the same point across multiple samples.

4.3. Discussion

4.3.1. Production of strains with genome-integrated targets

To allow use of the system on a single target, integration of that target onto the genome of the chassis strain was required. This would guarantee a single copy is present at all times (excepting during replication events, an exception which may safely be ignored) and ensures stability of that target in the host. The device and targets used above were still viable for use in a genome-based system, requiring only that the targets be adapted into a form amenable for integration.

It was decided that the CRIM system developed by Haldimann & Wanner (2001) was the easiest and most fitting method to use. A plasmid-based system, it had already been used in a synthetic biology context (Anderson 2010) and involved only routine laboratory protocols – cloning, plasmid propagation, transformation and curing, all of which had established and optimised work-flows in the lab.

The CRIM protocol relies on co-transformation of the target construct plus a resistance marker on a plasmid which is unable to replicate in normal bacterial strains, with a helper plasmid that causes recombination of sequences flanking the target with specific sites on the genome of *E. coli*. Thus, only by successful integration will the cells exhibit resistance to the chosen antibiotic. The helper plasmid contains a heat-sensitive origin

of replication, and may be cured post-transformation by an elevated temperature incubation and negative selection for its antibiotic resistance.

In this way, the GFP and XFP targets were successfully integrated into BH156 at single copy by cloning them into the integration plasmid (propagating the resulting construct in a *pir*+ strain which allows replication of the plasmid) and carrying out the co-transformation and selection/curing procedure. The resulting strains were called BH156-GFP and BH156-XFP.

Transformation of device plasmids into these strains, and the subsequent culturing procedure, followed the protocols developed for the multi-copy analysis work-flow.

4.3.2. Flow cytometry

The flow cytometer allows a population to be assessed with single-cell precision in a much more reliable way than via e.g. dilution and plating, with the added advantages of far greater sample sizes and elimination of possible mutations during the post-culture incubation phase. Since the fluorescence detector matches the emission spectra, this allowed the use of the GFP constructs without alteration – and by quickly examining tens to hundreds of thousands of individual cells, patterns in mutation can be detected easily by looking for distributions of cells within sub-populations defined by their fluorescent characteristics.

The positive control run with T7 RNAP alone used to calibrate the cytometer showed effectively zero loss of fluorescence after a 16 hour incubation, and this pattern was replicated in all later runs with the T7 RNAP device plasmid. This was consistent with the preliminary data from the plasmid-based plate assays and demonstrated that, while transcription-induced mutation is a recognised effect (Beletskii 1996; Francino 1996), and T7 RNAP transcribes at an elevated rate (with a single-copy target, we expect almost continuous transcription along much of the gene), any resulting knockouts arising from said mutation are negligible within the context of the overall population and thus that any contribution to knockout populations in the runs using other devices may be safely ignored.

The plasmids incorporating an AID moiety consistently conferred a mutator phenotype, as evidenced by a drastic increase in the number of knockouts seen within the populations examined on the flow cytometer. The fusion showed a smaller peak of non-fluorescent cells after 24 hours culture (3 repeated dilution/growth cycles), with most cells retaining the fluorescent phenotype; this was routinely mirrored in the unfused device plasmid samples, with most cells having lost fluorescence and a smaller population retaining it post culture. After 24 hours approximately 50% of cells cultured with the unfused device had lost fluorescence (figures 16 and 17).

These data suggested that the free AID device was the more proficient at causing mutation, as it was consistently effective at knocking out fluorescence in the genome-integrated GFP target. To elucidate further the exact rates and distributions of mutation events, examination of the sequence of the genomic target post-culture was required.

4.3.3. Sequencing

Post culture and analysis, remaining volumes of samples were plated. This allowed colonies to be selected for colony PCR amplification of the GFP gene from the genome; these were then purified and sent for sequencing to allow direct analysis of the mutation profile of a number of colonies – both fluorescent and non-fluorescent – for each device plasmid. This work is ongoing, but data collected so far are examined below along with analysis and discussion of how they might inform future work on the project.

Every confirmed mutation event (displayed in figure 18) occurred in colonies containing the fusion device plasmid. This was an unexpected result; the flow cytometry data demonstrated clearly that the unfused T7 and AID protein expression strategy was more effective at knocking out fluorescence more quickly. This does not seem to be a

result of more mutation within the actual GFP gene, however, given the lack of mutation events within the target confirmed by sequencing.

This distribution appears to show that while the unfused strategy is more effective at disrupting the expression of functional GFP (as seen in the flow cytometry data), it does not necessarily accomplish this by directly interfering with the GFP sequence itself – it may be causing mutation elsewhere in the genome and indirectly causing elimination of the fluorescent phenotype. Importantly, the degree of targeting accomplished by the fusion of the AID to the T7 RNAP appears very high if the results of the sequencing run are representative of the wider action of the devices.

Most of the mutation events recorded were from the strains which exhibited a loss of fluorescence; of the 19 events, 14 were from colonies that had lost function and 5 from those still expressing functional GFP at the time the target gene was amplified for sequencing. This is in line with the assumption that much of the loss of fluorescence for the fusion device results from direct disruption of the GFP protein; of the ten fusion knockout sequences, eight display at least one mutation within GFP and 6 have at least one mutation resulting in an amino acid change.

Figure 19 shows the distribution of mutations within the GFP gene, and highlights the prevalence of mutations which are C to T (or G to A), and therefore may likely be

attributed to AID activity. Mutation events appear evenly spaced throughout the sequence, with the possible exception of two clusters; one (at the 5 o'clock position in figure 19) where 2 mutation events were detected at the same base, with 2 more within 3 amino acids of that site, and another (at the 9 o'clock position) where a C to T substitution occurred at the same base pair in 4 independent cultures. This may be due to an uneven transcription rate along the DNA; if transcription progresses slowly or stalls at a point, that area is exposed to assault by AID for longer.

The lack of mutations at the promoter site is a promising result for the tractability of the system, but the large number of knockouts seen in the unfused samples are concerning – as figure 11 showed, the fused and unfused AID constructs exhibit similar levels of background mutation. Knocking out T7 RNAP functionality would be a boon to the cell, as it would remove the burden of T7 transcription placed on its metabolism. If this is happening with unfused AID, it may also be happening with the fusion. It is possible to ensure T7 RNAP functionality is retained, by expressing a required antibiotic from a T7 promoter; T7 RNAP knockouts would then not survive.

5. Implementation of the system toward an orthogonal signal/response pair

5.1.1. Aims

In this chapter, a proof of concept system is devised to apply the T7-AID fusion towards a real-world application. The chosen target gene is the LasR transcription factor from *Pseudomonas aeruginosa*, and the objective is to diversify its specificity for its cognate signal molecule. The first aim, therefore, is to integrate LasR into the *E. coli* genome behind P_{T7}.

Secondly, control circuitry is designed to allow the system to respond to and select for favourable mutations arising from the action of the fusion protein. This circuitry must operate *in vivo* and in real time, to maximise the chance of successful mutation and minimise the chance of reversion of that favourable mutation by continued attack from the fusion. The second aim is to construct the genes required for this system on plasmids, and to co-transform them with the T7-AID device plasmid.

Finally, the functionality of the control circuitry will be assessed.

5.2. System overview

5.2.1. The mutation target: a quorum-sensing system

As a test case for the system, and as the first attempt to utilise it toward improving a specific target for a specific need, it was decided that diversification of a signal-response pair would be both approachable and of benefit to the synthetic biology community. The chosen module was the *lasR*/P_{Las} system, which responds to levels of the acyl homoserine lactone (AHL) N-(3-oxododecanoyl)-homoserine lactone, or 3-oxo-C_{12}-HSL, with increasing transcription from the P_{Las} promoter (Passador 1993; Whiteley 2001). In the synthetic biology vernacular, the *lasR*/P_{Las} system is "AHL in, PoPs out", where PoPs stands for polymerases per second and represents a measurement unit for transcription rate at a specific DNA sequence. A brief overview of the *lasR*/P_{Las} system is given here.

One of the most widely-studied, and well understood, areas of population dynamics in bacteria is that of quorum sensing (Miller 2001). Quorum sensing is the mechanism by which bacteria assess the concentration of population in their local space. As the density of a population of bacteria increases, the bacteria must adapt; slowing growth and often producing structures or proteins of benefit to their neighbours. These can be as simple as a low level of antibiotic (Bainton 1992; Duerkop 2009) to attempt to

cleanse the area of competing species, to complex structures such as biofilms, which often signal the triggering of virulence factors in pathogenic species (Erickson 2002; Rutherford 2012). Without this behaviour, growing bacteria might use up all the resources in an area and cause starvation of the entire population, inhibiting growth or survival of that species.

Quorum sensing utilises a signal molecule – generally a chemical such as acyl homoserine lactone, or AHL, in the case of Gram-negative bacteria or an oligopeptide in the case of Gram-positive species – which is constantly expressed by all members of a population to gauge population density (Miller 2001). A high concentration of the signal molecule implies a high concentration of fellow bacteria in an area, and this can then trigger transcription of genes to take advantage of that via direct action on a promoter (Gram-negative) or two-component system (Gram-positive) (Bassler 1999).

In the case of the Las system, which is the quorum sensing system in *P. aeruginosa*, the key promoter is P_{Las} (Rust 1996; Whiteley 2001). It initiates transcription of *lasI* and *lasR*. *lasI* encodes a gene which produces 3-oxo-C_{12}-HSL, the quorum-sensing signal molecule which may diffuse out of the cell into the local environment, and *lasR* encodes a protein of the same name which, when bound to that AHL, activates the P_{Las} promoter. Thus in its native context there is a level of positive feedback; P_{Las} produces both LasR and the AHL required to allow it to activate P_{Las} (Seed 1995). If activated

beyond that baseline point, P_{Las} also triggers the transcription of virulence genes (Winson 1995).

These systems are of great interest to synthetic biologists for many reasons – chief among them, the hope that biosensor systems designed to detect and respond to these signal molecules may allow us to modify the behaviour of infectious bacteria. *P. aeruginosa*, for instance, is a common bacterium that can cause disease in humans and can be found in a range of environments including the soil, within buildings, and as part of human skin flora. It can cause problems in hospitals when it forms biofilms on indwelling medical devices such as catheters, for instance (Parsek 2003).

But there is a secondary reason for the interest in these types of systems, and that is that they provide an example of a device that could be of great use in designing genetic circuitry. Signal-response pairs allow communication not only within the cell itself, say in response to external addition of their inducer, but also between cells; this can be used to 'synchronise' a population, for instance, or to propagate a reaction (e.g. a target is detected locally by a subset of a population, which produces the signal molecule and propagates the signal out to the rest of the population allowing sensitive detection to produce observable output).

When building a genetic circuit, methods of communication between different parts of that circuit quickly become the limiting factor to complexity. This is due to crosstalk; one cannot utilise the same signal-response pair twice in a cell to different ends, as when one triggers it will unavoidably set off both. Thus a variety of devices allowing connection of two nodes of a circuit without repetition is paramount. Unfortunately, a limited number of these devices exist which also fulfil other usual requirements – reliability, non-leakiness, ease-of-use and so on.

The LuxI/LuxR pair from *Vibrio fischeri*, for instance, is commonly used in synthetic biology projects (Balagaddé 2005, Basu 2004, McMillen 2002, You 2004) but has a drawback in that it is relatively promiscuous; it responds to a number of different AHL molecules to different degrees, inhibiting the use of different quorum sensing systems together. Figure 20 shows the response curves of part BBa_F2620 (LuxR/P_{Lux}) when presented with various AHL molecules (Canton 2008). It can be seen that it responds well to a number of similar AHL molecules – if it could be mutated to lose the ability to respond to these, and to instead respond to 4AHL for example, an orthogonal signal-response pair would have been created which could be used in tandem with the original F2620 device in a circuit.

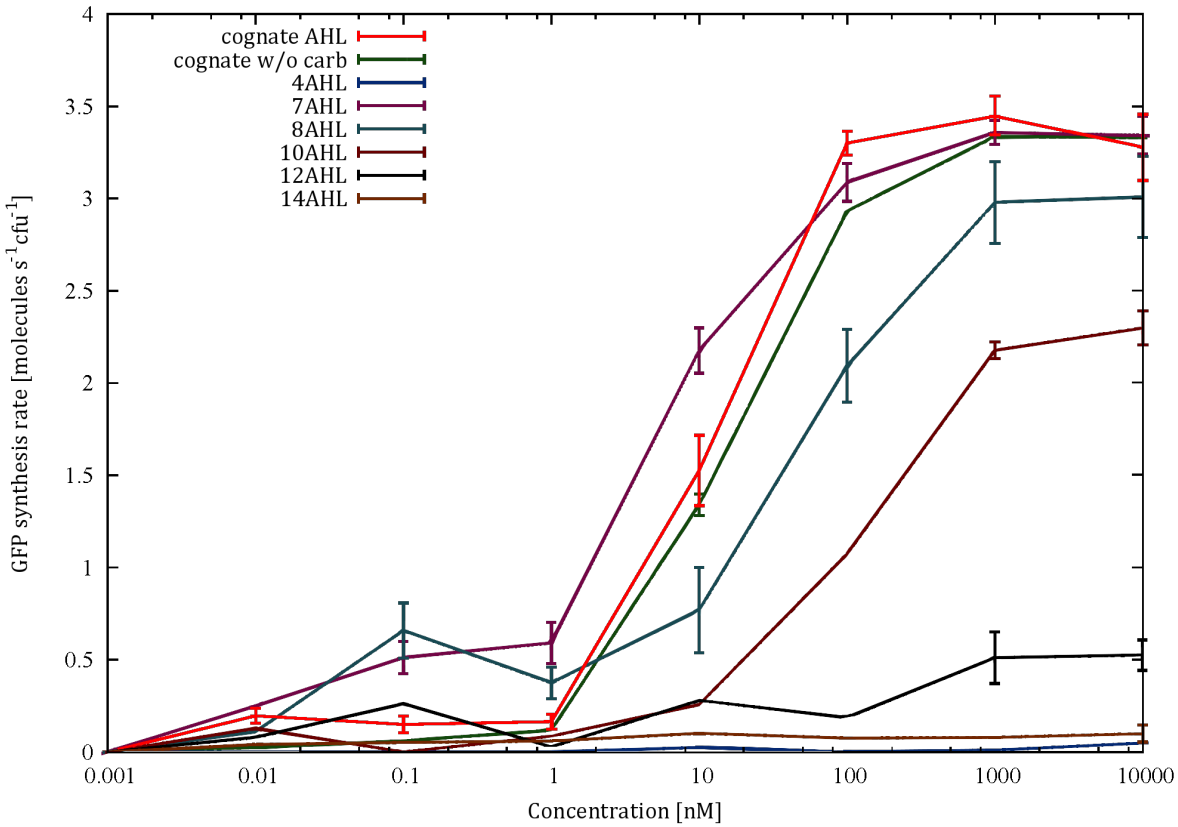


Figure 20: Response profile of the P_{Lux} system to a range of AHL molecules

Adapted from Canton 2008, this graph shows the response of the P_{Lux} inducible promoter to a range of AHL molecules at increasing concentrations.

Taking LasI/LasR as our test case, the ideal outcome would be to produce a signal-response pair that specifically detects a different AHL than its cognate signal molecule, 3-oxo-C_{12}-HSL. This then opens the door for evolution of multiple orthogonal variants of the system, each responding to a different inducer; these could be used in parallel with other AHL-based systems within the same population or the same cell, allowing complex communication and circuitry to be built on that. It is potentially even possible to produce systems that can detect non-AHL type molecules, and thus generate completely orthogonal bacterial signalling systems or new molecular recognition for use in biosensor devices.

To attack this problem, the target gene (*lasR*) would be placed behind the T7 promoter to expose it to transcription and mutation by the fusion protein. The next step is to design a mechanism whereby the evolution of a desired variant can be detected and selected for; a control system is needed.

5.2.2. The control circuit

One of the key benefits of the mutator system described in this report is the ability to continue mutation over long time periods. Eliminating the need to transfer the DNA from an *in vivo* to *in vitro* environment, and vice-versa, means that continuous generation of variation is possible; this is within the confines of the heightened

background mutation rates and assuming stability of growth, whether using batch culture, continuous flow, or other methodology.

This advantage can only be utilised effectively however, if there is some method of eliminating 'overmutation' – that is, mutating past the peak or phenotype which is aimed for. If culturing a large batch of cells and allowing mutation by the system, absence of real-time selection methods may cause any successful genetic variants to be rendered useless or unproductive by continued mutation.

If the phenotype being searched for is an obvious one, as in our test cases with GFP, it is relatively trivial to design a work-flow; a fluorescence-based cell-sorter could separate out successful reversions to function of GFP from samples of a culture, for instance, or be optimised to select out 'interesting' variants of GFP for example. But what of those products that are not so amenable to selection?

In this section a feedback loop is designed with an eye to flexibility of use. Primarily for use with the chosen exemplar case to be conducted with the mutator (AHL/LasR pair), it could be altered to function with any protein that can generate a PoPs output.

The primary design constraint is that upon successful binding of a new AHL to the mutated LasR protein, mutation should be halted or at least slowed drastically. To do this, a repressor system operating on the promoter producing the AID device is the

obvious choice. With that said, continued expression of the output molecule is required for operation of the loop, and thus an alternative polymerase (T7 RNAP alone) should be produced along with the repressor; this has the added benefit of immediately competing with the fusion protein for access to the T7 promoter, reducing transcription by the AID-carrying variant and resultant mutations. Degradation tags added to the fusion protein could complement and amplify this effect.

The secondary consideration is that of selection. Devising a way to determine which cells in a culture are 'successful' as far as reaching some target phenotype is important. With transfers in and out of cells eliminated the limiting step of this methodology, as far as maintaining the exploration of the greatest sequence space is concerned, is the screening process. Self-reporting, or self-selection, is a powerful tool in this regard.

To this end, the two most feasible outputs to screen for are fluorescence- and antibiotic selection-based. Both could allow selection followed by reintroduction to culture. Antibiotic selection-based methods seem more attractive, as they allow the elimination of all failures at a stroke, leaving behind only those candidates who are expressing the resistance marker, but they also have the downside of potentially killing those cells which are expressing the marker but at too low a level. It was decided that the system would be designed with both mechanisms, to allow them to be tested and weighed against each other; the use of both in parallel, elimination of the vast majority of cells

by antibiotic selection, and the examination of the rest using fluorescence-based tools, would be optimal in the end case.

The completed system design is shown in figure 22.

As the feedback loop initiates, the mutator and unfused T7 will compete for the T7 promoter. If the target is not producing a strong enough output (say, only very weakly binding the AHL molecule) then it is conceivable that the mutator may not be fully repressed until further mutations strengthen the target or eliminate its activity. This allows the mutator system to scale mutation as a target functionality is approached, a powerful additional benefit, though it does carry with it an attenuated risk of over-mutation. Such an effect may be tuned by the strength and reliability of the repressor system.

5.2.3. Leakiness

Inducible promoters are one of the most widely-available, and widely-used, parts in the synthetic biologist's tool kit. They add functionality to a circuit to allow output to be controlled by some input, often an inducer molecule. This control function may be binary (on/off in the presence or absence of inducer) or may follow a dose-response curve, allowing finer control and responsiveness for a circuit – as exemplified earlier in figure 20, the response curves of LasR/P_{Las} when presented with varying AHLs.

One common feature of inducible promoters which is generally seen as a drawback is leakiness. This is the baseline expression from the promoter in the absence of inducer; in a simple biosensor, for example, leakiness could cause continued low-level expression of the signal molecule which may build up and trigger a false-positive result. The mechanisms by which a promoter may be leaky vary, depending on the type of interaction that is blocking transcription normally – some promoters rely on physical structures within their DNA and associated operator regions to interfere or obscure the binding site until disrupted by some inducer; others may require some cofactor which can only bind in the presence of their inducer and so on. Thus leakiness may be realised by simple stochastic disruption of an inhibitor, by interference/competitive binding from a homologous molecule, or by low-level background levels of inducer and so on.

Several approaches designed to eliminate leakiness from, or produce tight alternatives to, available inducible promoters have been tried. Beyond specific engineering of promoter control sequences in an attempt to eliminate stochastic disruption (Alper 2005), or construction of novel promoters, some attempts have been made to alter the downstream bases into the RBS and gene (Guzman 1995). One promising and diverse avenue of research is RNA control; for instance, using a stretch of interfering/silencing RNA to inhibit ribosome binding by attaching to mRNA at that point and denying

translation (Morita 2006). These efforts progress to the more ambitious and innovative schemas such as the use of recombinases to invert a target, fully denying transcription until the gene is reoriented the correct way around (Ham 2006; Friedland 2009).

Cell-cell variability is the name given to the phenomenon whereby a range of phenotypes may arise in a given population of identical cells in an identical environment (Elowitz 2002, Pelkmans 2012). Stochastic effects combine with the cells' genetic 'programming' to allow diversification. This is evolutionarily favourable, allowing exploitation of various niches or substrates within that environment, or survival of some sub-population following a catastrophic event affecting the main population, but it is also a source of frustration in the laboratory; all cells are not equal, and population effects and single-cell effects may not be linked with ease.

Extrinsic noise, that arising from the environment, may be controlled for in the laboratory; intrinsic noise, however, the variability of expression inherent in any process reliant on biochemical reactions at the molecular scale, cannot. One major component of that noise, an amplifying factor, is the phenomenon of burst transcription. When an mRNA is produced it causes some variable number of translation events and each of these 'bursts' of resultant protein production are distributed randomly and independently, following a Poisson distribution (Spudich 1976, McAdams 1997).

Linking this idea back to leakiness, it can be seen that even relatively rare transcription events such as those arising from a leaky promoter may cause non-trivial production of protein and interfere with downstream circuitry.

In the case of the mutator system, this is of great relevance when designing the feedback circuitry which is supposed to shut down the mutator upon successful production of a target.

Control of leakiness in such a circuit is of paramount importance if that benefit is to be realised; continued expression of AID could cause loss of useful function generated by that molecule. Additionally, leakiness of the 'shut down' circuit when it is not required would cause a decrease in efficiency of the mutator system in culture.

Looking at the test case that was chosen – the AHL/LasR response pair – it could be seen that these promoters have a degree of leakiness and non-specificity (Chappell, 2013). It would be valuable, therefore, to be able to implement a control loop even under the constraints of the leaky promoters.

5.2.4. Bacterial logic gates, circuit design

To do this, an engineering solution to the problem was devised. Using a logic gate, a ubiquitous feature in computing and relatively commonly approached as a project by

synthetic biologists, it was hoped that the unique properties of a genetic circuit may allow construction of a novel gate device which could act as a 'high-pass' filter, filtering out low-level activation of a promoter while retaining the action at higher concentrations of inducer (including the dose-response profile).

At the simplest level, the proposed device is an AND gate with two identical inputs. A non-sensical device in electronics, several properties of genetic circuits give reason to believe that such an arrangement could have interesting properties in the genetic context: the stochastic nature of gene activation and transcription, discussed earlier, and the delay between that activation and expression of the output molecule key among them.

With a leaky promoter in front of an output gene, any stochastic firing from that promoter will result in some number of mRNAs. Each of those mRNAs may then result in some variable number of proteins being produced. Placing the gene behind two identical promoters in series may cause increased expression, both reading into the gene; placing them in parallel on the DNA, such that both must be activated to allow transcription, is not an available option. If we use a genetic logic gate, however, and decouple the input and output via a mediator, we can achieve the functional equivalent.

If two identical promoters trigger the production of two non-identical products, and if both those products are required to activate a promoter in front of that actual desired output, then only when both are triggered (within some time period dependent on the lifetime of the intermediate proteins) will any output be produced. Effectively, given that stochastic transcription is somewhat rare, the likelihood of producing output without induction decreases multiplicatively; whereas before any activation of a promoter would cause output, now only when both fire together will output be seen.

When we look at the case where the promoters are actively induced, however, there should be some stable level of both AND gate proteins in the cell at all times and output will be constant depending on that level. Thus the dose response curve and activity of the system at upper ranges can be conserved, while eliminating low-level expression occurring as a result of stochastic firing or noise. The described device should operate as a high-pass filter, but requires the following conditions be met:

1. Zero leakiness of the output promoter, the one triggered by the combined AND gate products;

2. AND gate products with a shorter lifetime than the median time between stochastic bursts;

3. Input promoters with a consistent activity when activated.

A review of the literature shows a number of AND gate systems have been built in a biological context (Anderson 2007; Ramalingam 2009; Sayut 2009). Of particular interest was one created by Wang et al. (2011). Along with meeting all the requirements outlined above, it had the added benefits of extensive characterisation, and convenience; the work was done at Imperial College, and all constructs and component parts were readily available.

Wang's AND gate (figure 21) comprises a sigma-54 promoter (P_{Hrp}) and the genes required to produce two proteins (HrpR and HrpS). Triggering transcription of those genes simultaneously allows P_{Hrp} to transcribe its target. Sigma-54 promoters are linked to the stress response in *E. coli*, and sigma-54, RNA polymerase and DNA very rarely form open complex spontaneously as compared to sigma-70 promoters (Buck 2006), meaning that the system displays minimal leakiness. Importantly, the Hrp system is native to *Pseudomonas syringae*, and does not interact with *E. coli* systems: when transported into the preferred host, its action is insulated from the chassis' background.

The characterisation of this system included transfer functions using two inducible inputs and a fluorescent output, and demonstrated extremely tight control of leakiness and specificity of operation (Wang 2011). The orthogonality provided by its use in an *E. coli* chassis was also attractive, as it matched the intended host strain. By replacing the input promoters – that is, those that serve to transcribe HrpR and HrpS – with the

P_{Las} promoter, it should be possible to exert tight control over the system and allow trial of the AND-gate-as-filter theory.

With these considerations in mind, the circuit was designed as follows (see figure 22).

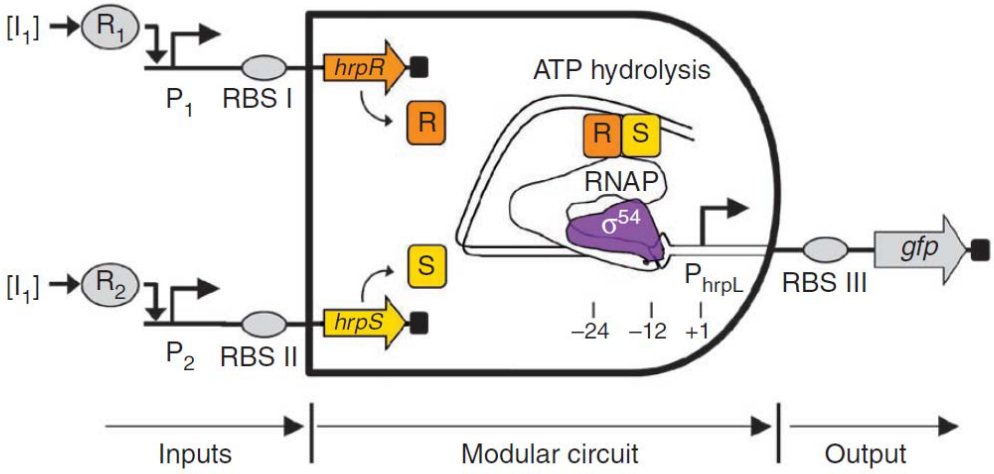


Figure 21: The HrpR/S-based AND gate

Adapted from Wang 2011, this genetic circuit diagram shows that any two inducible promoters (P_1 & P_2) may be linked to the production of HrpR and HrpS; upon simultaneous transcription of both, P_{hrp} may be activated to produce an output.

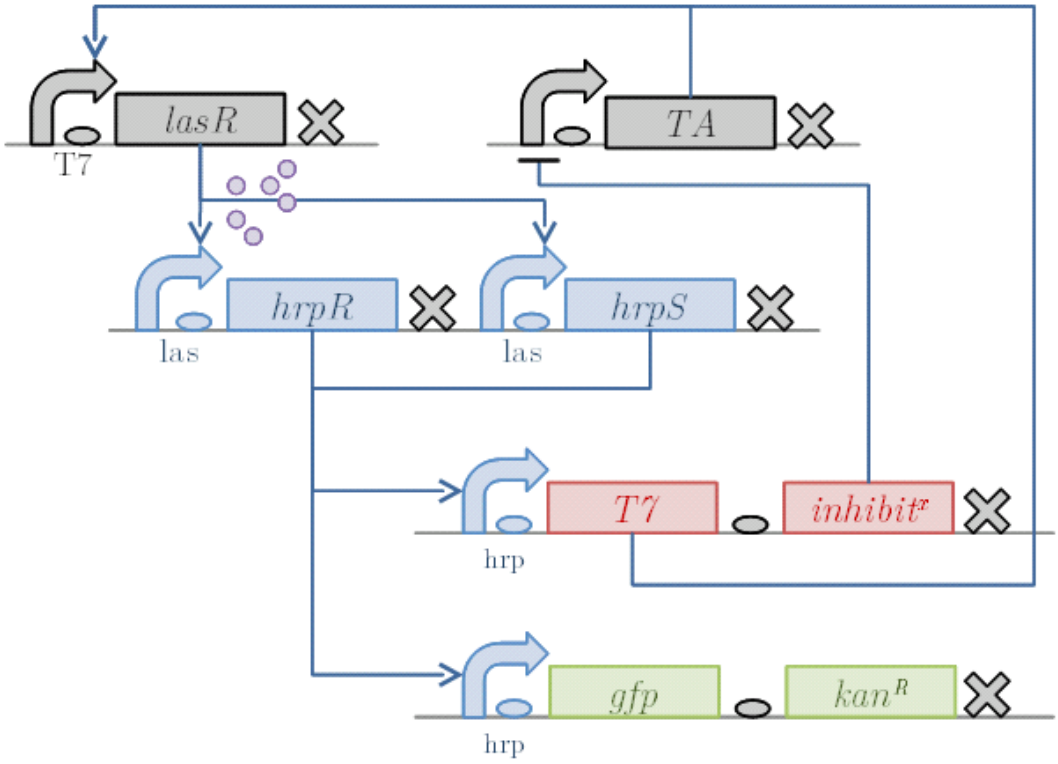


Figure 22: Schematic of AND gate and control circuitry

Parts in grey represent the mutator/target system. Those in blue are the AND gate and lasR/P_{Las} circuit, with red being the feedback loop to prevent overmutation and green the reporter genes to enable screening/selection.

5.3. Results

5.3.1. Construction

Construction of the target and control circuitry was performed in parallel. This allowed testing to begin on each part individually, in the order they were completed, before combinatorial testing began.

The P_{Las} testing construct was designed and constructed both with and without the AND gate described in the introduction to this chapter, and was inserted onto a high-copy backbone. The LasR target was inserted into pAH153 for integration. The mutator protein was transferred to and co-transformed on a low-copy backbone (pC5). The disparity in copy number between the mutator and the control circuitry should serve to allow tighter control of the mutator in the final system – in a mirror of the original construction logic outlined in section 2.4.2 which placed the mutator on the high copy backbone with the target on a low one, to ensure saturation of transcription. Since the testing constructs do not include the full circuitry, the high copy merely provides a more obvious output from successful expression.

All assembly was carried out using Gibson (7.2.6) or CPEC (7.2.7) cloning. Each of the four resulting plasmids was purified and sequence-verified (7.3.1) and are shown in figure 23.

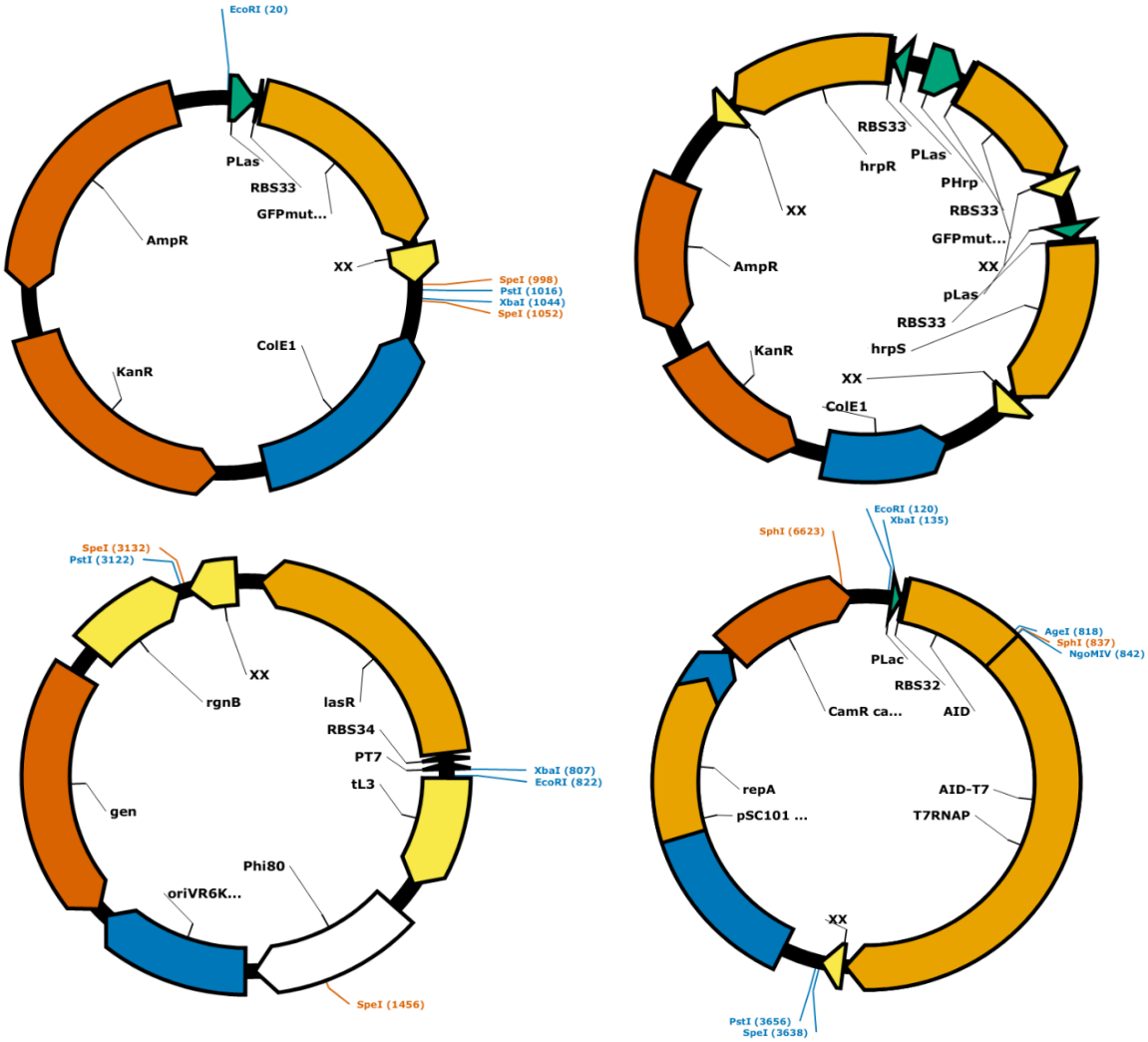


Figure 23: The P_{Las} diversification and reporting devices, with and without AND gate

Top, left: pA3:P_{Las}-GFP, containing GFP under P_{Las} control with the weak RBS33. Top, right: pA3:P_{Las}-AND-GFP, containing the components of the AND gate (hrpR/S) under P_{Las} control and GFP under P_{Hrp}. All modules use the weak RBS33.

Bottom, left: pCRIM:LasR, containing LasR under PT7 on the pAH153 plasmid for integration into the genome. Bottom, right: pC5:TA, the fusion protein on a low-copy backbone.

As the target, LasR was integrated into the GM31 genome following the standard CRIM procedure (7.2.12) and the integrated cassette was amplified and sequenced to ensure sequence veracity. GFP was also integrated into GM31, for use as a positive control in later stages. Upon confirmation of successful and correct integration, these strains was termed GM31-LasR and GM31-GFP.

5.3.2. Testing

Having been made electrocompetent (7.4.3), GM31-LasR and GM31-GFP were transformed as follows:

GM31-LasR

1. pUC19 (negative control)
2. pA3:P_{Las-}-GFP & pC5:T
3. pA3:P_{Hrp}-AND-GFP & pC5:T

GM31-GFP

1. pC5:T (positive control)

These permutations allowed comparison of both LasR-controlled GFP expression – both as a raw input and mediated by the Hrp-based AND gate – to benchmark constitutive or zero expression of the GFP molecule.

This was performed on a fluorescence microplate reader with later tests, such as examining leakiness of straight-through and AND-mediated systems which demanded greater granularity, to be performed on a flow cytometer.

The first round of experiments, cultures over 24 hours with 12-hour sampling times post-induction, yielded unexpected results (figures 24 & 26). They showed no discernable difference in output (as determined by the measured fluorescence, corrected for background and normalised for cell density by division of the corrected fluorescence value by the corrected OD_{600} value) between those cultures induced or uninduced by AHL. Indeed, no output at all was detected, with the exception of the positive control (GM31-GFP(pC5:T)).

To rule out growth differences between constructs, the growth curves from the above experiments were examined in greater detail. These revealed no significant differences between the different constructs (see figure 25). It was therefore considered that the most likely source of the problem was the AHL-induced LasR induction.

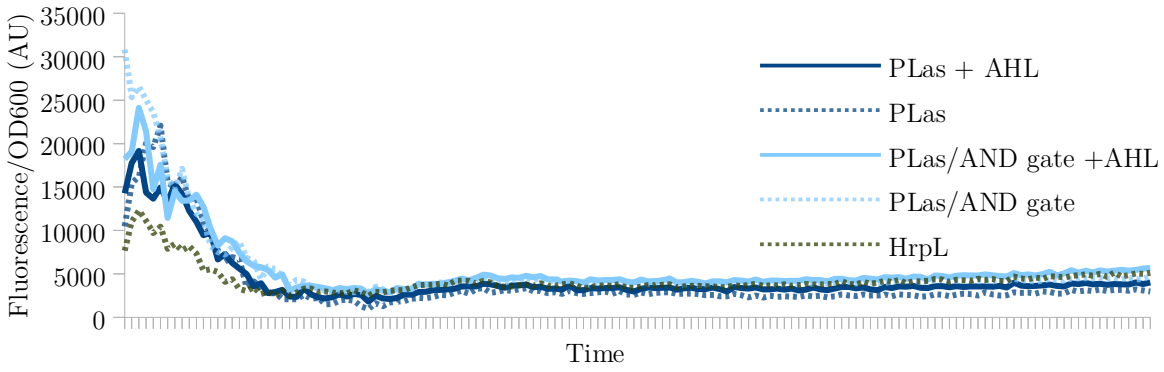


Figure 24: Induction of the LasR/P_{Las} system by AHL failed

Time runs from 0-12 h post-induction with 1 μM AHL for all figures on this page.

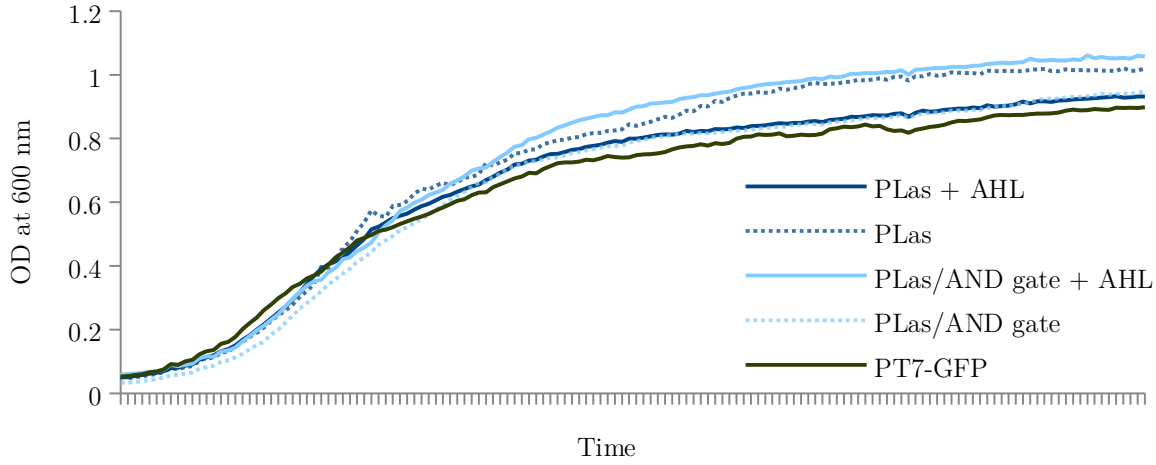


Figure 25: Growth curve for LasR testing constructs

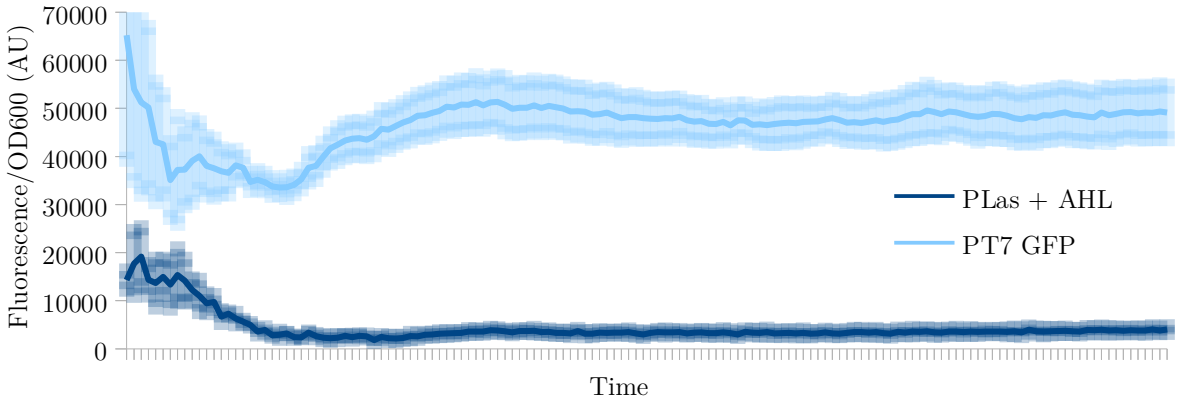


Figure 26: Fluorescence detection is validated with a positive control

Standard error of triplicate repeats is shown, demonstrating consistency between repeats.

Since LasR is chromosomally integrated there are fewer copies than on previously studied plasmid based systems (Chappell 2013). For comparison we therefore included a control plasmid that expresses LasR from the constitutive BBa_J23101 promoter (pA3-lasR-p_{Las}-GFP). This combination produced the expected result, shown in figure 27, demonstrating that the induction procedure was not at fault.

To assess whether the level of T7 RNAP in the cell was too low to generate enough flux through the circuit, GM31-GFP(pC5:T) was tested with a range of IPTG concentrations. The results (figure 28) demonstrate that the level of T7 RNAP produced merely by leakiness from the lac promoter preceding it is enough to cause full transcription of GFP behind the single P_{T7} in the cell – at all concentrations of inducer, including uninduced, there is an equal level of fluorescence output.

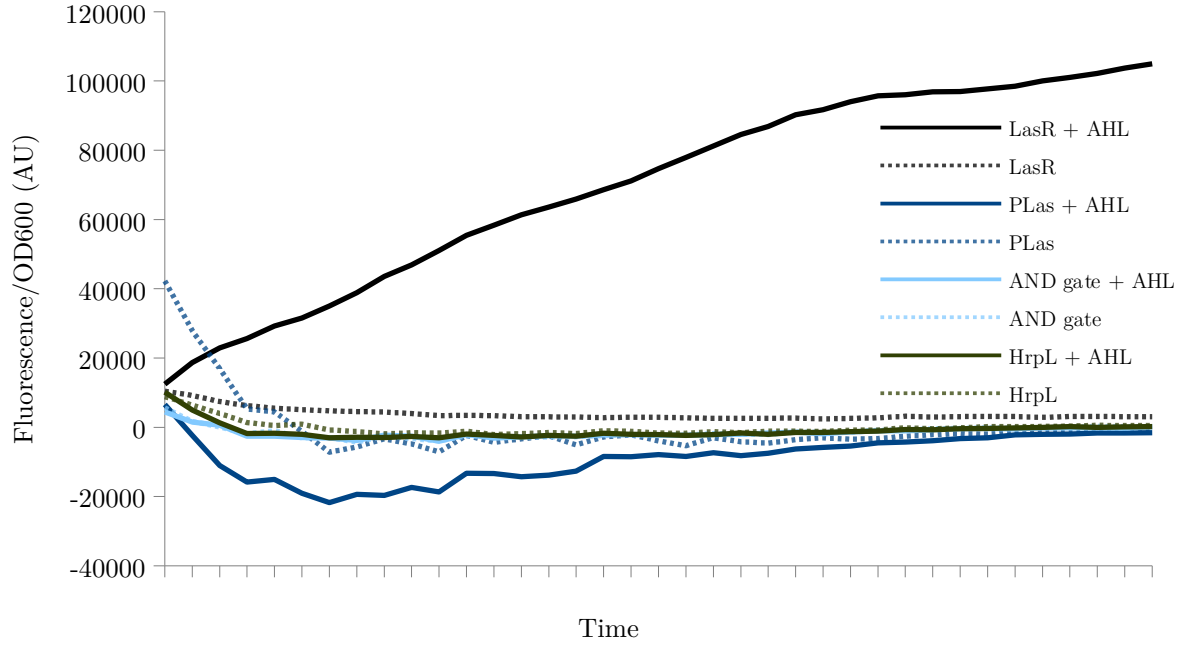


Figure 27: The original LasR system responds to induction with 1 μM AHL

Here, LasR is a positive control plasmid that constitutively expresses LasR and has GFP under a P_{Las} promoter. pA3:PLas-GFP and pA3:PLas-AND-GFP response to AHL induction are also shown, alongside HrpL. Time runs from 0-3 h post-induction.

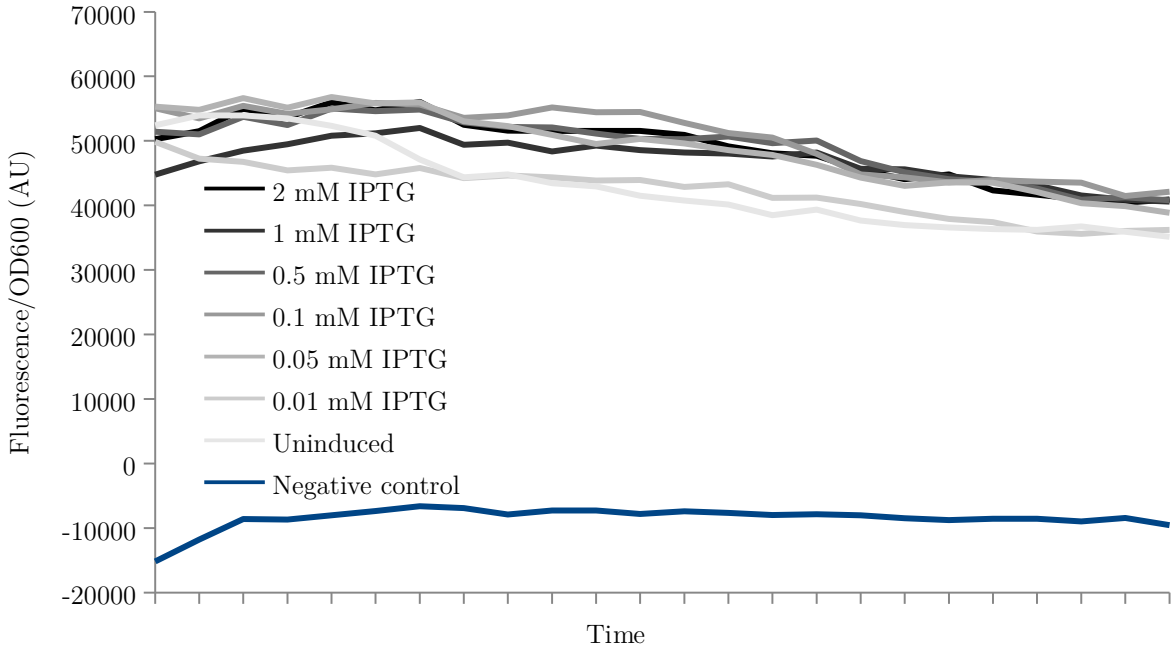


Figure 28: P_{Lac}-T7RNAP induced with IPTG shows strong expression

Time runs from 0-2 h post-induction.

To assess whether the level of LasR in the cell could be the cause, GM31-LasR(pA3:P_{Las}-GFP) & pC5:T) and GM31-LasR(pA3:lasR-p_{Las}-GFP) were both cultured, induced, then lysed and examined using SDS-PAGE. This gel is annotated and presented as figure 29. The control experiment shows a large band at 26.91 kDa (corresponding to GFP) in the induced sample, but no band in the uninduced sample. In the case of GM31-LasR(pA3:P_{Las}-GFP & pC5:T), a band is seen at 26.6 kDa in both uninduced and induced samples – this corresponds to overproduction of LasR by the cell, but no successful 'conversion' to GFP output upon induction. This ruled out LasR production levels being a limiting factor, and strongly indicated that the problem lay in the output circuitry.

Having thus eliminated a number of the most obvious issues which could cause failure of the system in the manner described, only a narrow field of choices remain – the most likely candidate for further scrutiny was the weak RBS (RBS33) preceding the outputs and control genes on pA3:P_{Las}-GFP/P_{Hrp}-GFP. These were marked for removal and replacement via CPEC, with primers being designed accordingly.

Upon successful replacement of the RBS with RBS34, the plasmids were re-transformed and GM31-LasR(pA3:P_{Las}-34-GFP & pC5:T7) was tested using flow cytometry. As seen in figure 30, these data showed the expected pattern, with the induced and uninduced sample peaks falling between the negative and positive controls. The leakiness

suggested by the presence of a peak in the uninduced samples is therefore the next point of interest to be examined; it is this leakiness that it is hoped will be ameliorated by the use of the AND gate.

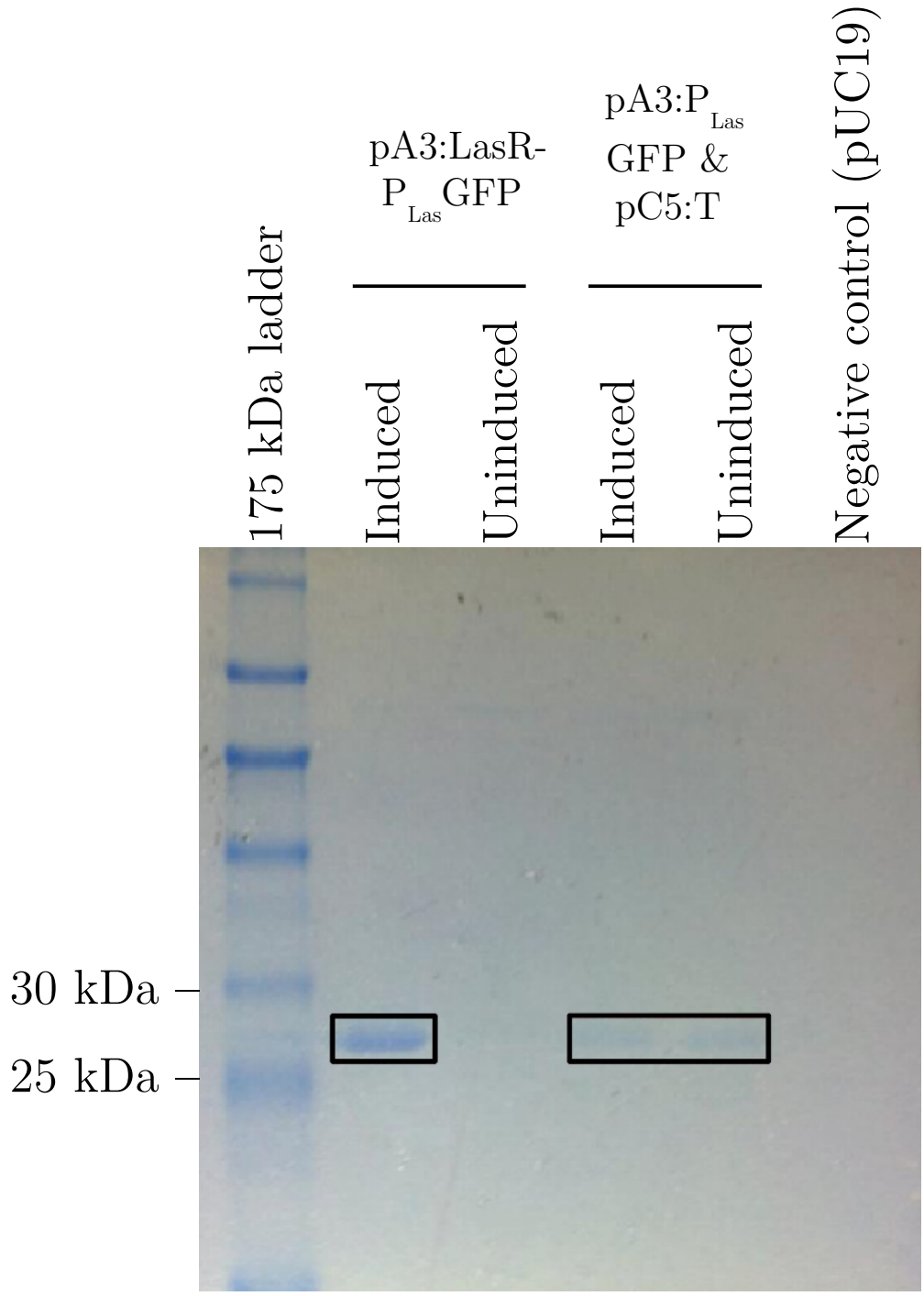


Figure 29: SDS-PAGE gel confirming expression of LasR

LasR shows a band at 26.6 kDa. GFP shows at 26.91 kDa – it is present in the control sample.

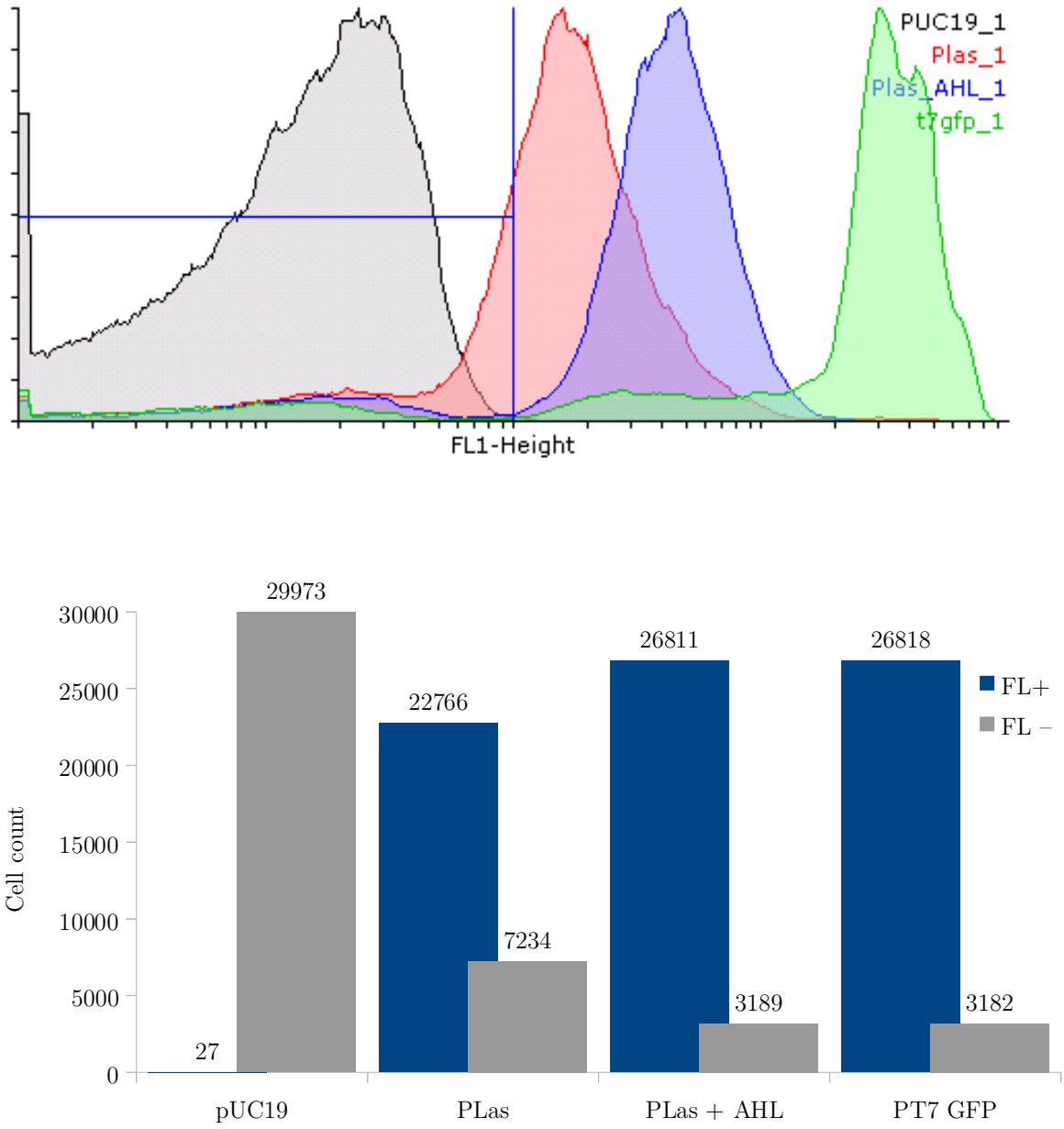


Figure 30: Flow cytometry data showing function of the LasR circuit, and response to induction

Top: negative control (black), GFP behind P_{Las} uninduced (red) and induced (blue) and GFP behind T7 RNAP positive (green) histograms are shown. The plot is divided into regions FL- and FL+.

Bottom: Cell count data showing the distribution of cells between FL- and FL+.

6. Conclusion

The primary aim of the project, to produce a system for the targeted directed evolution of proteins, has been achieved. Experiments involving the mutation of a single-copy target, integrated onto the genome of an *E. coli* chassis, have shown that the fusion of T7 RNA polymerase and AID brings to bear a significant targeted mutation effect in line with the objective of the device. Examination of the mutational landscape demonstrates an even spread of mutation along the target sequence which is far above background rate.

Significant progress has also been made toward the secondary aims, namely assessment of both the exact level of mutation and the degree to which mutation is targeted – loss of function assays have determined that masking may be ignored in any future gain-of function assay designed to calculate an accurate mutation rate for the device. The targets for such an assay have been produced. Integration of those targets and repetition of the single-copy analysis protocol, designed and optimised in this work, will allow such an accurate determination to be made.

The groundwork has been laid for diversification of LasR as a proof of concept, with technical issues of circuit design and development successfully solved. Data produced after the completion of the project, but using constructs and methods created and

optimised in this work demonstrated that the novel adaptation of an AND gate as a bandpass filter successfully reduces leaky expression in the designed control circuitry (figure 31). This will allow transformation of a self-contained mutation and screening device as designed here.

The focus of the project moving forward rests on augmentation of the system. Expressing a uracil DNA glycosylase inhibitor protein inducibly with the fusion in a non repair-deficient strain, and repressing it when the fusion is disabled by the control circuitry, could serve to reduce post-success mutation events while still offering the advantage of preserving mutations introduced by AID in the diversification stages.

Addition of a second AID molecule to the fusion, presenting greater opportunity for dimerisation to occur, could serve to increase the mutation rate of the system and is an avenue that demands exploration.

Tertiary aims, including the retention of compatibility of construction with Synthetic Biology techniques and the improvement of efficiency of work-flow in the Synthetic Biology labs at Imperial College were also achieved – the procurement of a liquid handler, and adaptation of the automation protocols developed during this project have led to streamlining of a number of tasks at the Centre for Synthetic Biology and Innovation, with high-throughput experimentation now routine.

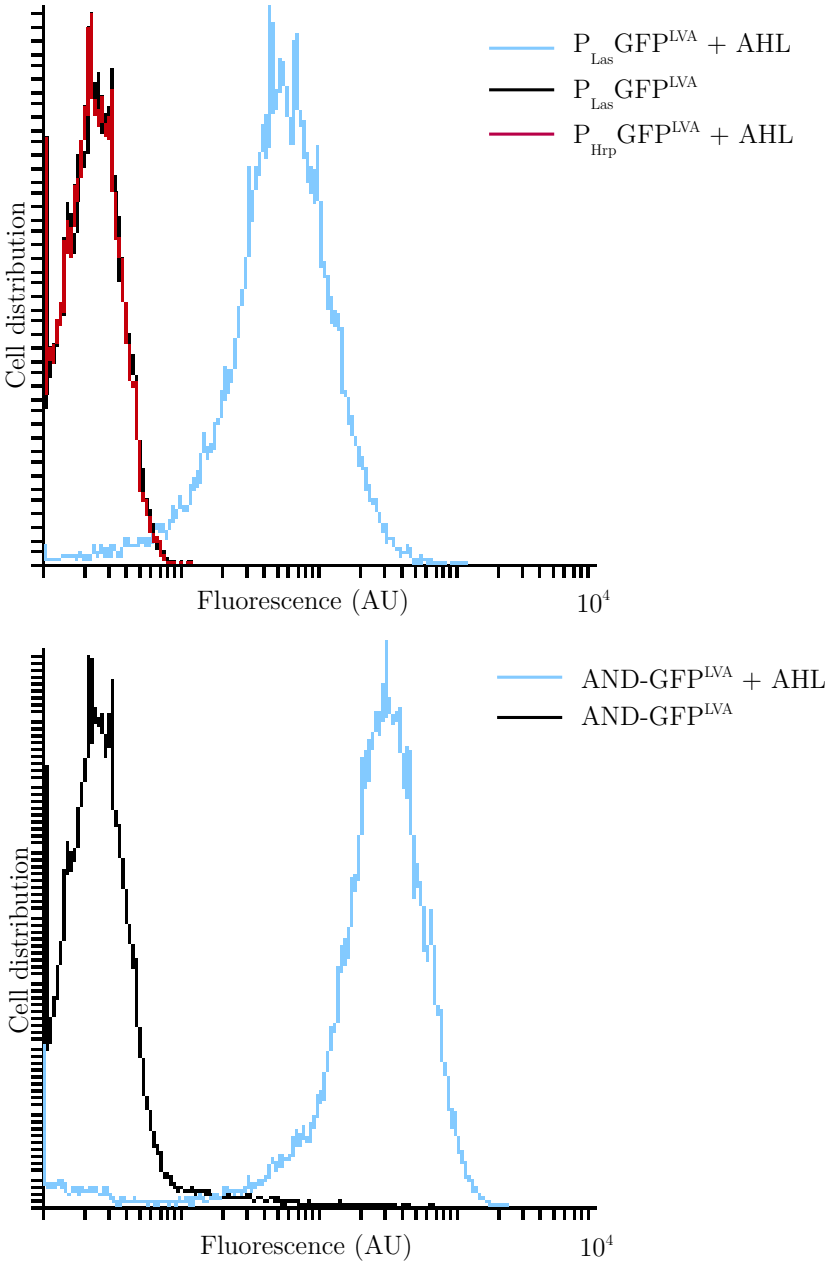


Figure 31: Flow cytometry showing the LasR circuit and AND gate circuit responding to AHL induction

Adapted from Storch 2013 (unpublished data). Response of the AND gate post-RBS replacement demonstrates functionality. Note GFP has an LVA degradation tag added, in an attempt to reduce build-up of leaky expression.

7. Methods

7.1. Tables

7.1.1. Strains

Strain	Genotype	Purpose	Reference
E. coli GM31	*F- Thr-1*, *araC14*, *leuB6*(Am), *fhuA31*, *lacY1*, *tsx-78*, *glnV44*(AS), *galK2*(Oc), *galT22*, λ⁻, *dcm-6*, *hisG4*(Oc), *rpsL136*(strR), *xylA5*, *mtl-1*, *thiE1*	Plasmid storage and isolation	Marinus 1973
E. coli BH156	*GM31 ung-1 tyrA*::*Tn10*	BER pathway knockout	Lutsenko 1999
E. coli BH156-GFP	*GM31 ung-1 tyrA*::*Tn10att*$_{\phi 80}$::*[gfp]*	Single-copy target analysis (loss of function)	This work
E. coli BH156-XFP	*GM31 ung-1 tyrA::Tn10 att*$_{\phi 80}$::*[xfp]*	Single-copy target analysis (reversion to function)	This work
E. coli GM31-LasR	*GM31 att*$_{\phi 80}$::*[lasR]*	Proof of concept diversification	This work
E. coli GM31-GFP	*GM31 att*$_{\phi 80}$::*[gfp]*	Proof of concept diversification	This work
E. coli BL21(DE3)	*F- ompT hsdS*$_B$ *(r*$_B$*-m*$_B$*-) gal dcm lon* λ*(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Native T7 RNA polymerase	NEB catalog #C2527
E. coli TransforMax EC100D pir+	F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL (StrR) nupG pir+(DHFR)	Allows propagation of CRIM plasmids	Epicentre Biotech-nologies

Table 3: Key strains used in this work

7.1.2. Plasmids

Name	Construct	Description	Reference
pCD		pACYC-Duet backbone; chloramphenicol resistance, p15A ori, 10-12 copies per cell.	Registry of Standard Biological Parts
pC5		pSB4C5 backbone; chloramphenicol resistance, pSC101 ori, ~5 copies per cell.	Registry of Standard Biological Parts
pA3		pSB1AK3 backbone; Amp/Kan resistance, pMB1 ori, 100-300 copies per cell.	Novagen
pA3:TA	P_{Lac}-RBS32-AID-T7RNAP-XX	AID-T7RNAP fusion	This work
pA3:T	P_{Lac}-RBS32-T7RNAP-XX	T7RNAP	This work
pA3:T+A	P_{Lac}-RBS32-AID-XX-P_{Lac}-RBS32-T7RNAP-XX	T7RNAP, AID	This work
pAH153 & pAH123		CRIM integration & helper plasmid pair	Haldimann 2001
pCD:PT7-GFP	P_{T7}-RBS34-GFP-XX	GFP under P_{T7}	This work
pC5:PT7-GFP	P_{T7}-RBS34-GFP-XX	GFP under P_{T7}	This work
pCD:PT7-XFP	P_{T7}-RBS34-XFP-XX	XFP under P_{T7}	This work
pC5:PT7-XFP	P_{T7}-RBS34-XFP-XX	XFP under P_{T7}	This work
pCRIM:GFP	P_{T7}-RBS34-GFP-XX	GFP under P_{T7}	This work
pA3:PLas-GFP	P_{Las}-RBS33-GFP-XX	GFP under P_{Las}	This work
pA3:PLas-AND-GFP	P_{Las}-RBS33-HrpR-XX-P_{Las}-RBS33-hrpS-XX-P_{Hrp}-RBS33-GFP-XX	GFP under AND gate control, with AND gate proteins under control of $P_{Las.}$	This work
pCRIM:LasR	P_{T7}-RBS34-LasR-XX	LasR under P_{T7}	This work
pC5:T	P_{Lac}-RBS32-T7RNAP-XX	T7 RNAP under P_{Lac}	This work

Table 4: Key plasmids used in this work

7.1.3. Primers

Name	Sequence	Purpose
VF2	TGCCACCTGACGTCTAAGAA	Sequencing constructs in BioBrick vectors
VR	ATTACCGCCTTTGAGTGAGC	
T7F	atgAACACGATTAACATCGCTAAGAAC	Extracting T7 in pSB1AK3 and adding a start codon
T7R	gttaatcgtgttCATCTAGTACTTTCCTGTGTGAC	
AIDF	TAATAATACTAGAGCCAGGCATC	Extracting AID in pSB1AK3 and adding a stop codon
AIDR	gctctagtattattaCTGACCGGTTAAACCCAG	
KanS_GF	gaggagaaatactagATGATTGAACAAGATGGATTG	Extract and add overhangs to *kan*^{S} for InFusion
KanS_GR	gatgcctggctctagtaTCAGAAGAACTCGTCAAG	
KanS_VF	TACTAGAGCCAGGCATC	Remove GFP target from vector, linearise for *kan*^{S} InFusion
KanS_VR	CTAGTATTTCTCCTCTTTAATCTC	
GFPtoXFP_SDM_F	CTTGTCACTACTTTCGGTCATGGTGTTCAATGCTTTGC	Site-directed mutagenesis (GFP-XFP)
GFPtoXFP_SDM_R	GCAAAGCATTGAACACCATGACCGAAAGTAGTGACAAG	
CRIM_Gene_F	gagccctactccatacgGTTGATCGGGCACGT	Extract GFP/XFP/kan^S targets for CPEC into pAH153
CRIM_Gene_R	ccgccttgaacgaattGGTCTGACAGCTCTAGC	
CRIM_153_F	gctagagctgtcagaccAATTCGTTCAAGGCGG	Linearise pAH153 for CPEC of GFP/XFP/kan^S targets
CRIM_153_R	acgtgcccgatcaacCGTATGGAGTAGGGCTC	
Phi80_P1	CTGCTTGTGGTGGTGAAT	Phi80-specific integration check
CRIM_P2	ACTTAACGGCTGACATGG	CRIM integration check
CRIM_P3	ACGAGTATCGAGATGGCA	

CRIM_T3	ACGGTGGTATATCCCGTG	Modified CRIM_P3 for use with pAH153 variants made in this work
Phi80_P4	TAAGGCAAGACGATCAGG	Phi80-specific integration check
SB_vector_CPECF	agcggtatcaGCTCACTCAAAGGCGGTAAT	Linearisation of SB vectors for CPEC/Gibson
SB_vector_CPECR	catgataataatggtTTCTTAGACGTCAGGTGGCA	
LasR_F	gcttctagagtaatacgactcactataTACTAGAGAAAGAGGAGAAATACTAGATGG	Extract RBS34-LasR for putting in pCRIM
LasR_R	tcatccatgccatgtgtaatcccaGCCTGGCTCTAGTATTATTAGAGAG	
pCRIM_F	cctctctaataatactagagccaggcTGGGATTACACATGGCATGGATGA	Linearise pCRIM for CPEC/Gibson
pCRIM_R	catctagtatttctcctctttctctagtaTATAGTGAGTCGTATTACTCTAGAAGC	
Rsh_GFP_in_153	TTGCATGCCTGCAGC	Sequencing of target in pAH153/amplification of genomic targets
Fsh_GFP_in_153	GCCTCGCTTTGTAACGG	
Flo_GFP_in_153	GCCTCGCTTTGTAACGG	As above but further from the target
Rlo_GFP_in_153	ACAGGCTTATGTCAACTGG	

Table 5: Key primers used in this work

Sequences in lower case are overhangs which do not initially bind to the target DNA sequence; sequence in capitals corresponds to the DNA-binding sequence.

7.2. DNA construction and manipulation

7.2.1. Plasmid isolation

For DNA construction purposes, plasmid DNA was purified from 3 ml overnight cell cultures using the E.Z.N.A. plasmid mini kit (Omega Bio-tek) or QIAprep mini kit (QIAGEN), following the manufacturer's protocol. Automated methods are described in 7.5.2.

The QIAprep protocol was adapted to avoid the use of spin columns when sample number was high. The modified protocol was as follows:

Cells were pelleted in a microcentrifuge tube in a bench-top centrifuge, and resuspended in 150 µl buffer P1. 150 µl buffer P2 was added, and the tube mixed by inversion. This was repeated with 210 µl buffer N3, and the resulting mixture was spun down in a benchtop microcentrifuge for 10 minutes. 450 µl supernatant was mixed with 900 µl ethanol by inversion, before a second 10 minute centrifuge step. Supernatant was discarded and the pellet washed by vortexing in 0.5 ml 70% ethanol. A final spin step of 1 minute was carried out, and the supernatant discarded. The pellet was dried and resuspended in 30-50 µl ddH_20.

7.2.2. PCR

Standard PCR was carried out in reactions containing 1 U Pfu Turbo Cx (Agilent Genomics), 0.2 mM of each dNTP, 1× PC2 buffer, 0.4 µM each primer, 50 ng template DNA, and ddH_2O up to a final volume of 25 µl.

PC2 buffer (Barnes 1994) was composed of 20 mM Tris-HCl (pH 8.5), 150 mg/ml bovine serum albumin, 16 mM $(NH_4)_2SO_4$, and 3.5 mM $MgCl_2$.

The PCR program used had the following steps: (i) initial denaturation for 2 min at 95°C; (ii) 30 cycles consisting of 30 s denaturing at 95°C, 30 s annealing at 55-65°C, 1 min/kb extension at 72°C; (iii) final extension for 2 min at 72°C. An Applied Biosystems 2720 thermal cycler was used.

For downstream applications requiring high-fidelity amplification, such as Gibson or CPEC DNA assembly, Phusion High-Fidelity enzyme (NEB) was used instead of Pfu Turbo Cx. The protocol described above was followed with two modifications: 1× Phusion HF Buffer was used instead of PC2 Buffer, and extension time was calculated as 15 s/kb plus an additional 15 s.

7.2.3. Colony PCR

Colony PCR was carried out using the same protocol as described in section 7.2.2, with the following modifications: (i) single colonies were picked and re-suspended in 20 µl ddH_2O, then 1 µl was used as DNA template; (ii) the initial denaturation step was increased to 5 min at 95°C.

7.2.4. Site-directed mutagenesis

Site-directed mutagenesis was conducted following the manufacturer's instructions in the Stratagene QuikChange Site-directed Mutagenesis kit (now supplied by Agilent). The thermal cycling program used was programmed as follows: initial 30 s at 95°C, then 16 cycles of 30 s at 95°C, 1 min at 48-57°C, and 1 min/kb at 72°C. The PCR product was incubated with 10 U DpnI (NEB) for 1 h at 37°C to digest template DNA. It was then transformed into *E. coli.*

7.2.5. In-Fusion assembly

The In-Fusion HD Cloning Plus kit (Clontech) was used for scarless assembly of DNA fragments , according to the manufacturer's instructions.

7.2.6. Gibson isothermal assembly

Gibson isothermal assembly protocol was applied as described by Gibson et al (2009) following their 'one-step isothermal DNA assembly protocol'. Briefly, DNA fragments were amplified by PCR with primers with 5' overhangs to generate approximately 40 bp overlap sequences. 100 ng of the largest DNA molecule and equimolar mass of smaller molecules were incubated with the reagent-enzyme mix, prepared as described (Gibson 2009), at 50°C for 1 h. 1 µl of the assembly reaction was used to transform electro-competent *E. coli* cells.

7.2.7. CPEC

Circular Polymerase Extension Cloning (CPEC) was performed as described by Quan and Tian (2009). The DNA molecules with overlapping end sequences were generated using the same method as in Gibson assembly (section 7.2.6). Electro-competent *E. coli* cells were transformed with 0.5 µl of the CPEC.

7.2.8. BioBrick cloning

The BioBrick assembly standard is a standardised restriction digest cloning strategy suitable for assembling parts from the Parts Registry (parts.igem.org). The protocol requires double-enzyme digestion of DNA parts and vector, followed by dephosphorylation of digested vector, and ligation of parts with vector (Shetty 2008).

7.2.9. Digestion

All restriction enzymes used were supplied by NEB. Generally, 10 U enzyme was used per 500 ng DNA, and incubated for 1 h at 37°C in the appropriate 1× buffer (chosen based on manufacturer's instructions).

7.2.10. Dephosphorylation

Linearised vectors were treated with Antarctic Phosphatase (NEB) to reduce self-ligation, following the manufacturer's protocol.

7.2.11. Ligation

DNA insert and vector were ligated using T4 DNA ligase (NEB). A 3:1 equimolar ratio of insert to vector was used, with 50 ng vector. DNA was added to a 20 µl reaction mix, with 400 U T4 DNA ligase and 1× T4 DNA ligase buffer, and incubated for 2 h at room temperature. Ligase was inactivated by incubation at 65°C for 10 minutes.

7.2.12. CRIM integration

Targets were integrated in to the *E. coli* genome using the Conditional-Replication, Integration, and Modular (CRIM) plasmid system. The protocol was followed as described in Haldimann et al (2001), with the pir+ cell strain used for plasmid propagation described in section 7.1.1.

Confirmation of successful integration was carried out with the 4-way PCR described in Haldimann 2001, but with a slightly modified 'P3' primer (CRIM_T3 in table 5).

7.3. Quantification

7.3.1. DNA concentration & sequence verification

The concentration of DNA was measured using the NanoDrop 1000 spectrophotometer (Thermo Scientific) to determine absorbance at 260 nm. Sequence verification of purified DNA was performed by MWG Eurofins Operon.

7.3.2. Gel electrophoresis/extraction

Agarose gels were prepared with 1% w/v agarose and 1× Tris-Borate-EDTA (TBE) buffer. DNA samples were loaded with 1× SYBR Green (Life technologies). Gels were run at 120 V, and 1 kB DNA ladder (NEB) was used as a size reference. Bands were visualised using one of two imagers; a LAS-3000 or a FLA-5000 (both Fuji). LAS-3000 employed 460 nm LED light source and Y515Di filter. FLA-5000 employed 47s nm laser light source with Y510 filter.

For isolation of DNA from agarose gels, the QIAQuick Gel Extraction kit (Qiagen) was used according to manufacturer's protocol.

7.3.3. SDS-PAGE

Cells were cultured in LB media to an OD at 600 nm of 0.6. They were then induced as appropriate and cultured overnight. 500 µl of culture was spun down in a bench top centrifuge and resuspended to a volume of 200 µl. This sample was heated to 95C for 0.5 h, then spun down. The supernatant was added to 50 µl of loading buffer (at 5x stock). 30 µl samples were loaded onto the gel.

The gel itself consisted of a 12% resolving gel (5 ml consisting of 1.6 ml ddH_2O; 2 ml 30/0.8% acrylamide/bis; 1.27 ml 1.5 M Tris at pH 8.8; 50 µl 10% SDS; 67 µl 10% ammonium persulfate and 7.1 ml TEMED) and a 5% stacking gel (5 ml consisting of 3.4 ml ddH_2O; 0.83 ml 30%/0.8 acrylamide/bis; 0.63 ml 1 M Tris at pH 6.8; 50 µl 10% SDS; 50 µl 10% ammonium persulfate and 10 µl TEMED), and was run for 15 minutes at 80 V and 80 minutes at 160 V. Staining and destaining used a solution of 40% methanol, 7% acetic acid, with 0.025% Coomassie Brilliant Blue added for the staining step which consisted of 2 minutes heating at low power in the microwave followed by 20 minutes shaking incubation in the dark.

7.4. Cultures & propagation

7.4.1. Cultures

Media used was either LB (tryptone at 10 g/l; yeast extract at 5 g/l; NaCl at 5 g/l in ddH_2O) or autoinducing media from Formedium (made up according to manufacturer's directions, 37.5 g/l). For agar plates, 7.5 g/l agar was added. Antibiotics were used as appropriate, with final concentrations being as follows: Amp – 100 µg/ml; Cam – 34 µg/ml; Kan – 25 µg/ml; Gent – 20 µg/ml; Rif – 50 µg/ml.

For general purposes, cells were diluted 1:1000 into, or picked from colonies using a sterile pipette tip which was ejected into, a glass test tube containing 3 ml of either LB or auto-inducing media as required. This media was supplemented with appropriate antibiotic from 1,000× stock, and the tube capped and incubated at 37°C with shaking at 222 rpm unless otherwise stated.

For the each culture of the multi-copy and single-copy assessment runs, 1 ml pre-warmed auto-inducing media supplemented with the appropriate antibiotic was inoculated from an individual colony picked from a transformation plate. These were incubated at 37°C for 2 hours, with shaking at 222 rpm. At this time, 10 µl of culture was diluted 1:100 and 1:10,000 in fresh media, 100 µl of which was plated on AI plates

with the appropriate antibiotic. The remainder of the culture was spun down in a bench-top centrifuge and stored at 4°C overnight, while the plates were stored at 37°C.

The dilution plates were scored the following morning to assess the number of colony-forming units (CFU) per ml of culture. This value was used to determine the volume of stored culture required to inoculate a 13 ml 'starter culture', consisting of LB with appropriate antibiotic with 20% glycerol, to ensure that the initial CFU of cultures in the runs was low enough (100-150) to conform to the assumptions listed in table 1.

Starter cultures were kept at -20°C, and used as seed stocks to inoculate each run.

7.4.2. Chemically-competent cells/chemical transformation

Chemically-competent cells were prepared based on a modified protocol from Sambrook and Russell, third edition (2001). A starter culture was prepared by inoculating a single colony into 5 ml of LB medium and incubating for 16 h at 37°C with shaking at 222 rpm. 100 ml LB medium was inoculated with 1 ml of the starter culture, and then grown at 37°C until OD_{600} reached 0.6-0.8. The culture was pre-chilled for 15 min on ice before centrifugation for 5 min at 4,000 RCF and 4°C. The supernatant was removed and the pellet re-suspended in 10 ml ice-cold 0.1 M $CaCl_2$ and incubated on ice for 20 min. The centrifuge step was then repeated, the supernatant discarded and cells were

resuspended in 5 ml ice-cold 0.1 M $CaCl_2$ with 15% glycerol. The cells were stored in aliquots of 210 µl at -80°C.

For transformation, 0.5-1 µl plasmid DNA and 100 µl cells were placed in a pre-chilled microcentrifuge tube on ice, and incubated for 30 minutes. The microcentrifuge tube was transferred to a waterbath or heatblock at 42°C for 2 minutes, then returned to the ice. 900 µl pre-warmed LB medium was added to the tube, which was incubated for 30 minutes at 37°C with 222 rpm shaking. Dilutions of the cells were plated on the appropriate antibiotic-supplemented media and incubated overnight at 37°C.

A variant of the transformation protocol for use on an automated platform is described in 7.5.3.

7.4.3. Electro-competent cells/electroporation

Electrocompetent cells were prepared based on a modified protocol from Sambrook and Russell, third edition (2001). A starter culture was prepared by inoculating a single colony into 50 ml of LB medium and incubating for 16 h at 37°C with 222 rpm. 1 l LB medium was inoculated to an OD_{600} of 0.1 using the starter culture, and then grown at 37°C until OD_{600} reached 0.6-0.8. The culture was pre-chilled for 15 min on ice before centrifugation for 15 min at 4,000 RCF and 4°C. The supernatant was removed and the pellet re-suspended in 1 L ice-cold ddH_2O. The centrifuge step was repeated three

times, with subsequent washes in 0.5 L ddH_2O, 20 ml sterile 10% glycerol, then 2.5 ml 10% glycerol. The cells were stored in aliquots of 125 µl at -80°C.

For transformation, 0.5-1 µl DNA and 40 µl cells were placed in a pre-chilled 1 mm electroporation cuvette. The sample was electroporated using a BioRad GenePulser with the conditions 1.8 kV, 200 Ω, and 25 µF. 1 ml pre-warmed LB medium was added to the cuvette, then the mixture was transferred to a microcentrifuge tube and incubated for 1 h at 37°C with 222 rpm shaking. Dilutions of the cells were plated on the appropriate antibiotic-supplemented media and incubated overnight at 37°C.

7.4.4. Glycerol stocks

For long-term storage of bacterial strains, overnight cultures were prepared by inoculation of a single colony in 3 ml LB media supplemented with the appropriate antibiotic. Glycerol was added after 18 h growth to a final concentration of 20% in 1 ml total volume, and the stocks were stored at -80°C.

7.5. Automation

The protocols for the automated versions of several lab procedures which were developed are described in general terms below; the programs themselves may be found

in appendix 9.2. All automation was performed on a TheONYX liquid handler (Aviso GMBH).

7.5.1. Preparation

Culturing was performed as described in section 7.4.1 with inoculation at 150 CFU. 2 ml of each culture was transferred to 96-well plates with well volumes of 2.2 ml for use on the handler, and 1 ml was transferred to a separate plate with well volumes of 1.2 ml for storage. The culture plates were centrifuged to pellet the cells, then placed on the liquid handler.

7.5.2. Plasmid purification

A script was written and optimised which instructed the platform to resuspend the cells, and perform a plasmid purification procedure using the magnetic bead-based kit according to manufacturer's directions. Wash steps were performed in 0.01 KOH for the plasmid purification procedures and 2 M KOH for the transformation steps. These solutions were changed between runs.

The DNA concentration was measured in the inbuilt platereader, allowing failed runs to be identified.

7.5.3. Transformation

40 µl of chemically competent cells prepared as in 7.4.2 were pipetted into a 96-well PCR plate on the cooled position of the handler, kept at 1°C. 0.5 µl DNA was transferred by the liquid handler and incubated for 30 minutes, then the plate moved to the PCR machine at 42°C for 30 seconds before being returned to the cooled position; mimicking the heat-shock protocol described in section 7.4.2. 100 µl pre-warmed LB media was added to the cells and they were returned to the PCR machine for a 2-hour incubation at 37°C to recover.

Post-recovery, the cells were plated for observation and colony counting.

7.5.4. Semi-automated continuous culture

For the longer culture runs examined by flow cytometry, the handler was used to sample cultures at regular intervals over the course of several days, and reinoculate cultures into fresh media.

Cultures were grown, from seed stocks as described in 7.4.1, in 96-well plates incubated at 37°C and agitated at 300 rpm in a heated shaker inbuilt to the robotic platform. Every 8 hours, the liquid handling platform inoculated a pre-warmed fresh plate of media with a sample from each well of the previous plate, and culture from the

previous plate was diluted 1:100 in ddH_2O and saved on a cooled position at 4°C, for later analysis by flow cytometer.

It should be noted here that while the miniprep procedure in particular is built around a kit which has since been discontinued by the manufacturer, it is easily adaptable to use with other, similar kits. Since the time of use, they have been adapted – for miniprep, PCR purification and transformation protocols – and applied to several other liquid handlers within the Centre for Synthetic Biology and Innovation at Imperial College and work with them is ongoing.

7.6. Analysis

7.6.1. Rifampicin reversion assays

Cells were transformed with appropriate plasmids and grown up overnight. Colonies were picked as soon as they could be identified, and used to inoculate 3 ml pre-warmed LB in a glass test-tube. These were incubated at 37°C for 2 hours, then divided and stored at -80°C in 20% glycerol in 1 ml aliquots. For the assays, cultures were inoculated from stocks derived from independent colonies 1:1000 into 5 ml pre-warmed auto-inducing media. These were incubated at 37°C with 222 rpm shaking for 18 hours. OD at 600 nm was assessed for a 10× dilution of each culture, and the cultures were normalised by addition of fresh auto-inducing media to an OD_{600} of 0.5. 2 ml of this

normalised culture was plate on LB-agar plates containing rifampicin and incubated at 37°C for 16 hours before being scored.

7.6.2. Fluorimetry

All samples for analysis by fluorescence plate reader were initially cultured from glycerol stocks of the relevant cell line. 3 µl of stock was taken and used to inoculate 3 ml pre-warmed (37°C) LB media in round-bottomed test tubes. These were cultured overnight at 37°C, with shaking at 222 rpm before dilution 1:100 into 3 ml of pre-warmed fresh LB media. After exponential phase was reached, cultures were diluted to an OD at 600 nm of 0.1 and allowed to grow for 1 hour. 3 repeats of 200 µl of each culture were distributed into a pre-warmed 96-well plate. Where appropriate, inducer was added at this stage. Plates were sealed with clear film and loaded into the plate reader, a BMG POLARstar Omega plate reader (BMG Labtech). Absorbance at 600 nm and fluorescence of the cells (485 nm/510 nm excitation/emission) were measured every 5 minutes. The reader maintained a constant temperature of 37°C for the duration of all assays, with shaking at 222 rpm between reads.

Fluorimetry data was analysed by determining background values via measurement of blank media. This was subtracted from values of both fluorescence and absorbance.

Fluorescence was then normalised with regard to cell density as measured by absorbance at 600 nm.

7.6.3. Flow cytometry

Cell culture was diluted 1:100 in 2 ml ddH_2O. This was loaded onto a Becton-Dickinson FACScan flow cytometer equipped with an Argon laser (excitation wavelength 488 nm), and able to detect green fluorescence (emission wavelength parameter FL1 at 530 nm). Gain was set to the maximum available level, to best differentiate fluorescent and non-fluorescent cells. Data was collected according to manufacturer's instructions, and analysed using either CyfLogic (Cyflogic.com) or Flowing Software (flowingsoftware.com). All gates and regions were defined post-sample collection.

8. Bibliography

Alper, H.; Fischer, C.; Nevoigt, E. & Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(36), 12678-12683. doi: 10.1073/pnas.0504604102

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. *Applied and Environmental Microbiology*, *64*(6), 2240-2246.

Anderson, J.C.; Voigt, C.A. & Arkin, A.A. (2007) Environmental signal integration by a modular AND gate. *Molecular Systems Biology*, *3*, 133. doi: 10.1038/msb4100173.

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. *Journal of Biological Engineering*, *4*(1), 1. doi: 10.1186/1754-1611-4-1.

Angers-Loustau, A.; Rainy, J. & Wartiovaara, K. (2007) PlasmaDNA: a free, cross-platform plasmid manipulation program for molecular biology laboratories. *BMC Molecular Biology*, *8*, 77. doi: 10.1186/1471-2199-8-77.

Arakawa, H. & Buerstedde, J. (2009) Activation-induced cytidine deaminase-mediated hypermutation in the DT40 cell line. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1517), 639-644. doi: 10.1098/rstb.2008.0202.

Arakawa, H.; Kudo, H.; Batrak, V.; Caldwell, R. B.; Rieger, M. A.; Ellwart, J. W.; Buerstedde, J. (2008) Protein evolution by hypermutation and selection in the B cell line DT40. *Nucleic Acids Research*, *36*(1), e1. doi: 10.1093/nar/gkm616.

Asteris, G. & Sarkar, S. (1996) Bayesian procedures for the estimation of mutation rates from fluctuation experiments. *Genetics*, *142*(1), 313-326.

Bainton, N.J.; Stead, P.; Chhabra, S.R.; Bycroft, B.W.; Salmond, G.P.; Stewart, G.S. & Williams, P. (1992) N-(3-oxohexanoyl)-L-homoserine lactone regulates carbapenem antibiotic production in *Erwinia carotovora*. *Journal of Biochemistry, 288*(3), 997-1004.

Balagaddé, F.K.; You, L.; Hansen, C.L.; Arnold, F.H. & Quake, S.R. (2005) Long-Term Monitoring of Bacteria Undergoing Programmed Population Control in a Microchemostat. *Science*, *309*(5731), 137-140. doi: 10.1126/science.1109173.

Barnes, W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proceedings of the National Academy of Sciences, 91*(6), 2216-2220.

Bassler, B.L. (1999). How bacteria talk to each other: regulation of gene expression by quorum sensing. *Current Opinion in Microbiology*, *2*(6), 582-587.

Basu, S.; Gerchman, Y.; Collins, C.H.; Arnold, F.H. & Weiss, R. (2004) A synthetic multicellular system for programmed pattern formation. *Nature*, *434*, 1130-1134. doi: 10.1038/nature03461.

Beletskii, A. & Bhagwat, A. (1996) Transcription-induced mutations: Increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13919-13924

Beletskii, A.; Grigoriev, A.; Joyce, S. & Bhagwat, A. S. (2000) Mutations induced by bacteriophage T7 RNA polymerase and their effects on the composition of the T7 genome. *Journal of Molecular Biology*, *300*(5), 1057-1065. doi: 10.1006/jmbi.2000.3944.

Bornscheuer, U. T.; Altenbuchner, J. & Meyer, H. H. (1998) Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epothilones. *Biotechnology and Bioengineering*, *58*(5), 554-559. doi: 10.1002/(SICI)1097-0290(19980605)58:5<554::AID-BIT12>3.0.CO;2-B.

Bransteitter, R.; Prochnow, C. & Chen, X. S. (2009) The current structural and functional understanding of APOBEC deaminases. *Cellular and Molecular Life Sciences: CMLS*, *66*(19), 3137-3147. doi: 10.1007/s00018-009-0070-y.

Brar, S. S.; Sacho, E. J.; Tessmer, I.; Croteau, D. L.; Erie, D. A. & Diaz, M. (2008) Activation-induced deaminase, AID, is catalytically active as a monomer on single-stranded DNA. *DNA Repair*, *7*(1), 77-87. doi: 10.1016/j.dnarep.2007.08.002.

Buck, M.; Bose, D.; Burrows, P.; Cannon, W.; Joly, N.; Pape, T.; Rappas, M.; Schumacher, J.; Wigneshweraraj, S. & Zhang, X. (2006) A second paradigm for gene activation in bacteria. *Biochemical Society Transactions*, *34*, 1067-1071. doi: 10.1042/BST0341067.

Camps, M.; Naukkarinen, J.; Johnson, B. P. & Loeb, L. A. (2003) Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(17), 9727-9732. doi: 10.1073/pnas.1333928100.

Canton, B.; Labno, A. & Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology*, *26*(7), 787-793. doi: 10.1038/nbt1413.

Chamberlin, M.; McGrath, J. & Waskell, L. (1970) New RNA polymerase from *Escherichia coli* infected with bacteriophage T7. *Nature*, *228*(5268), 227-231.

Chamberlin, M. & Ring, J. (1973) Characterization of T7-specific ribonucleic acid polymerase. 1. General properties of the enzymatic reaction and the template specificity of the enzyme. *The Journal of Biological Chemistry*, *248*(6), 2235-2244.

Chappell, J. E. (2013) Cell-free systems for synthetic biology: development of characterisation tools and biosensors. Unpublished doctoral dissertation, Imperial College London, UK.

Cobb, R. E.; Sun, N. & Zhao, H. (2013) Directed evolution as a powerful synthetic biology tool. *Methods*, *60*, 81-90. doi: 10.1016/j.ymeth.2012.03.009.

Cormack, B.P.; Valdivia, R.H. & Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP) *Gene*, 173(1), 33-38. doi: 10.1016/0378-1119(95)00685-0.

Dickerson, S. K.; Market, E.; Besmer, E. & Papavasiliou, F. N. (2003) AID mediates hypermutation by deaminating single stranded DNA. *The Journal of Experimental Medicine*, *197*(10), 1291-1296. doi: 10.1084/jem.20030481.

Duerkop, B.A.; Varga, J.; Chandler, J.R.; Peterson, S.B.; Herman, J.P.; Churchill, M.E.A.; Parsek, M.R.; Nierman, W.C. & Greenberg, E.P. (2009) Quorum-sensing control of antibiotic synthesis in *Burkholderia thailandesis*. *Molecular Biology of Pathogens, 191* (12), 3909-3918. doi: 10.1128/JB.00200-09.

Elowitz, M. B.; Levine, A. J.; Siggia, E. D. & Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science, 297*, 1183-1186. doi:10.1126/science.1070919

Erickson, D.L, Endersby, R.; Kirkham, A.; Stuber, K.; Vollman, D.D.; Rabin, H.R.; Ian, M. & Storey, D.G. (2002) *Pseudomonas aeruginosa* Quorum-Sensing Systems May Control Virulence Factor Expression in the Lungs of Patients with Cystic Fibrosis. *Infection and Immunity, 70*(4), 1783-1790. doi: 10.1128/IAI.70.4.1783-1790.2002.

Fernandez-Gacio, A.; Uguen, M. & Fastrez, J. (2003) Phage display as a tool for the directed evolution of enzymes. *Trends in Biotechnology*, *21*(9), 408-414. doi: 10.1016/S0167-7799(03)00194-X.

Foster, P.L. (2006) Methods for determining spontaneous mutation rates. *Methods in Enzymology, 409,* 195-213. doi: 10.1016/S0076-6879(05)09012-9.

Francino, M. P.; Chao, L.; Riley, M. A. & Ochman, H. (1996) Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science*, *272*(5258), 107-109. doi: 10.1126/science.272.5258.107

Frederico, L. A.; Kunkel, T. A. & Shaw, B. R. (1990) A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry*, *29*(10), 2432-2537. doi: 10.1021/bi00462a015

Friedland, A. E.; Lu, T. K.; Wang, X.; Shi, D.; Church, G. & Collins, J. J. (2009) Synthetic gene networks that count. *Science*, *234*(5931), 1199-1202. doi: 10.1126/science.1172005

Gibson, D.G.; Young, L.; Chuang, R.Y.; Venter, J.C.,;Hutchison, C.A. III; Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods, 6,* 343-345. doi: 10.1038/nmeth.1318.

Golomb, M. & Chamberlin, M. (1974) Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major *in vitro* transcripts by gel electrophoresis. *The Journal of Biological Chemistry*, *249*(9), 2858-2863.

Greener, A. & Callahan, M. (1994) XL1-Red: a highly efficient random mutagenesis strain. *Strategies*, *7*, 32-34.

Greener, A.; Callahan, M. & Jerpseth, B. (1996) An efficient random mutagenesis technique using an *E. coli* mutator strain. *Methods in Molecular Biology (Clifton, N.J.)*, *57*, 375-385. doi: 10.1385/0-89603-332-5:375.

Guzman, L. M.; Belin, D.; Carson, M. J. & Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, *177*(14), 4121-4130.

Haldimann, A. & Wanner, B. L. (2001) Conditional-Replication, Integration, Excision, and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria. *Journal of Bacteriology*, *183*(21), 6384-6393. doi: 10.1128/JB.183.21.6384-6393.2001.

Ham, T. S.; Lee, S. K.; Keasling, J. D. & Arkin, A. P. (2006) A tightly regulated inducible expression system utilizing the fim inversion recombination switch. Biotechnology and *Bioengineering*, *94*(1), 1-4. doi: 10.1002/bit.20916.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology*, *166*(4), 557-580. doi: 10.1016/S0022-2836(83)80284-8.

Hanes, J. & Plückthun, A. (1997) *in vitro* selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 4937-4942.

Heim, R.; Prasher, D. C. & Tsien, R. Y. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(26), 12501-12504.

Hilvert, D. (2013) Design of Protein Catalysts. *Annual Review of Biochemistry*, *82*, 447-470. doi: 10.1146/annurev-biochem-072611-101825

Horton, R. M.; Hunt, H. D.; Ho, S. N.; Pullen, J. K. & Pease, L. R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, *77*(1), 61-68.

Jeng, S. T.; Gardner, J. F. & Gumport, R. I. (1990) Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators. *The Journal of Biological Chemistry*, *265*(7), 3823-3830.

Jin, D. J. & Gross, C. A. (1988) Mapping and sequencing of mutations in the Escherichia coli rpoB gene that lead to rifampicin resistance. *Journal of Molecular Biology*, *202*(5), 45-58.

Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D. & Houk, K. N. (2013) Computational enzyme design. *Angewandte Chemie International Edition*, *52*(22), 5700-5725. doi: 10.1002/anie.201204077

Kuderova, A.; Nanak, E.; Truksa, M. & Brzobohaty, B. (1999) Use of rifampicin in T7 RNA polymerase-driven expression of a plant enzyme: rifampicin improves yield and assembly. *Protein Expression and Purification*, *16*, 405-409 .

Larijani, M.; Petrov, A. P.; Kolenchenko, O.; Berru, M.; Krylov, S. N. & Martin, A. (2007) AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Molecular and Cellular Biology*, *27*(1), 20-30. doi: 10.1128/MCB.00824-06.

Lea, D.E. & Coulson, C.A. (1949) The distribution of the number of mutants in bacterial populations. *Journal of Genetics, 49*(3), 264-285.

Lieber, A.; Kiessling, U. & Strauss, M. (1989) High level gene expression in mammalian cells by a nuclear T7-phase RNA polymerase. *Nucleic Acids Research*, *17*(21), 8485-8493.

Lindahl, T. & Nyberg, B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, *13*(16), 3405-3410. doi: 10.1021/bi00713a035

Luria, S.E. & Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics, 28*(6), 491-511.

Lutsenko, E. & Bhagwat, A. S. (1999) The role of the *Escherichia coli* mug protein in the removal of uracil and 3,N(4)-ethenocytosine from DNA. *The Journal of Biological Chemistry*, *274*(43), 31034-31038. doi: 10.1074/jbc.274.43.31034.

Lutz, S. & Patrick, W. M. (2004) Novel methods for directed evolution of enzymes: quality, not quantity. *Current Opinion in Biotechnology*, *15*, 291-297. doi: 10.1016/j.copbio.2004.05.004.

Lyakhov, D. L.; He, B.; Zhang, X.; Studier, F. W.; Dunn, J. J. & McAllister, W. T. (1997) Mutant bacteriophage T7 RNA polymerases with altered termination properties. *Journal of Molecular Biology*, *269*(1), 28-40. doi: 10.1006/jmbi.1997.1015.

McAdams, H. H. & Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(3), 814-819.

McMillen, D.; Kopell, N.; Hasty, J. & Collins, J.J. (2002) Synchronizing genetic relaxation oscillators by intercell signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(2), 679-684. doi: 10.1073/pnas.022642299.

Miller, M.B. & Bassler, B.L. (2001) Quorum sensing in bacteria. *Annual Review of Microbiology, 55*, 165-199. doi: 10.1146/annurev.micro.55.1.165.

Morey, K. J.; Antunes, M. S.; Barrow, M. J.; Solorzano, F. A.; Havens, K. L.; Smith, J. J. & Medford, J. (2012) Crosstalk between endogenous and synthetic components – synthetic signaling meets endogenous components. *Biotechnology Journal*, *7*(7), 846-855. doi: 10.1002/biot.201100487.

Morita, T.; Mochizuki, Y. & Aiba, H. (2006) Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(13), 4858-4863. doi: 10.1073/pnas.0509638103

Muramatsu, M.; Sankaranand, V. S.; Anant, S.; Sugai, M.; Kinoshita, K.; Davidson, N. O.; Honjo, T. (1999) Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of Biological Chemistry*, *274*(26), 18470-18476. doi: 10.1074/jbc.274.26.18470.

Nakashima, N. & Tamura, T. (2009) Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. *Nucleic Acids Research*, *37*(15), e103. doi: 10.1093/nar/gkp498.

Needleman, S.B. & Wunsch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology, 48*(3), 443-453. doi: 10.1016/0022-2836(70)90057-4

Parsek, M.R. & Singh, P.K. (2003) Bacterial Biofilms: An emerging link to disease pathogenesis. *Annual Review of Microbiology, 57*, 677-701. doi: 10.1146/annurev.micro.57.030502.090720.

Passador, L.; Cook, J.M.; Gambello, M.J.; Rust, L. & Iglewski, B.H. (1993) Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication. *Science, 260*(5111), 1127-1130.

Pelkmans, L. (2012) Using Cell-to-Cell Variability– A New Era in Molecular Biology. *Science, 336*(6080), 425-426. doi: 10.1126/science.1222161

Petersen-Mahrt, S. K.; Harris, R. S. & Neuberger, M. S. (2002) AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*, *418*(6893), 99-103. doi: 10.1038/nature00862.

Pham, P.; Bransteitter, R.; Petruska, J. & Goodman, M. F. (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature*, *424*, 103-107. doi: 10.1038/nature01760.

Prochnow, C.; Bransteitter, R.; Klein, M. G.; Goodman, M. F. & Chen, X. S. (2007) The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature*, *445*(7126), 447-451. doi: 10.1038/nature05492.

Quan, J. & Tian, J. (2009) Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways. (P. L. Ho, Ed.)*PLoS ONE*, *4*(7), e6441. doi: 10.1371/journal.pone.0006441.

Ramalingam, K.I.; Tomshine, J.R.; Maynard, J.A. & Kaznessis, Y.N. (2009) Forward engineering of synthetic biological AND gates. *Biochemical Engineering Journal*, *47*(1-3), 38-47. doi: 10.1016/j.bej.2009.06.014.

Ramiro, A. R.; Stavropoulos, P.; Jankovic, M. & Nussenzweig, M. C. (2003) Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature Immunology*, *4*(5), 452-456. doi: 10.1038/ni920.

Romero, P. A. & Arnold, F. H. (2009) Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology 10*, 866-876. doi: 10.1038/nrm2805.

Rosche, W. & Foster, P. L. (2000) Determining Mutation Rates in Bacterial Populations. *Methods*, *20*(1), 4-17. doi: 10.1006/meth.1999.0901.

Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.;; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S. & Baker, D. (2008) Kemp elimination catalysts by computational enzyme design. *Nature*, *453*, 190-195. doi: 10.1038/nature06879

Rust, L.; Pesci, E.C. & Iglewski, B.H. (1996) Analysis of the *Pseudomonas aeruginosa* elastase (LasB) regulatory region. *Journal of Bacteriology, 178*(4), 1134-1140.

Rutherford, S.T. & Bassler, B.L. (2012) Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control. *Cold Spring Harbor Perspectives in Medicine, 2*(11). doi: 10.1101/cshperspect.a012427.

Sacho, E. J.; Tessmer, I.; Diaz, M. & Erie, D. A. (2008) Letter to the Editor: Response to "Is AID a monomer in solution". *DNA Repair*, *7*(3), 351352. doi: 10.1016/j.dnarep.2007.11.004.

Sambrook, J. & Russell, D.W. (2001) Molecular Cloning: A laboratory manual Vol 1-3. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Sarkar, S.; Ma, W. T. & Sandri, G. H. (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica*, *85*(2), 173-179.

Sayut, D.J.; Niu, Y. & Sun, L. (2009) Construction and enhancement of a minimal genetic AND logic gate. *Applied Environmental Microbiology*, *75*(3), 637.642. doi: 10.1128/AEM.01684-08.

Schreier, B.; Stumpp, C.; Wiesner, S. & Höcker, B. (2009) Computational design of ligand binding is not a solved problem. *Proceedings of the National Academy of Sciences of the United States of America, 106*(44), 18491-18496.

Seed, P.C.; Passador, L. & Iglewski, B.H. (1995) Activation of the *Pseudomonas aeruginosa lasI* gene by LasR and the Pseudomonas autoinducer PAI: an autoinduction regulatory hierarchy. *Journal of Bacteriology, 177*(3), 654-659.

Shetty, R. P.; Endy, D. & Knight, T. F. (2008) Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*, *2*(1), 5. doi: 10.1186/1754-1611-2-5.

Siddappa, N. B.; Avinash, A.; Venkatramanan, M. & Ranga, U. (2007) Regeneration of commercial nucleic acid extraction columns without the risk of carryover contamination. *BioTechniques*, *42*(2), 186, 188-192. doi: 10.2144/000112327.

Sleight, S. C.; Bartley, B. A.; Lieviant, J. A. & Sauro, H. M. (2010) In-Fusion BioBrick assembly and re-engineering. *Nucleic Acids Research*, *38*(8), 2624-2636. doi: 10.1093/nar/gkq179.

Sohail, A.; Klapacz, J.; Samaranayake, M.; Ullah, A. & Bhagwat, A. S. (2003) Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Research*, *31*(12), 2990-2994. doi: 10.1093/nar/gkg464.

Sousa, R.; Chung, Y. J.; Rose, J. P. & Wang, B. C. (1993) Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature*, *364*(6438), 593-599. doi: 10.1038/364593a0.

Spirin, A. S. (2004) High-throughput cell-free systems for synthesis of functionally active proteins. *Trends in Biotechnology*, *22*(10), 538-545. doi: 10.1016/j.tibtech.2004.08.012.

Spudich, J. L. & Koshland Jr., D. E. (1976) Non-genetic individuality: chance in the single cell. *Nature, 262*, 467-471. doi: 10.1038/262467a0

Studier, F. W. & Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), 113-130. doi: 10.1016/0022-2836(86)90385-2.

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*(7), 1272-1281. doi: 10.1016/j.cell.2009.04.048

Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L. & Baker, D. (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature, 501*, 212-218. doi: doi:10.1038/nature12443

Tupin, A.; Gualtieri, M.; Roquet-Banères, F.; Morichaud, Z.; Brodolin, K. & Leonetti, J.P. (2010) Resistance to rifampicin: at the crossroads between ecological, genomic and medical concerns. *International Journal of Antimicrobial Agents, 35*, 519-523. doi: 10.1016/j.ijantimicag.2009.12.017.

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*, 894-898. doi: 10.1038/nature08187.

Wang, B.; Kitney, R.I.; Joly, N.; & Buck, M. (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, *2*, 508. doi: 10.1038/ncomms1516.

Wang, B. & Buck, M. (2012) Customizing cell signaling using engineered genetic logic circuits. *Trends in microbiology*, *20*(8), 376-384, doi: 10.1016/j.tim.2012.05.001.

Wehrli, W. & Staehelin, M. (1971) Actions of the Rifamycins. *Bacteriological Reviews*, *35*(3), 290-309

Whiteley, M. & Greenberg, E.P. (2001) Promoter Specificity Elements in *Pseudomonas aeruginosa* Quorum-Sensing-Controlled Genes. *Journal of Bacteriology*, *183*(19), 5529-5534. doi: 10.1128/JB.183.19.5529-5534.2001.

Winson, M. K.; Camara, M.; Latifi, A.; Foglino, M.; Chhabra, S. R.; Daykin, M.; Bally, M.; Chapon, V.; Salmond, G. P. & B. W. Bycroft (1995) Multiple N-acyl-L-homoserine lactone signal molecules regulate production of virulence determinants and secondary metabolites in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(20), 9427– 9431

Xu, L.; Aha, P.; Gu, K.; Kuimelis, R. G.; Kurz, M.; Lam, T.; Lim, A.C,; Liu, H,; Lohse, P.A.; Sun, L.; Weng, S.; Wagner, R.W.; Lipovsek, D. (2002) Directed evolution of high-affinity antibody mimics using mRNA display. *Chemistry & Biology*, *9*(8), 933-942.

You, L.; Cox, R.S. III, Weiss, R. & Arnold, F. H. (2004) Programmed population control by cell-cell communication and regulated killing. *Nature*, *429*, 868-871. doi: 10.1038/nature02491.

9. Appendices

9.1. Plate counts for automated plasmid segregation

9.1.1. Automated run

Negative Control

	Repeat 1			Repeat 2			Repeat 3		
Culture	GFP+ve	GFP-ve	% GFP-ve	GFP+ve	GFP-ve	% GFP-ve	GFP+ve	GFP-ve	% GFP-ve
1				50	0	0.00%	19	2	9.52%
2	10	0	0.00%	28	0	0.00%	41	6	12.77%
3	61	0	0.00%	29	0	0.00%	13	0	0.00%
4	13	0	0.00%	7	0	0.00%			
5	13	0	0.00%	10	0	0.00%			
6	18	0	0.00%	8	0	0.00%	4	3	42.86%
7	20	0	0.00%	40	0	0.00%	18	3	14.29%
8	16	0	0.00%	28	0	0.00%	16	0	0.00%
9	14	0	0.00%	32	0	0.00%	29	0	0.00%
10	5	0	0.00%	3	0	0.00%			
11									
12	6	0	0.00%	21	0	0.00%			
13	4	0	0.00%	12	0	0.00%	45	0	0.00%
14	1	0	0.00%				99	0	0.00%
15	10	0	0.00%	26	0	0.00%			
16	5	0	0.00%						
17	11	0	0.00%	16	0	0.00%			
18	20	0	0.00%				1	0	0.00%
19	150	0	0.00%	2	0	0.00%			
20	51	0	0.00%	38	0	0.00%	2	0	0.00%
21	6	0	0.00%	1	0	0.00%	6	0	0.00%
22	3	0	0.00%	7	0	0.00%	9	0	0.00%
23	16	0	0.00%						
24	47	0	0.00%	16	0	0.00%	19	0	0.00%
Totals	**500**	**0**	**0.00%**	**374**	**0**	**0.00%**	**321**	**14**	**4.18%**

T7 RNAP alone

	Repeat 1			Repeat 2			Repeat 3		
Culture	GFP+ve	GFP-ve	% GFP-ve	GFP+ve	GFP-ve	% GFP-ve	GFP+ve	GFP-ve	% GFP-ve
1	14	0	0.00%	46	0	0.00%			
2	62	0	0.00%						
3	46	0	0.00%	1	0	0.00%			
4	1	0	0.00%	2	0	0.00%			
5				14	0	0.00%			
6	32	0	0.00%	45	0	0.00%			
7	46	0	0.00%	29	15	34.09%			
8	41	0	0.00%	13	0	0.00%			
9	24	0	0.00%	45	0	0.00%			
10				40	0	0.00%			
11	97	2	2.02%	95	0	0.00%			
12	3	0	0.00%	2	0	0.00%			
13				51	0	0.00%			
14				173	0	0.00%			
15	66	0	0.00%	20	0	0.00%			
16				2	0	0.00%			
17	117	0	0.00%	58	0	0.00%			
18	54	0	0.00%						
19	1	0	0.00%						
20	93	1	1.06%	15	0	0.00%			
21	28	0	0.00%	8	0	0.00%			
22	14	0	0.00%	31	0	0.00%			
23	106	0	0.00%	13	13	50.00%			
24	38	0	0.00%	1	1	50.00%			
Totals	**883**	**3**	**0.34%**	**704**	**29**	**3.96%**	**0**	**0**	

T7-AID fusion

	Repeat 1			Repeat 2			Repeat 3		
Culture	GFP+ve	GFP-ve	% GFP-ve	GFP+ve	GFP-ve	% GFP-ve	GFP+ve	GFP-ve	% GFP-ve
1	10	0	0.00%	9	9	50.00%			
2	30	0	0.00%	7	0	0.00%			
3	60	0	0.00%	15	0	0.00%	12	0	0.00%
4	20	0	0.00%	11	0	0.00%	35	0	0.00%
5	2	0	0.00%	17	0	0.00%	7	0	0.00%
6	18	0	0.00%	24	3	11.11%	5	0	0.00%
7	15	0	0.00%	2	0	0.00%	8	0	0.00%
8	127	0	0.00%	28	0	0.00%	16	0	0.00%
9	50	0	0.00%	18	0	0.00%			
10	34	4	10.53%	65	0	0.00%	24	0	0.00%
11	55	0	0.00%	120	0	0.00%	1	0	0.00%
12	14	0	0.00%	50	0	0.00%	3	0	0.00%
13	3	0	0.00%	3	0	0.00%			
14	85	0	0.00%				4	3	42.86%
15	21	0	0.00%						
16				1	0	0.00%	11	0	0.00%
17	48	0	0.00%				13	0	0.00%
18	112	0	0.00%	12	0	0.00%	7	0	0.00%
19	51	0	0.00%	4	0	0.00%	5	0	0.00%
20	20	0	0.00%	8	0	0.00%	55	0	0.00%
21	68	0	0.00%				1	0	0.00%
22	91	0	0.00%	59	0	0.00%	29	0	0.00%
23	39	1	2.50%	43	0	0.00%			
24	2	0	0.00%	2	0	0.00%	31	0	0.00%
Totals	**975**	**5**	**0.51%**	**498**	**12**	**2.35%**	**267**	**3**	**1.11%**

T7 RNAP & AID unfused

	Repeat 1			Repeat 2			Repeat 3		
Culture	GFP+ve	GFP-ve	% GFP-ve	GFP+ve	GFP-ve	% GFP-ve	GFP+ve	GFP-ve	% GFP-ve
1				9	0	0.00%	0	0	
2				67	0	0.00%	57	0	0.00%
3	236	0	0.00%	106	0	0.00%	38	22	36.67%
4	158	0	0.00%	28	0	0.00%	21	21	50.00%
5	120	0	0.00%	104	0	0.00%			
6				134	22	14.10%	40	0	0.00%
7	1	0	0.00%	185	1	0.54%	105	0	0.00%
8				80	0	0.00%	196	9	4.39%
9	197	0	0.00%						
10	321	0	0.00%	31	3	8.82%	215	0	0.00%
11	105	0	0.00%	227	9	3.81%			
12	7	0	0.00%	12	0	0.00%	93	6	6.06%
13	258	0	0.00%	0	0		10	10	50.00%
14				84	40	32.26%			
15	100	0	0.00%	47	45	48.91%	232	0	0.00%
16				8	2	20.00%	1	0	0.00%
17				20	3	13.04%	167	77	31.56%
18	130	11	7.80%	92	0	0.00%	58	0	0.00%
19				36	0	0.00%	61	0	0.00%
20	266	0	0.00%	100	0	0.00%	266	21	7.32%
21				48	12	20.00%	101	18	15.13%
22				1	0	0.00%	9	0	0.00%
23	148	24	13.95%	73	23	23.96%			
24	7	0	0.00%	31	0	0.00%	104	0	0.00%
Totals	**2054**	**35**	**1.68%**	**1523**	**160**	**9.51%**	**1774**	**184**	**9.40%**

Figure 32: Colony counts of the semi-automated multi-copy mutation assays

Plasmids were purified from Bh156 and transformed into BL21 for plating and scoring. Colony counts are separated into GFP +ve (fluorescent) and GFP -ve (non-fuorescent). Top, left: pCD:GFP alone, no device plasmid. Top, right: pCD:GFP & pA3:T. Bottom, left: pCD:GFP & pA3:TA. Bottom, right: pCD:GFP & pA3:T+A

9.1.2. Semi-automated run

T7-AID fusion									
	Repeat 1			Repeat 2			Repeat 3		
Culture	GFP +ve	GFP -ve	% GFP -ve	GFP +ve	GFP -ve	% GFP -ve	GFP +ve	GFP -ve	% GFP -ve
1	6	0	0.00%	48	0	0.00%	74	0	0.00%
2	494	0	0.00%	218	2	0.91%	96	0	0.00%
3	661	7	1.05%	180	0	0.00%	351	0	0.00%
4	387	0	0.00%	338	5	1.46%	322	0	0.00%
5	1117	7	0.62%	696	0	0.00%	575	0	0.00%
6	110	0	0.00%	186	3	1.59%	137	0	0.00%
7	184	0	0.00%	548	0	0.00%	336	0	0.00%
8	186	0	0.00%	792	0	0.00%	838	0	0.00%
9	334	17	4.84%	1003	0	0.00%	276	5	1.78%
10	223	0	0.00%	287	5	1.71%	231	0	0.00%
11	16	0	0.00%	506	0	0.00%	460	0	0.00%
12	203	6	2.87%	193	0	0.00%	192	4	2.04%
13	504	0	0.00%	762	0	0.00%	1196	1	0.08%
14	970	4	0.41%	1119	0	0.00%	1325	0	0.00%
15	423	0	0.00%	73	0	0.00%	294	0	0.00%
16	229	0	0.00%	359	0	0.00%	879	0	0.00%
17	475	0	0.00%	344	0	0.00%	858	0	0.00%
18	257	0	0.00%	940	5	0.53%	225	0	0.00%
19	950	0	0.00%	403	0	0.00%	300	1	0.33%
20	1537	0	0.00%	1923	0	0.00%	997	0	0.00%
21	139	0	0.00%	1012	0	0.00%	1638	0	0.00%
22	476	0	0.00%	635	0	0.00%	433	0	0.00%
23	451	0	0.00%	1908	0	0.00%	654	0	0.00%
24	774	5	0.64%	668	3	0.45%	266	0	0.00%
Totals	**11106**	**46**	**0.41%**	**15141**	**23**	**0.15%**	**12953**	**11**	**0.08%**

T7 RNAP & AID unfused									
	Repeat 1			Repeat 2			Repeat 3		
Culture	GFP +ve	GFP -ve	% GFP -ve	GFP +ve	GFP -ve	% GFP -ve	GFP +ve	GFP -ve	% GFP -ve
1	266	189	41.54%	2337	3	0.13%	943	0	0.00%
2	1723	29	1.66%	789	0	0.00%	1473	0	0.00%
3	1349	78	5.47%	1671	12	0.71%	1756	0	0.00%
4	757	0	0.00%	1769	4	0.23%	1376	0	0.00%
5	799	923	53.60%	1730	3	0.17%	2797	0	0.00%
6	767	23	2.91%	1605	0	0.00%	1586	0	0.00%
7	1896	280	12.87%	1603	0	0.00%	458	0	0.00%
8	1284	36	2.73%	904	0	0.00%	2463	0	0.00%
9	1179	471	28.55%	730	1	0.14%	1568	0	0.00%
10	911	322	26.12%	1265	30	2.32%	1158	0	0.00%
11	623	545	46.66%	1800	2	0.11%	1432	0	0.00%
12	316	72	18.56%	2070	0	0.00%	3919	0	0.00%
13	915	268	22.65%	2021	0	0.00%	154	0	0.00%
14	194	0	0.00%	830	0	0.00%	1111	0	0.00%
15	1619	6	0.37%	935	0	0.00%	0	0	
16	861	112	11.51%	1785	0	0.00%	1112	0	0.00%
17	2253	370	14.11%	1198	0	0.00%	1638	0	0.00%
18	1564	23	1.45%	644	0	0.00%	3097	0	0.00%
19	3797	735	16.22%	284	0	0.00%	1545	0	0.00%
20	419	888	67.94%	981	0	0.00%	1158	0	0.00%
21	1415	347	19.69%	1440	0	0.00%	1446	28	1.90%
22	1003	702	41.17%	1181	0	0.00%	1761	0	0.00%
23	605	24	3.82%	1236	0	0.00%	1668	0	0.00%
24	1926	207	9.70%	0	0		0	0	
Totals	**28441**	**6650**	**18.95%**	**30808**	**55**	**0.18%**	**35619**	**28**	**0.08%**

Figure 33: Colony counts of the automated multi-copy mutation assays

Plasmids were purified from Bh156 and transformed into BL21 for plating and scoring. Colony counts are separated into GFP +ve (fluorescent) and GFP -ve (non-fluorescent). Top, pC5:GFP & pA3:TA. Bottom, pC5:GFP & pA3:T+A

9.2. Automation script

9.2.1. Miniprep & transformation

```
Included C:\RoboManager\Config\TFB Config.rss
0001  Initialize System
      >
      >Qiagen MagAttract 96 Protocol
      >
0005  Put Empty Plate "Lysate Plate" on B3 / Shaker
      >
      >Please define your sample layout before starting, by using the below command.
      >vvvvvvvvvvvvvvvvvvvvvvvvvv
0009  Define Selection on B3 / Shaker
      > ^^^^^^^^^^^^^^^^^^^^^^^^^^
      >
0012  Put Empty Plate "Magnetic Bead Separation" on D4 Magnet
0013  Put Empty Plate "Purified Plasmid DNA Plate" on B6
0014  Put Empty Plate "Reagents Plate" on C2
0015  Put Empty Plate "Compression Pad" on D2 / Compression Pad Position
0016  Put Empty Plate "Cells plate" on C4 / 96 cooled Pipetting Position
      >
0018  User Input : Layout, beads, PLATE TYPE? (y/n)
0019  User Input : Just the miniprep, or a subsequent transformation as well? (m/t)
0020  Jump To Label "Layout Check No" if Datakey "UserInput <Layout Check>" Contains n.
0021  Jump To Label "Layout Check No" if Datakey "UserInput <Layout Check>" Contains N.
0022  Jump To Label "Layout Check Yes" if Datakey "UserInput <Layout Check>" Contains y.
0023  Jump To Label "Layout Check Yes" if Datakey "UserInput <Layout Check>" Contains Y.
      >test stuff - ignore!
0025  Jump To Label "test" if Datakey "UserInput <Layout Check>" Is Equal to test.
      >test stuff - ignore!
0027  Jump To Label "Layout Check No" if Datakey "" Is Equal to .
0028  :: test
0029  Loop Start
0030      Initialize System
0031      Activate Shaker for 5 s at Speed 500, Shake Direction OFF
0032      FlushTips
0033  Loop End
0034  Jump To Label "Layout Check No" if Datakey "" Is Equal to .
0035  :: Layout Check Yes
0036  Initialize System
      >Miniprep protocol starting, using pre-defined layout...
0038  Jump To Label "No transformation" if Datakey "UserInput <Protocol>" not Contains t.
0039  Set Temperature of Cooled Position(s)
0040  Message : Make sure to put the cells in a TSP plate on C4 when it's cooled!
0041  :: No transformation
0042  Distribute 90.0 µl from C2 to B3 / Shaker
0043  Activate Shaker for 300 s at Speed 500, Shake Direction CW
0044  Wait for Shaker
0045  Activate Shaker for 300 s at Speed 1000, Shake Direction ANTI_CW
0046  FlushTips
0047  Wait for Shaker
0048  Distribute 100.0 µl from C2 to B3 / Shaker
0049  Activate Shaker for 180 s at Speed 1000, Shake Direction CW
0050  FlushTips
0051  Wait for Shaker
0052  Sleep for 1 Minutes and 0 Seconds
0053  Activate Shaker for 135 s at Speed 1000, Shake Direction ANTI_CW
0054  Wait for Shaker
0055  Distribute 125.0 µl from C2 to B3 / Shaker
0056  Distribute 125.0 µl from C2 to B3 / Shaker
0057  FlushTips
0058  Sleep for 0 Minutes and 30 Seconds
0059  Activate Shaker for 120 s at Speed 1000, Shake Direction CW
0060  Wait for Shaker
0061  Move Plate from B3 / Shaker to C3
0062  Loop Start
0063      Transfer 200.0 µl from C3 to D4 Magnet
0064      Move Plate from D4 Magnet to B3 / Shaker
0065      Activate Shaker for 300 s at Speed 1000, Shake Direction CW
0066      FlushTips
0067      Wait for Shaker
0068      Move Plate from B3 / Shaker to D4 Magnet
0069      Sleep for 1 Minutes and 0 Seconds
0070      Waste 300.0 µl from D4 Magnet
0071  Loop End
0072  Distribute 100.0 µl from C2 to D4 Magnet
0073  Move Plate from D4 Magnet to B3 / Shaker
0074  Activate Shaker for 60 s at Speed 1000, Shake Direction CW
0075  Wait for Shaker
0076  Move Plate from B3 / Shaker to D4 Magnet
0077  Sleep for 1 Minutes and 0 Seconds
0078  Waste 300.0 µl from D4 Magnet
      >Below is a 'loop' of PE wash x3, but seperated so it uses diff. wells from the reagents plate
0080  Loop Start
0081      Distribute 100.0 µl from C2 to D4 Magnet
0082      Move Plate from D4 Magnet to B3 / Shaker
0083      Activate Shaker for 60 s at Speed 1000, Shake Direction CW
0084      Wait for Shaker
0085      Move Plate from B3 / Shaker to D4 Magnet
0086      Sleep for 2 Minutes and 0 Seconds
0087      Waste 300.0 µl from D4 Magnet
0088      Distribute 100.0 µl from C2 to D4 Magnet
0089      Move Plate from D4 Magnet to B3 / Shaker
0090      Activate Shaker for 60 s at Speed 1000, Shake Direction CW
0091      Wait for Shaker
0092      Move Plate from B3 / Shaker to D4 Magnet
0093      Sleep for 2 Minutes and 0 Seconds
0094      Waste 300.0 µl from D4 Magnet
0095      Distribute 100.0 µl from C2 to D4 Magnet
0096      Move Plate from D4 Magnet to B3 / Shaker
0097      Activate Shaker for 60 s at Speed 1000, Shake Direction CW
0098      Wait for Shaker
0099      Move Plate from B3 / Shaker to D4 Magnet
0100      Sleep for 2 Minutes and 0 Seconds
0101      Waste 300.0 µl from D4 Magnet
0102  Loop End
0103  Waste 300.0 µl from D4 Magnet
0104  Distribute 75.0 µl from C2 to D4 Magnet
0105  Move Plate from D4 Magnet to B3 / Shaker
0106  Activate Shaker for 300 s at Speed 1000, Shake Direction CW
0107  Wait for Shaker
0108  Move Plate from B3 / Shaker to D4 Magnet
0109  Sleep for 2 Minutes and 0 Seconds
      >'Elute' DNA
0111  Open Thermo Cycler "Primus 1"
0112  Transfer 50.0 µl from D4 Magnet to B6
0113  Move Plate from B6 to Primus 1
0114  Close Thermo Cycler "Primus 1"
0115  Start PCR Program "HOLD80.CYC" on Thermo Cycler "Primus 1"
0116  Park Robot
0117  Wait for Thermo Cycler "Primus 1"
      >Transform DNA or infinite 4 degree hold!
0119  Jump To Label "No transformation 2" if Datakey "UserInput <Protocol>" not Contains t.
      >Transformation!
0121  Open Thermo Cycler "Primus 1"
0122  Start PCR Program "HOLD4OPE.CYC" on Thermo Cycler "Primus 1"
0123  Sleep for 1 Minutes and 0 Seconds
0124  Put Empty Plate "Evaporated Plasmid DNA Plate" on Primus for CP only Handler
0125  Move Plate from Primus for CP only Handler to B6
0126  Sleep for 1 Minutes and 0 Seconds
0127  Transfer 5.0 µl from B6 to C4 / 96 cooled Pipetting Position
0128  FlushTips
0129  Sleep for 30 Minutes and 0 Seconds
0130  Move Plate from C4 / 96 cooled Pipetting Position to Primus for CP only Handler
0131  Set Temperature of Cooled Position(s)
0132  Wait For Temperature of Cooled Position(s)
0133  Move Plate from Primus for CP only Handler to C4 / 96 cooled Pipetting Position
0134  Sleep for 0 Minutes and 45 Seconds
0135  Move Plate from C4 / 96 cooled Pipetting Position to Primus for CP only Handler
0136  Set Temperature of Cooled Position(s)
0137  Sleep for 5 Minutes and 0 Seconds
0138  Move Plate from Primus for CP only Handler to C4 / 96 cooled Pipetting Position
0139  Distribute 150.0 µl from C2 to C4 / 96 cooled Pipetting Position
0140  Move Plate from C4 / 96 cooled Pipetting Position to Primus for CP only Handler
0141  Set Temperature of Cooled Position(s)
0142  Initialize System
0143  Open Thermo Cycler "Primus 1"
0144  Remove Plate/TipRack from Primus for CP only Handler
0145  Remove Plate/TipRack from Primus 1
0146  Put Empty Plate "Transformation Plate" on Primus 1
0147  Move Plate from D2 / Compression Pad Position to Primus for CP only Handler
0148  Close Thermo Cycler "Primus 1"
0149  Start PCR Program "37TO4INF.CYC" on Thermo Cycler "Primus 1"
0150  Jump To Label "Transfer DNA and Finish" if Datakey "" Is Equal to .
0151  :: No transformation 2
0152  Move Plate from D2 / Compression Pad Position to Primus for CP only Handler
0153  Close Thermo Cycler "Primus 1"
0154  Start PCR Program "HOLD4INF.CYC" on Thermo Cycler "Primus 1"
0155  Jump To Label "Transfer DNA and Finish" if Datakey "" Is Equal to .
      >Put DNA into final miniprep plate
0157  :: Transfer DNA and Finish
0158  Message : Welcome back! Hit enter to end the protocol; you can check on the PCR status using the graphic to the left.
0159  Initialize System
0160  Open Thermo Cycler "Primus 1"
0161  Move Plate from Primus for CP only Handler to D2 / Compression Pad Position
0162  Jump To Label "Finish" if Datakey "" Is Equal to .
      >Finish up
0164  :: Finish
0165  Set Temperature of Cooled Position(s)
0166  FlushTips
0167  Park Robot
0168  Wait for Thermo Cycler "Primus 1"
0169  :: Layout Check No
0170  Park Robot
```

Figure 34: The automation script for combined plasmid purification and transformation

9.3. Details of mutations detected by sequencing (4.3.3)

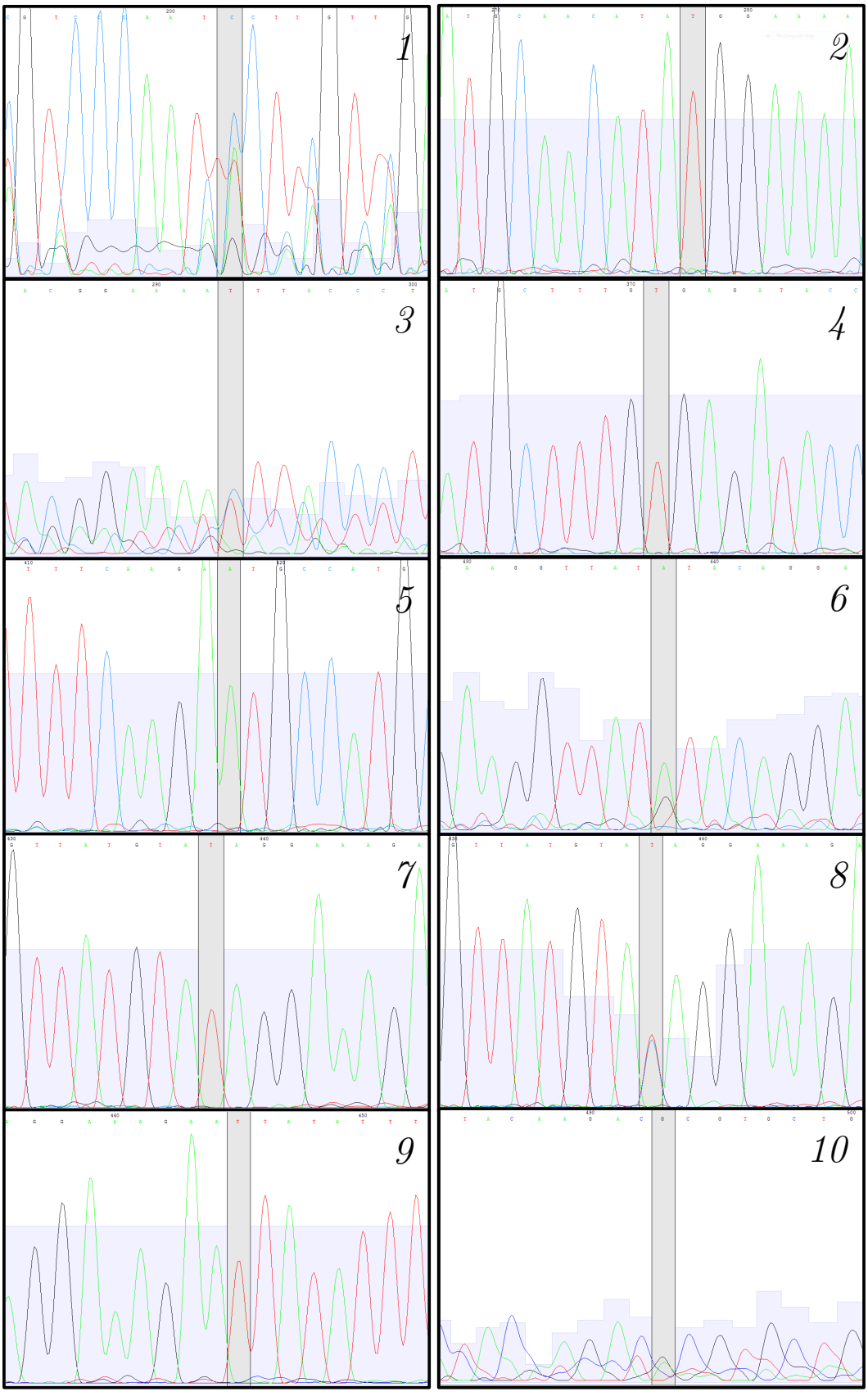
9.3.1. Mutation information

No.	***Mutation***	***Sequence***	***Trace position***	***GFP position***	***AA result***
1	*ATT–ATC*	*F,-,6*	*203*	*42*	*Silent*
2	*TAC–TAT*	*F,-,10*	*278*	*117*	*Silent*
3	*CTT–TTT*	*F,-,2*	*293*	*124*	*Leu–Phe*
4	*GCG–GTG*	*F,-,4*	*371*	*215*	*Ala–Val*
5	*AGT–AAT*	*F,-,10*	*418*	*257*	*Ser–Asn*
6	*GTA–ATA*	*F,-,8*	*438*	*277*	*Val–Ile*
7	*CAG–TAG*	*F,-,7*	*438*	*280*	*Gln–STOP*
8	*CAG–TAG*	*F,+,6*	*438*	*280*	*Gln–STOP*
9	*ACT–ATT*	*F,-,3*	*445*	*290*	*Thr–Ile*
10	*ACA–ACG*	*F,-,2*	*493*	*324*	*Silent*
11	*AGA–ATA*	*F,-,2*	*534*	*365*	*Arg–Ile*
12	*AAC–AAT*	*F,+,7*	*596*	*438*	*Silent*
13	*AGC–AGT*	*F,-,8*	*686*	*525*	*Silent*
14	*AGC–AGT*	*F,-,7*	*683*	*525*	*Silent*
15	*AGC–AGT*	*F,+,6*	*683*	*525*	*Silent*
16	*AGC–AGT*	*F,-,3*	*680*	*525*	*Silent*
17	*TAC–TAT*	*F,+,8*	*763*	*600*	*Silent*
18	*ACA–ATA*	*F,+,10*	*830*	*674*	*Thr–Ile*
19	*TAC–TAT*	*F,-,1*	*867*	*711*	*Silent*

Table 6: Further information on mutation events

Sequence key – X,Y,Z where X is F for the fusion, U for unfused T7-AID and T for T7 RNAP alone; Y is – or + corresponding to knockouts or fluorescent colonies; Z is the sequence number from 1-10.

9.3.2. Trace files



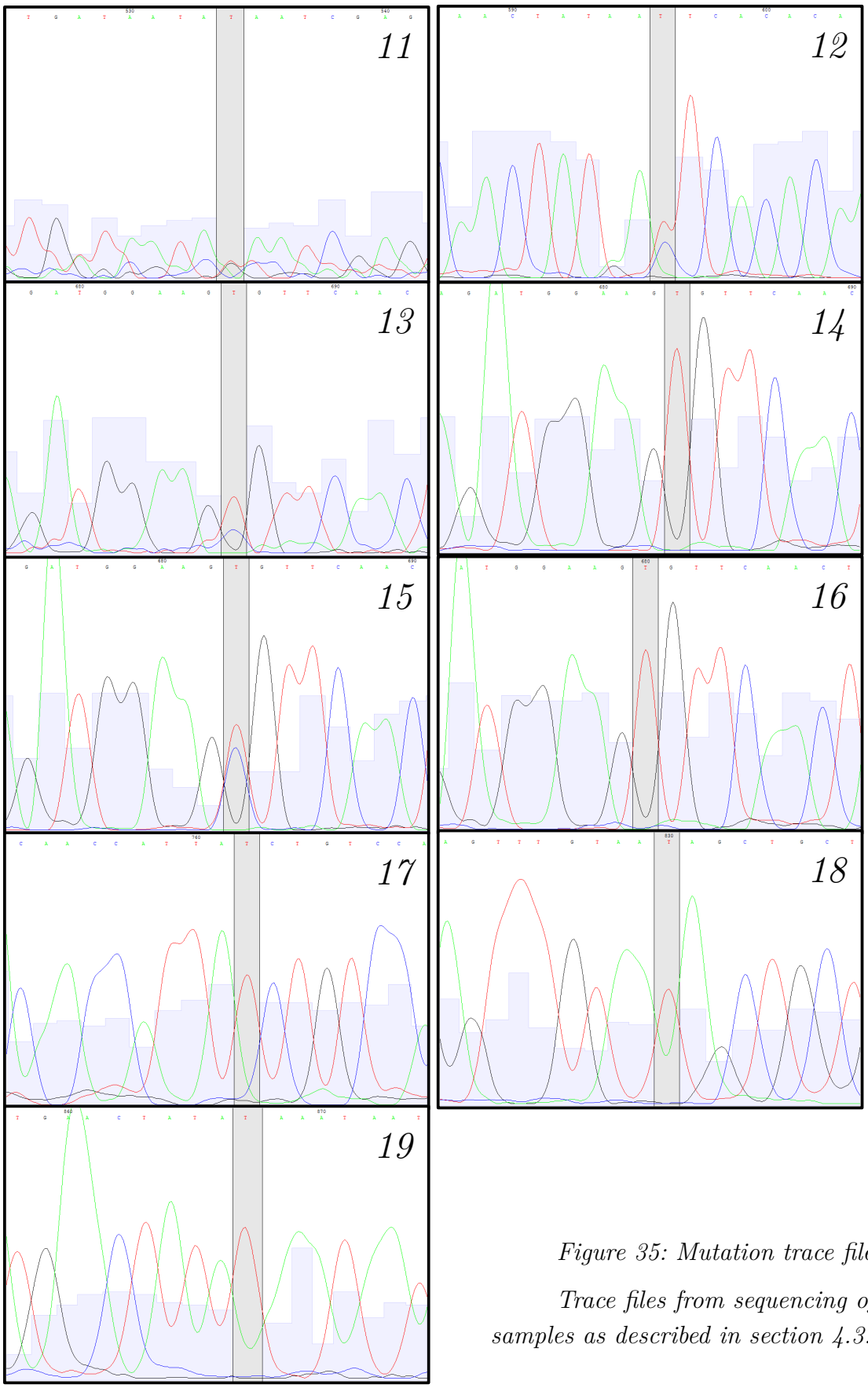


Figure 35: Mutation trace files

Trace files from sequencing of samples as described in section 4.3.3

9.4. DNA sequences of construct parts

9.4.1. Sequence information

Name	Description	Sequence
P_{T7}	T7 RNAP promoter	TAATACGACTCACTATA
P_{Lac}	Lac-inducible promoter	AATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACA
RBS32	Moderate-strength RBS	TCACACAGGAAAG
RBS33	Weak RBS	TCACACAGGAC
RBS34	Strong RBS	AAAGAGGAGAAA
B0015 (XX)	Double transcriptional terminator	CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACG CTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA
AID	Activation-Induced Deaminase	ATGGATAGCCTGCTGATGAATCGTCGCAAATTTCTGTATCAGTTTAAAAATGTGCGTTGGGCCAAAGGTCGTCGT GAAACCTATCTGTGCTATGTTGTGAAACGTCGTGATAGCGCAACCAGCTTTAGCCTGGATTTTGGTTATCTGCGC AATAAAAATGGTTGTCATGTGGAGCTGCTGTTTCTGCGTTATATTAGCGATTGGGATCTGGATCCGGGTCGTTGT TATCGTGTTACCTGGTTTACCAGCTGGTCACCGTGTTATGACTGTGCACGTCATGTTGCAGATTTTCTGCGTGGT AATCCGAATCTGAGCCTGCGTATTTTTACCGCACGTCTGTATTTTTGCGAAGATCGTAAAGCAGAACCGGAAGGT CTGCGTCGTCTGCATCGTGCAGGTGTTCAGATTGCCATTATGACCTTTAAAGATTATTTTTATTGCTGGAATACCT TTGTGGAAAATCATGAACGCACCTTTAAAGCATGGGAAGGCCTGCATGAAAATAGCGTTCGTCTGTCTCGCCAGC TGCGTCGTATTCTGCTGCCGCTGTATGAAGTTGATGATCTGCGTGATGCCTTTCGTACCCTGGGTTTATAA
T7 RNAP	RNA Polymerase from T7 bacteriophage	ATGAACACGATTAACATCGCTAAGAACGACTTCTCTGACATCGAACTGGCTGCTATCCCGTTCAACACTCTGGCTG ACCATTACGGTGAGCGTTTAGCTCGCGAACAGTTGGCCCTTGAGCATGAGTCTTACGAGATGGGTGAAGCACGCT TCCGCAAGATGTTTGAGCGTCAACTTAAAGCTGGTGAGGTTGCGGATAACGCTGCCGCCAAGCCTCTCATCACTA CCCTACTCCCTAAGATGATTGCACGCATCAACGACTGGTTTGAGGAAGTGAAAGCTAAGCGCGGCAAGCGCCCGA CAGCCTTCCAGTTCCTGCAAGAAATCAAGCCGGAAGCCGTAGCGTACATCACCATTAAGACCACTCTGGCTTGCC TAACCAGTGCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTCGGGCCATTGAGGACGAGGCTCGCT TCGGTCGTATCCGTGACCTTGAAGCTAAGCACTTCAAGAAAAACGTTGAGGAACAACTCAACAAGCGCGTAGGGC ACGTCTACAAGAAAGCATTTATGCAAGTTGTCGAGGCTGACATGCTCTCTAAGGGTCTACTCGGTGGCGAGGCGT GGTCTTCGTGGCATAAGGAAGACTCTATTCATGTAGGAGTACGCTGCATCGAGATGCTCATTGAGTCAACCGGAA TGGTTAGCTTACACCGCCAAAATGCTGGCGTAGTAGGTCAAGACTCTGAGACTATCGAACTCGCACCTGAATACG CTGAGGCTATCGCAACCCGTGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTCCAACCTTGCGTAGTTCCTCCTA AGCCGTGGACTGGCATTACTGGTGGTGGCTATTGGGCTAACGGTCGTCGTCCTCTGGCGCTGGTGCGTACTCACA GTAAGAAAGCACTGATGCGCTACGAAGACGTTTACATGCCTGAGGTGTACAAAGCGATTAACATTGCGCAAAACA CCGCATGGAAAATCAACAAGAAAGTCCTAGCGGTCGCCAACGTAATCACCAAGTGGAAGCATTGTCCGGTCGAGG ACATCCCTGCGATTGAGCGTGAAGAACTCCCGATGAAACCGGAAGACATCGACATGAATCCTGAGGCTCTCACCG CGTGGAAACGTGCTGCCGCTGCTGTGTACCGCAAGGACAGGGCTCGCAAGTCTCGCCGTATCAGCCTTGAGTTC ATGCTTGAGCAAGCCAATAAGTTTGCTAACCATAAGGCCATCTGGTTCCCTTACAACATGGACTGGCGCGGTCGT GTTTACGCCGTGTCAATGTTCAACCCGCAAGGTAACGATATGACCAAAGGACTGCTTACGCTGGCGAAAGGTAAA CCAATCGGTAAGGAAGGTTACTACTGGCTGAAAATCCACGGTGCAAACTGTGCGGGTGTCGATAAGGTTCCGTTC CCTGAGCGCATCAAGTTCATTGAGGAAAACCACGAGAACATCATGGCTTGCGCTAAGTCTCCACTGGAGAACACT TGGTGGGCTGAGCAAGATTCTCCGTTCTGCTTCCTTGCGTTCTGCTTTGAGTACGCTGGGGTACAGCACCACGGC CTGAGCTATAACTGCTCCCTTCCGCTGGCGTTTGACGGGTCTTGCTCTGGCATCCAGCACTTCTCCGCGATGCTC CGAGATGAGGTAGGTGGTCGCGCGGTTAACTTGCTTCCTAGTGAGACCGTTCAGGACATCTACGGGATTGTTGCT AAGAAAGTCAACGAGATTCTACAAGCAGACGCAATCAATGGGACCGATAACGAAGTAGTTACCGTGACCGATGAG AACACTGGTGAAATCTCTGAGAAAGTCAAGCTGGGCACTAAGGCACTGGCTGGTCAATGGCTGGCTCACGGTGT TACTCGCAGTGTGACTAAGCGTTCAGTCATGACGCTGGCTTACGGGTCCAAAGAGTTCGGCTTCCGTCAACAAGT GCTGGAAGATACCATTCAGCCAGCTATTGATTCCGGCAAGGGTCCGATGTTCACTCAGCCGAATCAGGCTGCTGG ATACATGGCTAAGCTGATTTGGGAATCTGTGAGCGTGACGGTGGTAGCTGCGGTTGAAGCAATGAACTGGCTTAA GTCTGCTGCTAAGCTGCTGGCTGCTGAGGTCAAAGATAAGAAGACTGGAGAGATTCTTCGCAAGCGTTGCGCTG TGCATTGGGTAACTCCTGATGGTTTCCCTGTGTGGCAGGAATACAAGAAGCCTATTCAGACGCGCTTGAACCTGA TGTTCCTCGGTCAGTTCCGCTTACAGCCTACCATTAACACCAACAAAGATAGCGAGATTGATGCACACAAACAGGA GTCTGGTATCGCTCCTAACTTTGTACACAGCCAAGACGGTAGCCACCTTCGTAAGACTGTAGTGTGGGCACACGA

		GAAGTACGGAATCGAATCTTTTGCACTGATTCACGACTCCTTCGGTACCATTCCGGCTGACGCTGCGAACCTGTT CAAAGCAGTGCGCGAAACTATGGTTGACACATATGAGTCTTGTGATGTACTGGCTGATTTCTACGACCAGTTCGC TGACCAGTTGCACGAGTCTCAATTGGACAAAATGCCAGCACTTCCGGCTAAAGGTAACTTGAACCTCCGTGACAT CTTAGAGTCGGACTTCGCGTTCGCGTAA
T7-AID	Fusion protein	ATGGATAGCCTGCTGATGAATCGTCGCAAATTTCTGTATCAGTTTAAAAATGTGCGTTGGGCCAAAGGTCGTCGT GAAACCTATCTGTGCTATGTTGTGAAACGTCGTGATAGCGCAACCAGCTTTAGCCTGGATTTTGGTTATCTGCGC AATAAAAATGGTTGTCATGTGGAGCTGCTGTTTCTGCGTTATATTAGCGATTGGGATCTGGATCCGGGTCGTTGT TATCGTGTTACCTGGTTTACCAGCTGGTCACCGTGTTATGACTGTGCACGTCATGTTGCAGATTTTCTGCGTGGT AATCCGAATCTGAGCCTGCGTATTTTTACCGCACGTCTGTATTTTTGCGAAGATCGTAAAGCAGAACCGGAAGGT CTGCGTCGTCTGCATCGTGCAGGTGTTCAGATTGCCATTATGACCTTTAAAGATTATTTTTATTGCTGGAATACCT TTGTGGAAAATCATGAACGCACCTTTAAAGCATGGGAAGGCCTGCATGAAAATAGCGTTCGTCTGTCTCGCCAGC TGCGTCGTATTCTGCTGCCGCTGTATGAAGTTGATGATCTGCGTGATGCCTTTCGTACCCTGGGTTTAACCGGTC AGGAAGTTGCATGCACCGCCGGCAACACGATTAACATCGCTAAGAACGACTTCTCTGACATCGAACTGGCTGCTA TCCCGTTCAACACTCTGGCTGACCATTACGGTGAGCGTTTAGCTCGCGAACAGTTGGCCCTTGAGCATGAGTCTT ACGAGATGGGTGAAGCACGCTTCCGCAAGATGTTTGAGCGTCAACTTAAAGCTGGTGAGGTTGCGGATAACGCT GCCGCCAAGCCTCTCATCACTACCCTACTCCCTAAGATGATTGCACGCATCAACGACTGGTTTGAGGAAGTGAAA GCTAAGCGCGGCAAGCGCCCGACAGCCTTCCAGTTCCTGCAAGAAATCAAGCCGGAAGCCGTAGCGTACATCACC ATTAAGACCACTCTGGCTTGCCTAACCAGTGCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTCGG GCCATTGAGGACGAGGCTCGCTTCGGTCGTATCCGTGACCTTGAAGCTAAGCACTTCAAGAAAAACGTTGAGGAA CAACTCAACAAGCGCGTAGGGCACGTCTACAAGAAAGCATTTATGCAAGTTGTCGAGGCTGACATGCTCTCTAAG GGTCTACTCGGTGGCGAGGCGTGGTCTTCGTGGCATAAGGAAGACTCTATTCATGTAGGAGTACGCTGCATCGAG ATGCTCATTGAGTCAACCGGAATGGTTAGCTTACACCGCCAAAATGCTGGCGTAGTAGGTCAAGACTCTGAGACT ATCGAACTCGCACCTGAATACGCTGAGGCTATCGCAACCCGTGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTC CAACCTTGCGTAGTTCCTCCTAAGCCGTGGACTGGCATTACTGGTGGTGGCTATTGGGCTAACGGTCGTCGTCCT CTGGCGCTGGTGCGTACTCACAGTAAGAAAGCACTGATGCGCTACGAAGACGTTTACATGCCTGAGGTGTACAAA GCGATTAACATTGCGCAAAACACCGCATGGAAAATCAACAAGAAAGTCCTAGCGGTCGCCAACGTAATCACCAAG TGGAAGCATTGTCCGGTCGAGGACATCCCTGCGATTGAGCGTGAAGAACTCCCGATGAAACCGGAAGACATCGAC ATGAATCCTGAGGCTCTCACCGCGTGGAAACGTGCTGCCGCTGCTGTGTACCGCAAGGACAGGGCTCGCAAGTC TCGCCGTATCAGCCTTGAGTTCATGCTTGAGCAAGCCAATAAGTTTGCTAACCATAAGGCCATCTGGTTCCCTTAC AACATGGACTGGCGCGGTCGTGTTTACGCCGTGTCAATGTTCAACCCGCAAGGTAACGATATGACCAAAGGACTG CTTACGCTGGCGAAAGGTAAACCAATCGGTAAGGAAGGTTACTACTGGCTGAAAATCCACGGTGCAAACTGTGCG GGTGTCGATAAGGTTCCGTTCCCTGAGCGCATCAAGTTCATTGAGGAAAACCACGAGAACATCATGGCTTGCGCT AAGTCTCCACTGGAGAACACTTGGTGGGCTGAGCAAGATTCTCCGTTCTGCTTCCTTGCGTTCTGCTTTGAGTAC GCTGGGGTACAGCACCACGGCCTGAGCTATAACTGCTCCCTTCCGCTGGCGTTTGACGGGTCTTGCTCTGGCATC CAGCACTTCTCCGCGATGCTCCGAGATGAGGTAGGTGGTCGCGCGGTTAACTTGCTTCCTAGTGAGACCGTTCAG GACATCTACGGGATTGTTGCTAAGAAAGTCAACGAGATTCTACAAGCAGACGCAATCAATGGGACCGATAACGAA GTAGTTACCGTGACCGATGAGAACACTGGTGAAATCTCTGAGAAAGTCAAGCTGGGCACTAAGGCACTGGCTGGT CAATGGCTGGCTCACGGTGTTACTCGCAGTGTGACTAAGCGTTCAGTCATGACGCTGGCTTACGGGTCCAAAGAG TTCGGCTTCCGTCAACAAGTGCTGGAAGATACCATTCAGCCAGCTATTGATTCCGGCAAGGGTCCGATGTTCACT CAGCCGAATCAGGCTGCTGGATACATGGCTAAGCTGATTTGGGAATCTGTGAGCGTGACGGTGGTAGCTGCGGTT GAAGCAATGAACTGGCTTAAGTCTGCTGCTAAGCTGCTGGCTGCTGAGGTCAAAGATAAGAAGACTGGAGAGATT CTTCGCAAGCGTTGCGCTGTGCATTGGGTAACTCCTGATGGTTTCCCTGTGTGGCAGGAATACAAGAAGCCTATT CAGACGCGCTTGAACCTGATGTTCCTCGGTCAGTTCCGCTTACAGCCTACCATTAACACCAACAAAGATAGCGAGA TTGATGCACACAAACAGGAGTCTGGTATCGCTCCTAACTTTGTACACAGCCAAGACGGTAGCCACCTTCGTAAGAC TGTAGTGTGGGCACACGAGAAGTACGGAATCGAATCTTTTGCACTGATTCACGACTCCTTCGGTACCATTCCGGC TGACGCTGCGAACCTGTTCAAAGCAGTGCGCGAAACTATGGTTGACACATATGAGTCTTGTGATGTACTGGCTGA TTTCTACGACCAGTTCGCTGACCAGTTGCACGAGTCTCAATTGGACAAAATGCCAGCACTTCCGGCTAAAGGTAA CTTGAACCTCCGTGACATCTTAGAGTCGGACTTCGCGTTCGCGTAA
GFPmut3b	Optimised Green Fluorescent Protein	ATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCAC AAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTG GAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCGAGATACCCAGA TCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAA GATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAA GGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACA TCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTC AACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTC CACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGC TGGGATTACACATGGCATGGATGAACTATACAAATAATAA
XFP	GFPmut3b with Y66H mutation	ATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCAC AAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTG GAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTCGGTCATGGTGTTCAATGCTTTGCGAGATACCCAG ATCATATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAA AGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAA AGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATAC ATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTT CAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGT CCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTG CTGGGATTACACATGGCATGGATGAACTATACAAATAA
LasR	Transcriptional activator from *P. aeruginosa*	ATGGCCTTGGTTGACGGTTTTCTTGAGCTGGAACGCTCAAGTGGAAAATTGGAGTGGAGCGCCATCCTCCAGAA GATGGCGAGCGACCTTGGATTCTCGAAGATCCTGTTCGGCCTGTTGCCTAAGGACAGCCAGGACTACGAGAACGC CTTCATCGTCGGCAACTACCCGGCCGCCTGGCGCGAGCATTACGACCGGGCTGGCTACGCGCGGGTCGACCCGAC GGTCAGTCACTGTACCCAGAGCGTACTGCCGATTTTCTGGGAACCGTCCATCTACCAGACGCGAAAGCAGCACGA GTTCTTCGAGGAAGCCTCGGCCGCCGGCCTGGTGTATGGGCTGACCATGCCGCTGCATGGTGCTCGCGGCGAAC TCGGCGCGCTGAGCCTCAGCGTGGAAGCGGAAAACCGGGCCGAGGCCAACCGTTTCATAGAGTCGGTCCTGCCG ACCCTGTGGATGCTCAAGGACTACGCACTGCAAAGCGGTGCCGGACTGGCCTTCGAACATCCGGTCAGCAAACCG

		GTGGTTCTGACCAGCCGGGAGAAGGAAGTGTTGCAGTGGTGCGCCATCGGCAAGACCAGTTGGGAGATATCGGT TATCTGCAACTGCTCGGAAGCCAATGTGAACTTCCATATGGGAAATATTCGGCGGAAGTTCGGTGTGACCTCCCG CCGCGTAGCGGCCATTATGGCCGTTAATTTGGGTCTTATTACTCTCTAATAA
P_{Las}	AHL- and LasR-dependent promoter	GACGCGGTTGCCCAAGCCGTTTCCGAAGGTGCCTCCTGAGCAGTTCAGATAGGTTGTCCGACCCCGAGTGGCAC GGCAGAGTGCCA
HrpR	Cooperative activator from *P. syringae*	ATGAGTACAGGCATCGATAAGGACGTCCGAGAGTGTTGGGGCGTAACTGCATTATCAGCGGGTCATCAAATTGCA ATGAATAGCGCGTTTCTGGATATGGACTTGCTGTTGTGCGGGGAAACCGGCACCGGCAAGGACACACTGGCCAAC CGCATTCACGAGTTGTCCAGCAGGTCGGGACCCTTTGTGGGCATGAACTGCGCCGCCATTCCCGAGTCGCTGGCA GAGAGCCAGTTATTCGGTGTGGTCAACGGTGCATTCACCGGCGTATGCCGGGCTCGCGAGGGCTACATAGAGGCC TCCAGTGGTGGCACCTTGTACCTGGATGAAATCGACAGCATGCCGTTGAGCCTGCAAGCCAAACTGCTGCGTGTG TTGGAGAGTCGAGGTATCGAGCGTCTGGGCTCGACCGAATTTATCCCGGTGGATCTGCGGATCATTGCCTCGGCC CAGCGGCCACTGGATGAACTGGTGGAACAAGGACTTTTCCGTCGCGACCTGTTTTTTCGGCTCAACGTGCTGAC GCTTCACTTGCCAGCCTTGCGCAAACGTCGTGAACAGATCCTGCCATTGTTCGACCAGTTCACCCAGGGTATCGC TGCCGAGTTCGGACGTCCCGCTCCTGCGCTGGACAGCGGGCGTGTGCAGCTGCTGCTCAGCCACGACTGGCCGG GCAACATCCGCGAATTGAAGTCTGCGGCCAAGCGCTTCGTACTCGGCTTCCCCTTGCTGGGCGCCGACCCTGTGG AAGCGCTTGACCCTGCCACGGGGCTGCGCACGCAAATGCGCATCATCGAGAAAATGCTCATCCAGGATGCCTTGA AGCGGCACAGGCACAATTTCGACGCGGTGCTTCAGGAGTTGGAGTTGCCAAGACGCACCCTGTATCACCGCATGA AGGAACTGGGAGTTGCAGCGCCGATCGCTGCGACGGCCGGGGTCTAA
HrpS	Cooperative activator from *P. syringae*	ATGAGTCTTGATGAAAGGTTTGAGGATGATCTGGACGAGGAGCGGGTTCCGAATCTGGGGATAGTTGCCGAAAG TATTTCGCAACTGGGTATCGACGTGCTGCTATCGGGTGAGACCGGCACGGGCAAAGACACGATTGCCCGACGGAT TCATGAGATGTCAGGCCGCAAAGGGCGCCTGGTGGCGATGAATTGCGCGGCCATTCCGGAGTCCCTCGCCGAGA GCGAGTTATTCGGCGTGGTCAGCGGTGCCTACACCGGCGCTGATCGCTCCAGAGTCGGTTATGTCGAAGCGGCGC AGGGCGGCACGCTGTACCTGGATGAGATCGATAGCATGCCGCTGAGCCTGCAAGCCAAATTGCTGAGGGTGCTGG AAACCCGAGCGCTTGAACGGCTGGGTTCGACGTCGACGATCAAGCTGGATATCTGCGTGATCGCCTCCGCCCAAT GCTCGCTGGACGACGCCGTCGAGCGGGGGCAGTTTCGTCGCGATCTGTATTTTCGCCTGAACGTCCTGACACTCA AGCTTCCTCCGCTACGTAACCAGTCTGATCGCATAGTTCCCCTGTTCACACGTTTTACGGCCGCCGCCGCGAGGG AGCTCGGTGTTCCCGTTCCCGATGTTTGCCCACTGCTGCACAAAGTGCTGCTGGGCCACGACTGGCCCGGCAATA TCCGTGAGCTCAAGGCGGCAGCCAAACGCCATGTGCTGGGTTTCCCCTTGCTGGGCGCCGAGCCGCAGGGCGAA GAGCACTTGGCCTGTGGGCTCAAATCGCAATTGCGAGTGATCGAAAAAGCCCTGATTCAGGAGTCGCTCAAGCG CCACGACAATTGTGTGGATTCGGTAAGCCTGGAACTGGACGTGCCACGCCGTACGCTCTATCGACGCATCAAAGA ATTGCAGATCTAA
P_{Hrp}	HrpR- and HrpS-dependent promoter	GCCGGATTATGTCCGCTGAGTGGGTCACGGTCCCGGATCAGTTCCCTTGCGAAGCTGACCGATGTTTTTGTGCCA AAAGCTGTTGTGGCAAAAAACGGTTTGCGCAAAGTTTTGTATTACAAAGAATTTCACATTTTAAAATATCTTTATA AATCAATCAGTTATTTCTATTTTTAAGCTGGCATGGTTATCGCTATAGGGCTTGTAC
kan^R	Kanamycin resistance	ATGAGCCATATTCAACGGGAAACGTCTTGCTCCAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGT ATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAG AGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGA CGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGAT CCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGT GTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAG GCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAA CAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTT GATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACC AGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGG TATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAA

Table 7: DNA sequences of parts and plasmids used in this work