Engineering Serine Integrase-based Synthetic Gene Circuits for Cellular Memory and Counting

Jia Zhao

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Institute of Molecular, Cell and Systems Biology

College of Medical, Veterinary and Life Sciences

University of Glasgow

Abstract

A cellular counting system based on synthetic gene circuits would enable complex biological programming and be used in many biotechnology applications. Although a variety of synthetic memory circuits have been constructed, basic modules that can be assembled into a counting system are lacking. This thesis focuses on engineering a binary counting module, which can alternate between two states in response to a single repeating input signal. The highly directional large serine bacteriophage integrases were utilised as the basis for the synthetic circuits constructed in this study. Integrases and their protein co-factors, the recombination directionality factor (RDF) can change the orientation of a specific DNA segment flanked by two recombination sites. Integrase alone switches the orientation in one direction, and this directionality is reversed by the addition of its corresponding RDF. The two orientations can be used to turn gene expression on and off, leading to distinct output states which can be thought of as representing a single binary digit (0 and 1) heritably stored in the DNA.

In this study, three different serine integrase-based synthetic gene circuits for cellular memory and counting were engineered and characterised. A set-reset latch was first constructed. By expressing ϕC31 integrase and co-expressing integrase with RDF Gp3 from two independent inducible systems, the orientation of the invertible DNA in the set-reset latch was inverted and restored respectively. This device demonstrated that ϕC31 integrase can successfully encode information into plasmid DNA. Next, a state-based latch was constructed, in which the *gp3* gene was placed inside the invertible DNA segment to couple its transcriptional regulation to the circuit state. Integrase expression triggered by one input signal resulted in inversion of the invertible DNA, placing the *gp3* gene in the correct orientation for transcription. Gp3 expression can then be triggered by another input signal to reverse the directionality of integrase, restoring the DNA back to its original configuration. By optimising the stoichiometry and kinetics of integrase and Gp3 expression, efficient switching of both multi-copy plasmid and single copy chromosomal DNA was achieved. Finally, the state-based latch was developed into a binary counting module by introducing a delay mechanism, in which *gp3* transcription was inhibited by a state-based repressor during recombination requiring the absence of Gp3. Placing expression of *gp3* under the control of the invertible DNA, allowed a single input signal controlling only integrase expression to switch the module between OFF (0) and ON (1). This is the first integrase-based module that generates different outputs in response to the same input signal and a fundamental step towards building a genetic binary counter with large counting capacity.

Table of Contents

Abstract 2
Table of Contents 3
List of Tables 8
List of Figures 9
Acknowledgements 13
Author's Declaration 14
Abbreviations 15
1 Introduction 16
1.1 Introduction to synthetic biology 16
1.1.1 Genetic parts, devices, and systems 16
1.1.2 Development of platform technologies 17
1.2 Synthetic gene circuits for cellular computation and memory 18
1.2.1 Engineering genetic switches for memory 18
1.2.2 Engineering synthetic oscillators for temporal control 24
1.2.3 Engineering genetic logic gates for processing multiple signals 26
1.2.4 Engineering genetic counting systems and their development 30
1.3 Introduction to genetic recombination 35
1.3.1 Homologous recombination 36
1.3.2 Site-specific recombination 36
1.3.3 Transpositional recombination (Transposition) 37
1.4 Classification of site-specific recombinases: tyrosine vs. serine recombinases 38
1.4.1 Tyrosine recombinases 39
1.4.2 Serine recombinases 40
1.4.3 Biological function of site-specific recombinases 40
1.5 Integrases and their applications 41
1.5.1 Tyrosine integrases 42
1.5.2 Applications of Tyrosine integrases 43
1.5.3 Serine integrases 44
1.5.4 Applications of serine integrases 46
1.6 Synthetic biological devices based on recombinases 47
1.7 Aims of the study and thesis outline 48
2 Materials and Methods 52
2.1 Bacterial Strains 52
2.2 Chemicals and buffer solutions 52

2.2.1 Chemicals....52
2.2.2 Buffer solutions....53
2.3 Antibiotics....54
2.4 Repressors and Inducers....54
2.5 Oligonucleotides....54
2.6 Custom DNA synthesis....59
2.7 Plasmids....59
2.8 Bacterial growth medium....64
2.9 Bacterial growth conditions....64
2.10 Generalised transduction in *E.coli* by bacteriophage P1....64
2.10.1 Preparation of P1 lysates....64
2.10.2 P1 transduction....65
2.11 Preparation of competent cells and transformation....65
2.11.1 Chemical transformation....65
2.11.2 Electroporation....66
2.12 Conjugation....67
2.13 Transposition....67
2.14 Plasmid DNA preparation....67
2.15 Gel Electrophoresis....67
2.15.1 Agarose gel electrophoresis....68
2.15.2 Denaturing polyacrylamide gel electrophoresis....68
2.16 Extraction of DNA from gels....68
2.16.1 Extraction of DNA from agarose gels....69
2.16.2 Extraction of DNA from polyacrylamide gels....69
2.17 Ethanol precipitation of DNA....69
2.18 Concentration measurement....70
2.19 Restriction endonuclease digestion of DNA....70
2.20 Ligation....70
2.21 Oligonucleotide insertion....70
2.22 Polymerase chain reaction (PCR)....70
2.23 TOPO TA cloning....71
2.24 Site-Directed mutagenesis....72
2.25 Sequencing....72
2.26 *In vitro* recombination reactions....72
2.27 Bacterial induction conditions....73

2.27.1 Overnight induction 73
2.27.2 Pulsed induction 73
2.27.3 Time course of induction recombination 74
2.27.3.1 Continuous induction time course 74
2.27.3.2 Short-pulse induction time course 74
2.28 Cell fixation 75
2.29 Cell fluorescence measurement 75
2.29.1 Typhoon Scanner 75
2.29.2 Blue light Transilluminator 75
2.29.3 Flow cytometry 75
2.29.4 Fluorescent microscopy 75
2.30 MacConkey assay 76
2.31 Production of substrate containing *attR* and *attL* 76
2.32 Random library construction and selection 77
2.33 Quantification by Quantity One 77
3 Preliminary characterisation and testing of ϕC31 integrase inversion recombination system 78
3.1 Introduction 78
3.2 Investigation of the inversion reaction mediated by ϕC31 integrase *in vitro* 79
3.2.1 Effects of ϕC31 integrase and Gp3 concentrations on recombination efficiency *in vitro* 80
3.2.1.1 Reactions with substrate plasmid containing *attB* and *attP* 80
3.2.1.2 Reactions with substrate plasmid containing *attR* and *attL* sites 81
3.2.2 Effects of reaction times on recombination efficiency 83
3.3 Investigation of the inversion recombination mediated by ϕC31 integrase *in vivo* 85
3.3.1 Construction and characteristics of a tightly regulated ϕC31 integrase expression system 85
3.3.2 Inversion recombination reaction with high copy-number plasmid substrates *in vivo* 88
3.3.2.1 Investigations on the *attB*×*attP* inversion reaction with the substrate plasmid pZJ16off 88
3.3.2.2 Investigations of *attB*×*attP* inversion reactions with substrate plasmid pZJ18off containing a *cer* recombination site 92
3.3.2.3 Investigations on the *attR*×*attL* inversion reaction with substrate plasmid pZJ18on containing a *cer* recombination site 95
3.3.2.4 Effects of Int expression levels on inversion reaction efficiency 97
3.3.3 Inversion recombination reaction with low copy-number plasmid substrates *in vivo* 100

3.3.3.1 Investigations on the inversion reaction mediated by Int with different expression levels 100
3.3.3.2 Effects of reaction time on inversion reaction efficiency 103
3.4 Conclusion and discussion 104
4 Design and engineering of plasmid-borne memory devices 107
4.1 Introduction 107
4.2 Engineering of the genetic set-reset latch 108
4.2.1 Design and construction of the genetic set-reset latch 108
4.2.2 Investigation of the recombination efficiency with the set-reset latch 110
4.3 Engineering of the state-based latch (two-signal controlled system) 114
4.3.1 Design and construction of the state-based latch 114
4.3.2 Characterisation of the state-based latch 120
4.3.2.1 Process of continuous time induction reaction 120
4.3.2.2 Reactions with different lengths of induction time 121
4.3.2.3 Reactions with different Int expression plasmids 122
4.3.2.4 Reactions in different strains 125
4.3.3 Further optimisation of the state-based latch by changing the expression levels of RDF gp3 127
4.3.4 Reliability test of the optimised state-based latch under multicycle set-reset operation using two different input signals 131
4.3.5 Test of the state-based latch using Bxb1 integrase instead of ϕC31 integrase . 133
4.4 Engineering of the binary counting module (single-signal controlled system) 135
4.4.1 Developing the state-based latch into a binary counting module to be controlled efficiently in strain DS941 135
4.4.2 Optimisation of the binary counting module by changing the expression levels of RDF gp3 138
4.4.3 Characterisation of the optimised binary counting module to determine the optimal induction time 140
4.4.4 Reliability test of the binary counting module under multicycle switch-on switch-off operation using a single input signal 144
4.5 Conclusion and discussion 148
4.5.1 Set-reset latch 149
4.5.2 State-based latch 150
4.5.3 Binary counting module 151
5 The design and engineering of chromosomal memory devices 155
5.1 Introduction 155
5.2 Chromosomal delivery of memory devices using ISY100 transposition 156
5.3 Methods used to analyse the performance of the chromosomal device 158

5.4 Investigation of factors affecting recombination efficiencies of chromosomal state-based latch 163

5.4.1 Effects of induction pulse length on recombination efficiencies 163

5.4.2 Effects of Int expression level and degradation rate on recombination efficiencies 165

5.5 Reliability test of the chromosomal state-based latch under multicycle set-reset operation 169

5.6 Test of the chromosomal devices with a single signal 171

5.7 Conclusion and discussion 177

5.7.1 Transposons as a tool to deliver a single copy substrate sequence into the *E.coli* chromosome 177

5.7.2 Efficient and reliable recombination reaction with the chromosomal state-based latch (two signals controlled system) 178

5.7.3 Ways to improve the single-signal controlled systems 180

6 Final discussion 183

6.1 Primary data suggested ϕC31 integrase was a good candidate for construction of genetic devices 185

6.2 Plasmid-borne devices behaved efficiently 186

6.2.1 The set-reset latch proved that ϕC31 integrase could efficiently set and reset information in DNA 187

6.2.2 The state-based latch demonstrated that expression of RDF could be controlled by the DNA inversion state 187

6.2.3 The binary counting module realised the goal of altering outputs using a single repeating input signal 188

6.3 Making single copy memory devices 189

References 193

List of Tables

Table 1. 1 The classification and function of recombinases mentioned in this study.....41
Table 2. 1 Bacterial Strains.....52
Table 2. 2 Chemicals.....52
Table 2. 3 Buffer Solutions.....53
Table 2. 4 Antibiotics.....54
Table 2. 5 Inducers and repressors.....54
Table 2. 6 Oligonucleotides designed in this study.....55
Table 2. 7 Plasmids used and constructed in this study.....60
Table 3. 1 Int expression plasmids.....98
Table 4. 1 Predicted promoters in *Gp3* and *GFP*.....120
Table 4. 2 Effect of gp3 RBS on reaction efficiencies.....130
Table 4. 3 Induction times and recombination efficiencies for different memory devices.....149
Table 5. 1 Int expression plasmids and recombination performance abtained.....168
Table 5. 2 Recombination efficiencies of chromosomal devices controlled by a single signal.....176
Table 5. 3 Comparison between the state-based latch and the rewritable recombinase addressable data module.....179

List of Figures

Figure 1. 1 Synthetic switches based on transcriptional regulators. 20
Figure 1. 2 Synthetic switches based on recombinases. 23
Figure 1. 3 Synthetic oscillators. 25
Figure 1. 4 Conventional symbols and truth tables of basic logic gates. 26
Figure 1. 5 Synthetic logic gates based on transcriptional regulation and RNA regulation 28
Figure 1. 6 Synthetic logic gates based on recombinase 30
Figure 1. 7 Synthetic counters. 32
Figure 1. 8 A synthetic push-on push-off switch. 33
Figure 1. 9 A synthetic toggle flip-flop. 34
Figure 1. 10 Outcomes from site-specific recombination in circular molecules. 37
Figure 1. 11 Mechanism of recombination by tyrosine recombinases. 39
Figure 1. 12 Mechanism of recombination by serine recombinases. 40
Figure 1. 13 Mechanism of integration and excision catalysed by phage integrase. 42
Figure 1. 14 Recently proposed model of the integration reaction catalysed by serine integrases. 46
Figure 1. 15 A binary counter based on DNA inversion 50
Figure 2. 1 Maps of synthetic plasmids. 59
Figure 2. 2 Flow diagram showing the overnight induction process. 73
Figure 2. 3 Flow diagram showing the short pulse induction process. 74
Figure 3. 1 Mechanism of ϕC31 integration, excision and inversion reaction. 79
Figure 3. 2 Expected change of substrate plasmids with the inversion recombination reaction. 80
Figure 3. 3 *attB*×*attP in vitro* reaction. 81
Figure 3. 4 *attR*×*attL in vitro* reaction. 82
Figure 3. 5 Kinetics of the *in vitro* inversion reaction. 84
Figure 3. 6 The MacConkey recombination assay. 86
Figure 3. 7 MacConkey recombination assay used to test the ϕC31 integrase expression system. 87
Figure 3. 8 Inversion reaction between *attP* and *attB*. 89
Figure 3. 9 Intramolecular and intermolecular recombination. 91
Figure 3. 10 Combination of integrase site-specific recombination and cer recombination to convert multimers to monomers. 93
Figure 3. 11 Comparison of the recombination between substrates without and with cer recombination site. 94
Figure 3. 12 Inversion reaction between *attR* and *attL*. 96
Figure 3. 13 Process of changing the RBS sequence for Int. 98
Figure 3. 14 Effects of different Int expression levels on *attB*×*attP* and *attR*×*attL* recombination on high copy number substrate plasmids. 99
Figure 3. 15 Effects of different Int expression levels on *attB*×*attP* and *attR*×*attL* recombination on low copy number substrate plasmids. 102
Figure 3. 16 Electrophoretic analysis of the *attB*×*attP* time course reaction. 103
Figure 4. 1 Diagram showing the design principle of set-reset latch. 109

Figure 4. 2 Electrophoretic analysis of the set-reset latch in the single directional reaction.. 111
Figure 4. 3 Electrophoretic analysis of the genetic set-reset latch in a full cycle of set-reset reaction. 113
Figure 4. 4 Diagram showing the design principle of the state-based latch. 115
Figure 4. 5 Optimisation of the data storage element of the state-based latch.....117
Figure 4. 6 Electrophoretic analysis of the state-based latch under continuous induction time course. 121
Figure 4. 7 Electrophoretic analysis of the state-based latch under pulsed induction time course. 122
Figure 4. 8 Set and reset reaction of the state-based latch with different expression levels of Int. 124
Figure 4. 9 Performance of state-based latch in different strains. 126
Figure 4. 10 Process of seeking optimal Gp3 expression levels in state-based latch. 128
Figure 4. 11 Fluorescent analysis of cells containing substrate with different Gp3 RBSs under set and reset reactions. 129
Figure 4. 12 Electrophoretic analysis of cells containing mutational state-based latch in a full cycle of set-reset reaction. 130
Figure 4. 13 Five cycles of set-reset operation with the optimised state-based latch. 132
Figure 4. 14 Reset or Set operation with the state-based latch based on Bxb1 integrase. .134
Figure 4. 15 The switch-on reaction of state-based latch in strain DS941.. 136
Figure 4. 16 The switch-on and switch-off reactions of the new device in strain DS941. 137
Figure 4. 17 Fluorescent analysis of cells containing substrate with different RBSs for Gp3 under set and reset reactions. 139
Figure 4. 18 Pulsed induction time course with the binary counting module by adding input signal after different preculture length.. 141
Figure 4. 19 Pulsed induction time course of the binary counting module.....143
Figure 4. 20 Multicycle operations with the binary counting module. 145
Figure 4. 21 Fluorescent analysis of cell population containing the binary counting module in multicycle reactions. 147
Figure 4. 22 Expected performance of the binary counting module with single induction pulse. 152
Figure 5. 1 Process of integrating the invertible substrate sequence into the chromosome. 158
Figure 5. 2 Methods used to check the recombination performance of the chromosomal device. 161
Figure 5. 3 Flow cytometry analysis of the chromosomal state-based latch under pulsed induction time course. 164
Figure 5. 4 Recombination efficiencies with different Int expression and degradation conditions. 167
Figure 5. 5 Five cycles of set-reset operation with chromosomal state-based latch. 170
Figure 5. 6 Pulsed induction time course with chromosomal state-based latch (without time delay from TetR) using a single repeating signal.....173
Figure 5. 7 Pulsed induction time course with chromosomal binary counting module (with time delay from TetR) using a single repeating signal 174

Figure 5. 8 Recombination process with chromosomal single signal controlled module and the equation used to estimate switch-on and switch-off reaction efficiencies. 175
Figure 5. 9 Predicted performance of modules with different reaction efficiencies.. 182
Figure 6. 1 Summary of the results presented in previous result chapters.....184
Figure 6. 2 A schematic diagram showing the hypothetical architecture of a two bit binary counter.....192

This thesis is dedicated with all my love to my parents

Shizhong Zhao and *Sumei Wang*

Acknowledgements

I would like to thank so many people for supporting and helping me over the past four years. Special thanks go to my parents for encouraging me to pursue my dream and their unreserved support. Thank you to Qingnan who plays a fabulous role in my life, for his patience, understanding, and constant support. None of this could have happened without their love.

I would also like to thank those people who have contributed to the progress of my project. Great thanks go to Prof. Marshall Stark for his valuable advice and generously offering me lab resources; Dr. Femi Olorunniji for supplying me purified proteins and helping with the *in vitro* assay. Thank you to Ms. Diane Vaughan for the training of flow cytometry, and Yanqing Song for the help on fluorescent microscope. Thank you to Dr. Martin Boocock, Dr. Alexandra Pokhilko, and Dr. Sally-Jane Rowland for all those interesting conversations in the office. I would like to thank people who have worked with me in Bower Building and helped me in many aspects: Dr. Mai-Britt Jensen, Miss Arlene McPherson, Dr. Steven Kane, Dr. Emanuele Conte and many others. Thank you to Prof. Gareth Jenkins and Prof. William Cushley for reviewing my work and providing help when I was in a difficult situation. Thank you to Prof. Geoffrey Baldwin and Prof. John Christie for examining my work and their suggestions for the future work.

Thank you to my amazing friend Dr. Menglin Cao for lighting my way whenever I felt confused. I am grateful to Dr. Christine Merrick for her help with my work and life, and her family for giving me a special memory of Christmas; Dr. Steph Holt for bringing me lots joy and her enormous contribution to the proofreading of this thesis. Thank you to Bin Jia for helpful discussion and suggestion. Special thanks go to Dr. Ben Zhang for giving me great inspiration and support, especially during my thesis writing, the toughest period in my life. My acknowledgements also go to my friends Min Hou, Carla Minguet, Johannes Mayer, Kevin Crawford, and Yujing Zhao whom I spent a wonderful time with in Glasgow.

Finally, I would like to express the great appreciation to my supervisors, Prof. Susan Rosser Dr. Sean Colloms, for giving me such a precious PhD experience. Thank you to Susan for offering me the opportunity to access to the cutting-edge research and work on such an exciting project, and her support and encouragement all along. Thank you to Sean for guiding me into the world of science, and his valuable suggestions and tremendous patience whenever I felt difficult to proceed. Thank you very much.

Author's Declaration

"I declare that the research presented in this thesis is my own work except where stated and it has not been submitted for any other degree".

Jia Zhao

1st September 2015

Abbreviations

AHL	acyl-homoserine lactone	NRI~P	phosphorylated NRI
APS	ammonium persulphate	PAGE	polyacrylamide gel electrophoresis
ara	arabinose	PCR	polymerase chain reaction
aTc	anhydrotetracycline	RBS	ribosome binding site
BAC	bacterial artificial chromosome	RD	recombinase domain
BP	*attB* and *attP*	RDF	recombination directionality factor
BSA	bovine serum albumin	RFP	red fluorescent protein
CC	coiled-coil motif	RL	*attR* and *attL*
CTD	C-terminal DNA binding domain	RNA	ribonucleic acid
DAP	diaminopimelic acid	RNAP	RNA polymerase
DTT	dithiothreitol	SDS	sodium dodecyl sulphate
dNTP	deoxynucleotide triphosphate	SIMM	single invertase memory module
EDTA	ethylenediaminetetraacetic acid	SKE	SDS/protease K/EDTA
Fis	factor for inversion stimulation	SSC	side scatter
FSC	forward scatter	TAE	Tris-acetate-EDTA
GFP	green fluorescent protein	taRNA	transactivating noncoding RNA
IDB	Integrase dilution buffer	TBE	Tris boric acid EDTA
IHF	integration host factor	TEMED	tetramethylethylenediamine
IPTG	isopropyl β-D-1-thiogalactopyranoside	TF	transcriptional factor
IRB	integrase reaction buffer	Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
IS	insertion sequence	tRNA	transfer RNA
L-Broth	luria broth	Xis	excisionase
NTD	N-terminal catalytic domain	YFP	yellow fluorescent protein
NRI	nitrogen regulator I	ZD	zinc-ribbon domain

1 Introduction

1.1 Introduction to synthetic biology

Synthetic biology has emerged as a discipline combining biology with engineering principles over the past decade (Andrianantoandro *et al*., 2006; Cheng and Lu, 2012). With the rapid development of synthetic biology, several definitions have been used to describe this discipline. The commonly accepted definition of synthetic biology is that it "involves the design and construction of novel biological parts, devices and systems, and the redesign of existing natural biological systems for useful purposes" (Roberts *et al*., 2013). Synthetic biology utilises abstraction, modularization and standardization concepts from engineering to design biological systems, and builds up these systems using more efficient methods (such as DNA synthesis and DNA assembly). The goal of synthetic biology is to engineer the behaviours of organisms for many different applications, such as processing information and producing useful products (for instance biofuels, pharmaceuticals, and materials), meanwhile, deepening the understanding of biological processes.

1.1.1 Genetic parts, devices, and systems

One of the primary goals of synthetic biology is to construct novel biological systems from characterised parts. Parts can be assembled to create devices or modules, then multiple devices can be connected to form systems, and systems can be further developed into large networks and synthetic chromosomes (Cheng and Lu, 2012).

A growing number of parts have been characterised, such as promoters (Davis *et al*., 2011; Kelly *et al*., 2009), terminators (Chen *et al*., 2013), ribosome binding sites (RBSs, Salis *et al*., 2009), and protein degradation tags (Andersen *et al*., 1998; Cameron and Collins, 2014). These parts perform specific functions in various biological processes (transcription and translation of genes, post transcriptional processes, and degradation of proteins), and the behaviours of these processes can be regulated via tuning the parameters of corresponding parts (Arpino *et al*., 2013).

By using well-characterised parts, a number of higher-order devices have been constructed, such as switches and memory elements (Jerome Bonnet *et al.*, 2012; Gardner *et al.*, 2000; Ham and Lee, 2006), oscillators (Atkinson *et al*., 2003; Danino *et al*., 2010; Elowitz and

Leibler, 2000; Stricker *et al*., 2008), logic gates (Bonnet *et al*., 2013; Elowitz *et al*., 2002; Moon *et al*., 2012; Siuti *et al*., 2013; Wang *et al*., 2011), and counters (Friedland *et al*., 2009).

As the number of available synthetic devices has increased, more sophisticated systems have been created by assembling several devices together, or linking devices with cellular activities to execute useful human-defined functions. For example, a synthetic program containing a genetic multiplexer can be used to control the activation of two chemosensory pathways (Moon *et al*., 2011), and synthetic RNA regulatory systems linked to cell growth cytokine targets can be used to control cell proliferation (Chen *et al*., 2010).

1.1.2 Development of platform technologies

To achieve the goal of efficient construction of genetic systems, platform technologies need to be developed. The term "platform technology" refers to a suite of tools and methods which can be applied in the field of synthetic biology (Kitney and Freemont, 2012). Platform technologies include methods for part characterisation, host cells, DNA assembly, and web-based information systems. All these aspects are introduced in more detail below.

As mentioned in section 1.1.1, in order to reliably engineer devices or systems from genetic parts, it is necessary to characterise the properties and functional behaviours of parts (Canton *et al*., 2008). To do this, standardized tools, techniques, and units of measurement are needed to facilitate the characterisation of parts by independent groups (Kelly *et al*., 2009). Detailed knowledge of how different genetic parts and devices contribute to phenotypes of host cells, such as growth characteristics, differences between single and population cell behaviours, and metabolic burden, needs to be obtained to predict their performance in different cellular contexts.

As more complex systems need to be assembled from multiple genetic parts, DNA assembly is becoming a key platform technology for synthetic biology (Ellis *et al*., 2011). A number of techniques of DNA assembly have been developed, including 1) restriction and ligation-based methods, such as BioBricks™ (Knight, 2003) and Golden Gate (lab of Sylvestre Marillonnet, Icon Genetics GmbH, Germany); 2) homology-based approaches, such as sequence and ligation-independent cloning (Li and Elledge, 2007), Gibson Assembly (Gibson *et al*., 2009), and yeast *in vivo* DNA assembly (Shao *et al*., 2009), and 3) site-

specific recombination-based assembly, such as site-specific recombination-based tandem assembly (Colloms *et al.*, 2013; Zhang *et al.*, 2011) and Gateway (Kirchmaier *et al*., 2013).

Web-based information systems are another type of platform technology that will be necessary in scaling up the complexity of synthetic biology designs. Synthetic Biology Open Language is an example of such a system (Galdzicki *et al*., 2014). These systems should incorporate information about the part characterisations, host cells, DNA assembly methods, models, and software tools (Kitney and Freemont, 2012).

1.2 Synthetic gene circuits for cellular computation and memory

As a main goal of synthetic biology, variations of synthetic gene circuits possessing human-defined functions have been constructed. These circuits endow cells with the ability to sense specific input signals, perform computations, commit to memory and respond with outputs, and have the potential to be used in a wide range of applications. For example, synthetic regulatory circuits can be used as intracellular sensing and control elements, providing novel strategies for disease diagnoses and therapies, and helping researchers understand more about biological systems in health and disease (Cheng and Lu, 2012).

This section focuses on the development of four types of synthetic gene circuits, which are genetic switches, oscillators, logic gates, and counters, named after their performances simulating electronic circuits. The limitation of current synthetic genetic counters and the significance of construction of an efficient counter are mainly discussed at the end of this section.

1.2.1 Engineering genetic switches for memory

In electrical engineering, a switch is used to control the current in an electric circuit. In a biological system, a genetic switch can be used to change the gene expression from one pattern to another in response to an environmental signal (Ptashne, 2004). Genetic switches have been found throughout natural systems, and widely used to support biotechnology applications (Bonnet and Endy, 2013). The simplest genetic switches are regulated promoters that can respond to specific signals to switch on and off gene expression. These promoters include the previously used promoter P_L and P_R (Elvin *et al*., 1990) and P_{LAC} (Yanisch-Perron *et al*., 1985), the later widely used promoter P_{Tet} (Skerra, 1994; Stanton *et*

al., 2014) and P_{BAD} (Guzman *et al*., 1995), as well as some of their derivatives (Lutz and Bujard, 1997).

Gene expression controlled by the regulated promoters mentioned above are transient. The state could only be maintained in the presence of the input signal, and the descendants of the cell have no memory that their ancestors had been exposed to any signal. Biological memory was defined as a sustained cellular response to a transient signal (Burrill and Silver, 2010). In practice, using regulated promoters that lack memory function as genetic switches has a number of limitations. For example, they cannot be used to detect transient signals, or control the expression of desired proteins after induction by a brief input. Compared to simple regulated promoters, the synthetic switches introduced in this section are more advanced in displaying memory. Typical examples of synthetic switches have been constructed by combining transcriptional regulation and feedback, or by using recombinases to record information into the DNA sequence.

An early example of a synthetic switch for memory is the toggle switch containing two repressible promoters that drive the expression of mutually inhibitory repressors (Fig. 1.1A, Gardner *et al*., 2000). In this toggle switch, each repressor negatively regulates the synthesis of the other repressor, thereby, creating two stable states. By addition of transient chemical or thermal induction, one repressor is inactivated selectively, switching the state of the device. This switch was demonstrated to retain state after removing the input signals.

In contrast to this early toggle switch, which only relied on negative feedback, subsequent switchable systems introduced positive feedback. In one example, a genetic switch containing both negative and positive feedback modules was constructed (Fig. 1.1B, Atkinson *et al*., 2003). The core element of this switch is a dual-input promoter $P_{lac\text{-}glnAp2}$, which is repressed by LacI (under the control of a constitutive promoter) and induced by NRI~P (phosphorylated form of glnG encoded nitrogen regulation protein NRI), and is set to drive the expression of NRI. NRI can induce its own expression from $P_{lac\text{-}glnAp2}$ promoter. There is a competition between the LacI and NRI~P on the $P_{lac\text{-}glnAp2}$ promoter. Switching of this circuit can be determined by controlling the level of effective LacI through varying the concentration of β-D-1-thiogalactopyranoside (IPTG), and is observed via the glutamine synthetase level produced from *glnA* gene (on the chromosome) driven by the toggle switch.

Later, a memory device in yeast based on autoregulatory transcriptional positive feedback was designed and constructed (Fig. 1.1C, Ajo-Franklin *et al*., 2007). The system consists of two elements, which are the "signal sensor" and the "autofeedback memory". A signal induces the "signal sensor" to express a transcription factor which triggers the expression of the same transcription factor in the "autofeedback memory". Then, the transcription factor starts to activate its own promoter even in the absence of input signal, resulting in memory. The fluorescent proteins RFP and YFP are used to label the "signal sensor" and the "autofeedback memory" respectively, and galactose is used as the input signal.

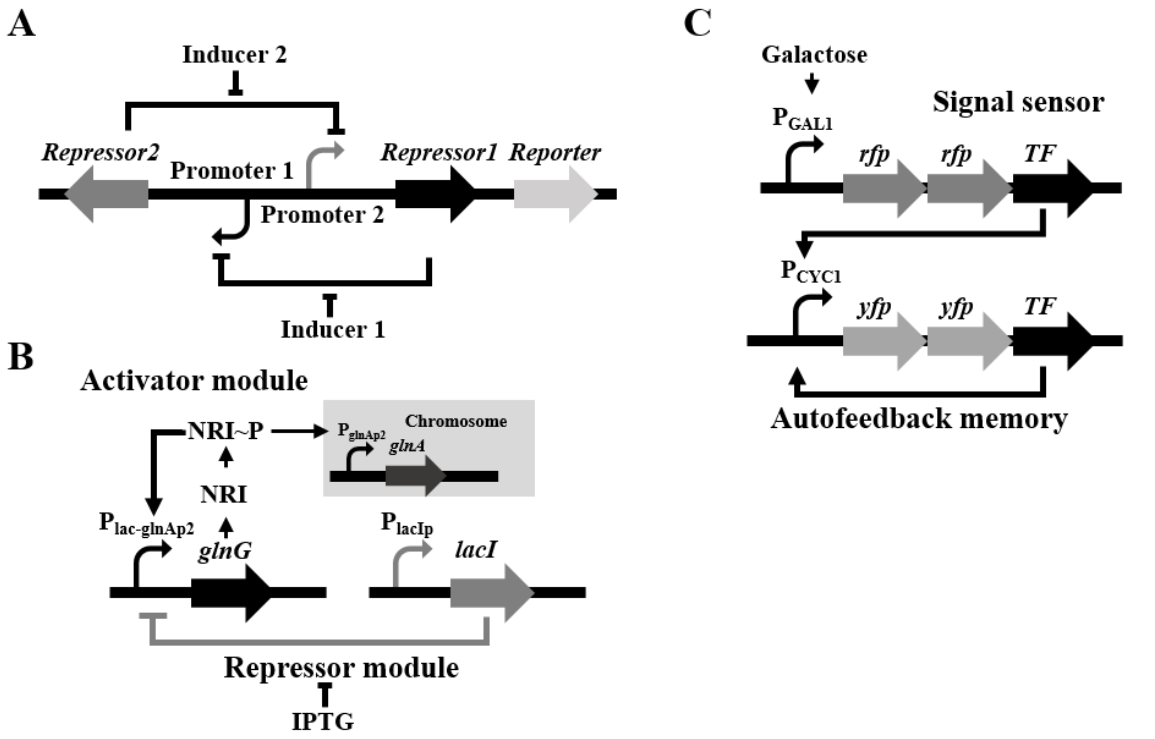


Figure 1. 1 Synthetic switches based on transcriptional regulators. (A) The toggle switch. Two repressible promoters are arranged to express their mutually acting repressors, resulting in bistable states in which only one of the two promoters is active at a given time. It can be flipped to the desired state using environmental inputs to disrupt the inhibition of the promoters. Adapted from Gardner *et al.* (2000). (B) The positive/negative feedback loop switch. One unit contains the dual-input promoter $P_{lac\text{-}glnAp2}$ fused to *glnG* (gene encoding NRI). The competition between the activator, NRI~P, and the repressor, LacI, on this promoter will determine the state of this device. The LacI is produced from a constitutive promoter in another unit, and its effect on the $P_{lac\text{-}glnAp2}$ can be controlled by changing the concentration of IPTG. The state was observed through glutamine synthetase level produced from *glnA* gene on the chromosome. Adapted from Atkinson *et al.* (2003). (C) The autoregulatory feedback loop switch. In the "signal sensor" unit, the expression of a transcriptional factor (TF) is controlled by an inducible promoter $P_{GAL1/10}$. In the "autofeedback memory" unit, the transcriptional factor can activate its own promoter, resulting in autoregulatory expression even in the absence of stimulus. Both present stimulus and previous exposure to stimulus will switch the states of the synthetic network from inactive to active. Adapted from Ajo-Franklin *et al.* (2007).

Another method to make genetic switches capable of memory is to encode memory into DNA sequence using the mechanism of site-specific recombination. Site-specific recombinases can recognise specific sequences (recombination sites) within a piece of DNA, catalyse recombination between these two sites, and flip the orientation of the intervening DNA sequence if the two sites are placed in inverted repeat (Stark *et al*., 1992). By using an inducible promoter to control the expression of a recombinase and change the substrate DNA conformation, information from the environmental or an intracellular signal can be written into the DNA circuit.

The early recombinase addressable switches were based on invertases, which are highly specialised in catalysing inversion recombination between sites in inverted repeat, while deletions between sites in direct repeat and recombination between sites on different molecules occur at very low frequency (van de Putte and Goosen, 1992). One such example is based on FimE inversion recombination system. Invertase FimE naturally inverts a DNA switch to control the expression of type I fimbriae in the *fim* system of *Escherichia coli* (Holden *et al*., 2007). The synthetic switch was constructed by placing a constitutive promoter between inverted repeat recombination sites (IRR^{on} and IRL^{on}) for invertase FimE and a target gene was cloned outside of the recombination sites in an opposite orientation to the constitutive promoter (Fig. 1.2A). When the expression of invertase FimE is activated by adding inducer, the orientation of the promoter is flipped via FimE-catalysed DNA inversion, resulting in two new sites IRR^{off} and IRL^{off} and expression of the target gene. This FimE inversion switch is a successful demonstration of recording information into a DNA sequence, and shows leak-less properties and persistence of state after removal of the inducer or recombinase input (Ham and Lee, 2006).

The same group constructed a more complex inversion switch based on two invertases FimB and Hin (Fig. 1.2B, Ham *et al*., 2008). The FimB invertase acts in the same *fim* system as FimE to control the expression of fimbriae in *E. coli*, but can invert the switch in both directions; from on to off and from off to on (Holden *et al*., 2007). The Hin invertase catalyses inversion recombination to control the expression of the flagellar gene H2 and the repressor gene of the flagellar gene H1 in the Hin system of *Salmonella typhimurium* (van de Putte and Goosen, 1992). The synthetic switch was constructed by integrating FimB and Hin recombination systems, in which two inversion switches were placed on the same DNA molecule in an overlapping manner. When the expressions of FimB and Hin are induced by addition of arabinose and anhydrotetracycline respectively in different orders, this device

switched to different states, and the states can be detected by PCR with chosen primers. This switch is the first artificial system that integrated two inversion systems into a single circuit and demonstrated the history dependent configuration of recombinase-based memory devices.

The recombinase FimE used in the earlier synthetic switch (Fig. 1.2A, Ham and Lee, 2006) inverts the DNA segment overwhelmingly from on (IRR^{on} and IRL^{on}) to off (IRR^{off} and IRL^{off}) orientation, whereas the FimB and Hin used in the latter switch (Fig. 1.2B, Ham *et al*., 2008) inverts the DNA segment in both directions (Holden *et al*., 2007; van de Putte and Goosen, 1992). There is not any controlled mechanism for these recombinases to invert and restore the state of target DNA sequence, and their recombination usually ends up with a mixture of two different states, which might limit their information storage capacity.

As an alternative to these invertases, the stability and capacity of the recombinase-based switches can be improved by using highly directional large serine integrases, which are able to invert and restore the orientation of a DNA segment under tight control. Two large serine integrases (from *Streptomyces* phage ϕC31 and from *M. tuberculosis* phage Bxb1) have been proved to catalyse efficient recombination between sites *attP* and *attB* and generate *attR* and *attL* (Ghosh *et al*., 2003; Thorpe *et al*., 2000). Whereas, in the presence of a corresponding phage encoded excisionase (Xis) or recombination directionality factor (RDF), the *attP* and *attB* recombination is inhibited, and integrase catalyses recombination between *attR* and *attL* to regenerate *attP* and *attB* (Khaleel *et al*., 2011; Ghosh *et al*., 2006).

The successful application of a serine integrase to create a synthetic switch was demonstrated by the rewritable recombinase addressable data module (Fig. 1.2C, Bonnet *et al*., 2012). In this device, a constitutive promoter was placed between recombination sites for Bxb1 integrase, facing towards one reporter gene (*gfp*) and away from another (*rfp*). The gene encoding Bxb1 integrase is under the control of a $P_{Ltet0\text{-}1}$ promoter so that its expression can be controlled by addition of aTc. Separately, on a different plasmid, the RDF gene, *gp47* together with Bxb1 integrase gene are under the control of a P_{BAD} promoter so that their expression can be activated by addition of arabinose. The unidirectional Bxb1 integrase catalyses inversion in one direction, flipping the promoter and changing reporter gene expression. With the assistance of the RDF gp47, Bxb1 integrase catalyses the reverse reaction, restoring the original reporter gene expression. It was demonstrated that this genetic switch flips repeatedly and reliably in response to alternate input signals aTc and arabinose

for multiple cycles. In addition, the switch is able to hold either state for approximately 100 cell doublings in the absence of any signal.

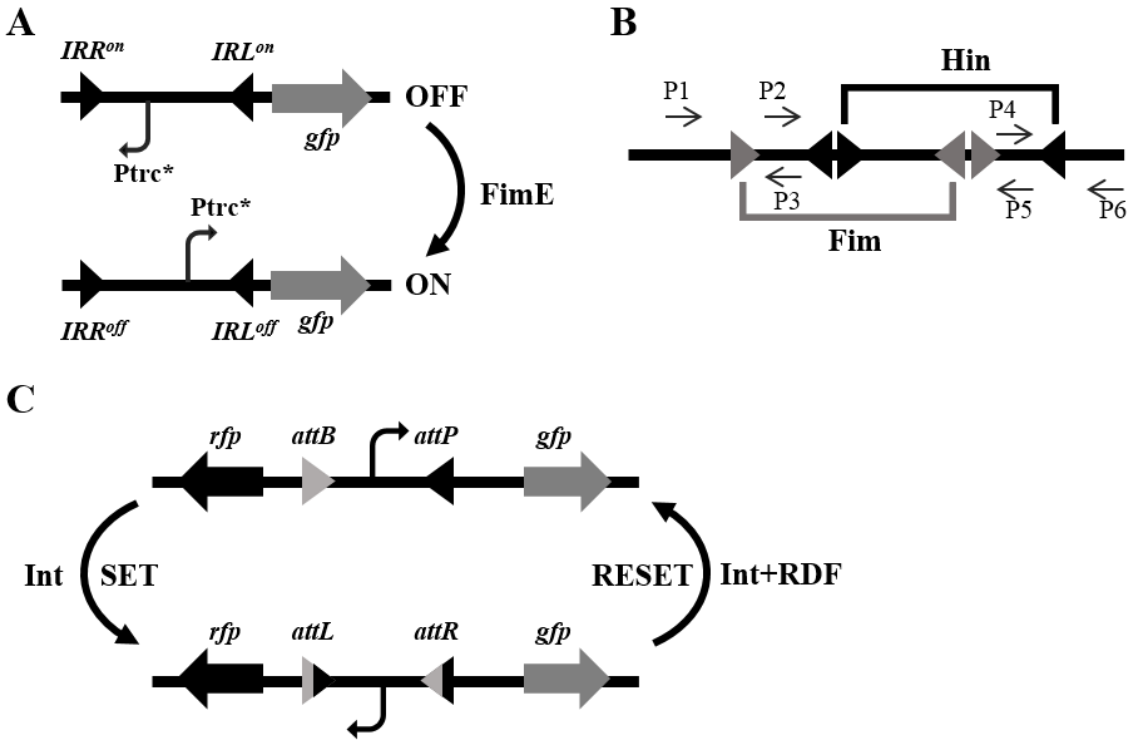


Figure 1. 2 Synthetic switches based on recombinases. (A) The FimE inversion recombination switch. FimE recognises inverted repeat sites IRR^{on} and IRL^{on}, and inverts the intervening constitutive promoter to the correct orientation for reporter gene expression. FimE expression is controlled by an inducible promoter. Adapted from Ham and Lee (2006). (B) The double inversion recombination switch. The inverted repeat sites with intervening constitutive promoter for FimB and Hin recombinase are placed in an overlapping manner. Several sets of primers (P1-P6) can be used to detect the order that FimB and Hin have been induced. Adapted from Ham *et al.* (2008). (C) The rewritable recombinase addressable data module. Bxb1 integrase recognises inverted repeat sites *attP* and *attB*, and recombines them to generate *attL* and *attR* sites. This inverts the intervening constitutive promoter, switching off expression of RFP and switching off expression of GFP. With the assistance of RDF gp47, Bxb1 integrase recognises inverted repeat sites *attL* and *attR*, and recombines them to generate *attB* and *attP* sites. This restores the intervening sequence to the original state. Adapted from Bonnet *et al.* (2012).

With the development of genetic circuits that can sense and remember environmental signals, some circuits have begun to be used for practical applications. One of the examples is the engineered *E.coli* which is able to detect and record an environment signal in the mammalian gut (Kotula *et al*., 2014). The memory device used in this system could be re-engineered to respond to disease-related signals (e.g. chemical signatures of inflammation, cancer, parasites, or environmental toxins in the gut) and produce diagnostic outputs (e.g. visual indicator) or therapeutic drugs (e.g. a specific antibacterial component), which may lead to the development of diagnostics and therapeutics.

1.2.2 Engineering synthetic oscillators for temporal control

Oscillations are omnipresent in living systems, controlling processes such as metabolism, signalling pathways, cardiac and circadian pacemakers, and learning processes (Aubel and Fussenegger, 2010). Inspired by natural devices controlling the oscillation processes of living cells, many different synthetic oscillators have been successfully constructed. Around the same time as the first synthetic toggle switch was reported, the first synthetic oscillator was constructed by arranging three repressor-operator modules in a cyclic negative feedback loop (Fig. 1.3A, Elowitz and Leibler, 2000). In this network, the lactose operon repressor protein, LacI from *E. coli*, inhibits transcription of the tetracycline repressor gene, *tetR* from the tetracycline-resistance transposon Tn10, whose gene product represses the transcription λ repressor *cI*. Finally, λ CI inhibits the expression of LacI. This network produces fluctuating levels of each repressor protein. The fluctuating can be visualised using the reporter protein GFP that is driven by the target promoter of a specific repressor in the loop (e.g. P_{TET}-*gfp*).

The oscillator introduced above was not robust and damped rapidly, which was thought to be a result of lack of positive feedback loops (Aubel and Fussenegger, 2010). The first synthetic oscillator combining positive and negative feedback loops was constructed a few years later (Fig. 1.3B, Atkinson *et al.*, 2003). The core element of this switch is a dual-input promoter $P_{lac\text{-}glnAp2}$, which is repressed by LacI and induced by NRI~P. The $P_{lac\text{-}glnAp2}$ promoter is set to drive the expression of *glnG*, whose gene product, NRI, will be phosphorylated to form NRI~P. Production of NRI~P induces its own expression and that of LacI, until LacI builds up enough to suppress NRI~P expression. At this point, LacI expression starts falling, allowing NRI~P expression to start a new cycle. Oscillation is observed via the β-galactosidase level produced from the lactose operon driven by the LacI controlled P_{lac} promoter.

With the rapid progress of synthetic gene circuits, a more precise, robust and tunable oscillator was constructed, which was also based on a dual-input promoter (Fig 1.3C, Stricker *et al*., 2008). In this network, a transcriptional repressor gene *lacI*, a transcriptional activator gene *araC*, and a reporter gene *yemGFP* are controlled by three identical copies of the dual-input promoter $P_{lac/ara\text{-}1}$, which is repressed by LacI in the absence of IPTG and activated by AraC in the presence of arabinose. In the presence of arabinose, the positive loop drives the expression of AraC, LacI, and YemGFP. While the LacI represses the

expression of these proteins in a negative feedback loop (in the absence of IPTG). The differential activity of the positive and negative feedback loops drives the oscillatory behaviour, and the period of the oscillator can be tuned by adjusting the levels of arabinose and IPTG. Cells containing this genetic circuit exhibits large-amplitude fluorescence oscillations and the oscillatory period is tunable.

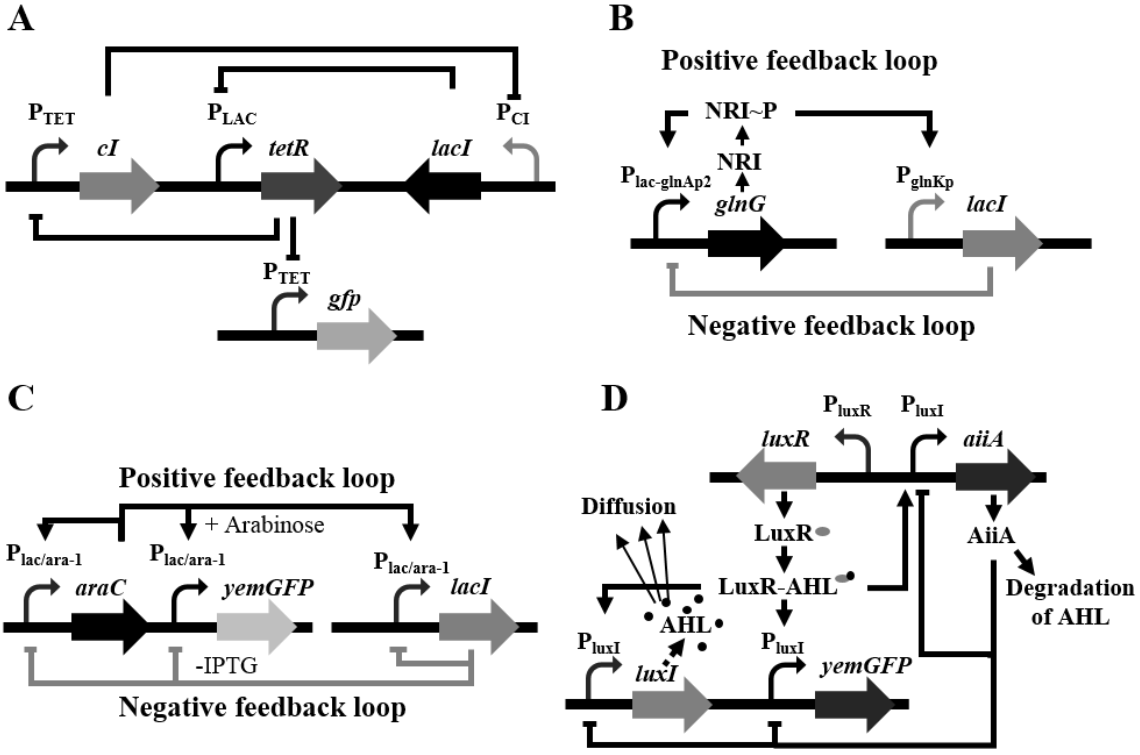


Figure 1. 3 Synthetic oscillators. (A) The repressilator. A negative feedback cycle consists of three repressors: TetR, LacI and CI. Each of them represses its cognate promoter (P_{TET}, P_{LAC} and P_{CI}) driving the expression of the next repressor. A reporter gene is driven by P_{TET} to indicate the state of the device. Adapted from Elowitz and Leibler (2000). (B) The positive/negative feedback loop oscillator. One unit contains the dual-input $P_{lac-glnAp2}$ promoter, which is activated by NRI~P and repressed by LacI, fused to *glnG* (the NRI encoding gene). LacI is produced from a NRI inducible promoter in another unit, and its effect on promoter $P_{lac-glnAp2}$ can be synchronized by growing cells in medium with IPTG, that can inactivate LacI repressor. The oscillation is reported via the fluctuating β-galactosidase level affected by the P_{LAC} driven lactose operon *lacZYA*. Adapted from Atkinson *et al.* (2003). (C) The precise, robust and tunable oscillator. The dual-input promoter $P_{lac/ara-1}$ is activated by the AraC protein in the presence of arabinose and repressed by the LacI protein in the absence of IPTG. The expression of AraC, LacI and YemGFP is driven by three separate dual-input promoter $P_{lac/ara-1}$. The oscillation is reported via the YemGFP output. Adapted from Stricker *et al.* (2008). (D) The synchronized oscillator. The expression of LuxI protein, AiiA repressor, and YemGFP are driven by the P_{luxI} promoter in three separate transcriptional modules. The expression of LuxR is driven by the constitutive promoter P_{luxR}. LuxI enzymatically produces the small molecular AHL, which can bind to the LuxR protein to activate the promoter P_{luxI} in a positive feedback loop, or be degraded by AiiA in a negative feedback loop. The AHL can diffuse outside of the cell membrane and into neighbouring cells, which is the core factor for the synchronized oscillation. Adapted from Danino *et al.* (2010).

However, none of the aforementioned oscillators can be synchronized, which is a limitation for the sensitivity and robustness of the dynamic response to external signals. This limitation

was overcome by a synchronized bacterial oscillator. This oscillator coupled the oscillatory network to a mechanism of quorum sensing system to enable population synchronization of oscillation (Fig. 1.3D, Danino *et al*., 2010). The expression of LuxI (a synthase that produces acyl-homoserine lactone, AHL), AiiA (a quorum-quenching component that mediates AHL degradation), and reporter YemGFP are under the control of the P_{luxI} promoter, which is induced by the signal molecule AHL bound to the LuxR regulatory protein (AHL-LuxR). The expression of LuxR is driven by the constitutive promoter P_{luxR}. AHL, a core component of the quorum sensing, can diffuse across the cell membrane and mediate intercellular communication. The combination of a positive feedback loop (LuxR–AHL complex actives the transcription of P_{luxI}) and a negative feedback loop (AiiA negatively regulates the transcription of P_{luxI} by catalysing degradation of AHL) results in the synchronized oscillation of cell populations at a critical cell density.

The oscillators discussed so far are completely independent from host metabolisms. The "metabolator" device coupled oscillation to glycolytic flux using the signalling metabolite acetyl phosphate, and demonstrated the possibility of using metabolic flux to control network oscillation (Fung *et al*., 2005). In addition, a recent work has demonstrated the functionality of a circadian oscillator in *E. coli*, a gut bacterium without its own native circadian rhythm. The oscillator system was transplanted from the autotrophic cyanobacteria *synechococcus elongatus*, and displayed a period that matched the geophysical day-night cycle. It can potentially be used in automated daily drug delivery, circadian control in industrial microbial process, and treatment of circadian rhythm disorders (Chen *et al*., 2015).

1.2.3 Engineering genetic logic gates for processing multiple signals

In some cases, it is desirable to take actions or make decisions based on multiple intracellular or environmental inputs. To fulfill this requirement, synthetic logic gates were constructed that can give specific phenotype in response to multiple signals. The most basic logic gates are defined by conventional symbols and their performances are represented using corresponding truth tables (Fig. 1.4). Examples of logic gates constructed with protein-based transcriptional regulators, RNA-based post-transcriptional regulators, and site-specific recombinases are introduced in this section.

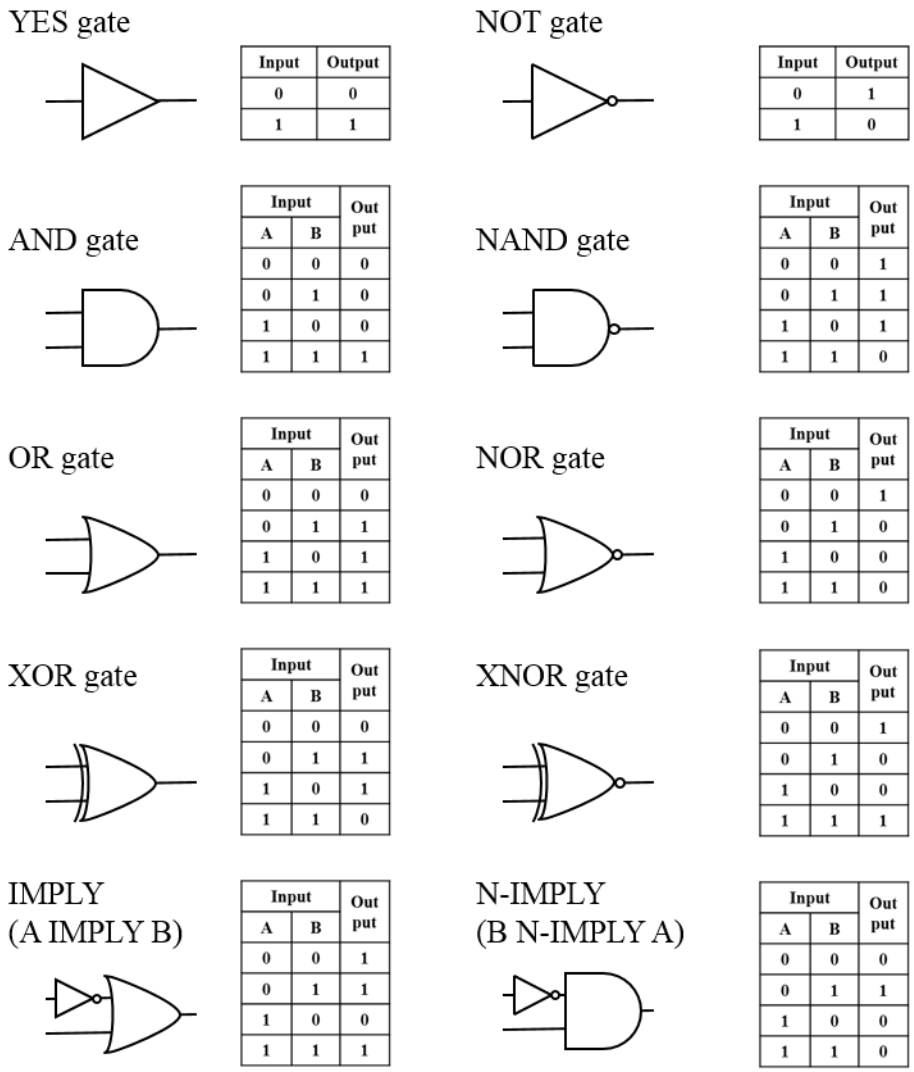
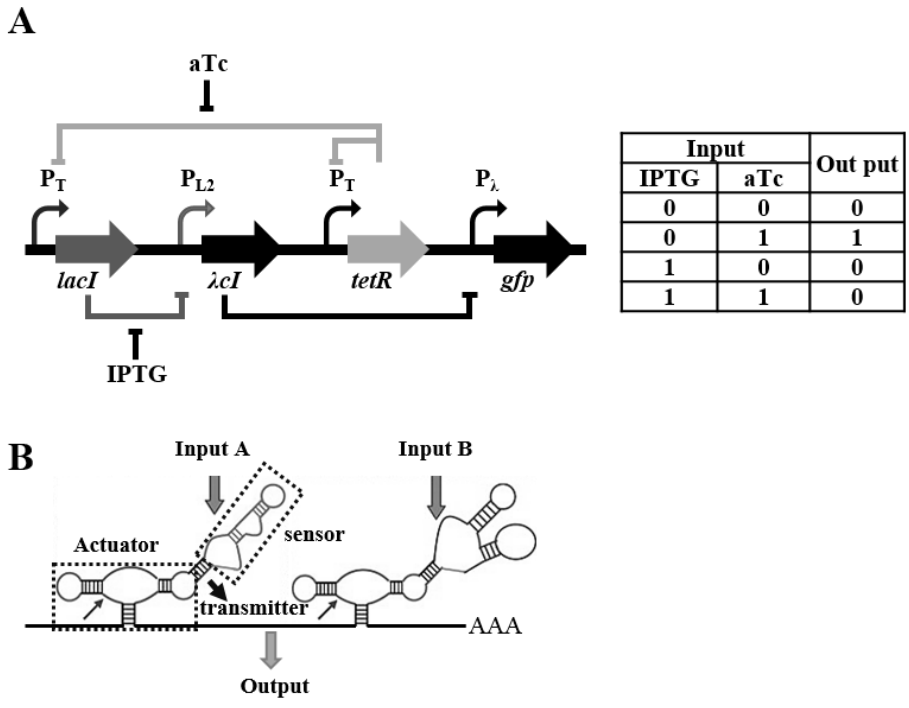


Figure 1. 4 Conventional symbols and truth tables of basic logic gates. When the combination of input signals leads to the state change of the gate the value is defined as "1". If not the value is "0".

Early genetic logic gates were constructed with protein-based transcriptional regulators (Elowitz *et al*., 2002). These networks consist of three natural transcriptional regulators LacI, TetR, and λ CI, as well as promoters that produce different responses when assembled in different configurations. The binding state of LacI and TetR can be changed with inducers, IPTG and aTc, respectively. Thus, these two inducers are used as the input signals to stimulate the networks to display different logical behaviours depending on the different arrangements of promoters and transcriptional regulators. As an example, the logic circuits showed in Figure 1.5A acts as N-IMPLY in a *lacI*$^{-}$ *E. coli* strain, producing GFP only in the absence of IPTG and in the presence of aTc.

Another approach to control the regulatory logic of gene expression is through RNA regulators. One such example was based on synthetic ribozymes (Fig. 1.5B, Win and Smolke, 2008). By coupling a ribozyme to the 3' untranslated region of a target gene, the ribozyme self-cleavage (active state of ribozyme) can inactivate the transcript and lower the gene expression. The synthetic ribozyme switch has three functional modules (actuator, sensor and transmitter), in which the "actuator" and "sensor" modules are connected by the "transmitter" module. When an input molecule binds to the sensor, the ribozyme's self-cleavage activity is changed and therefore influences the expression of the target gene. The synthetic ribozyme behaves like a single-input buffer or an inverter that is inactivated or activated in the presence of input, respectively, increasing or decreasing gene expression, respectively. When using two ribozymes to control transcription, the output is high only when both ribozymes are in their inactive states. By assembling different functional synthetic ribozymes in different combinations, the engineered devices can behave as AND, NOR, NAND, and OR logic gates. For example, the AND gate shown in Figure 1.5B consists of two single input buffers, which produces gene product only in the presence of both inputs A and B.



Input		Out put
IPTG	aTc	
0	0	0
0	1	1
1	0	0
1	1	0

Figure 1. 5 Synthetic logic gates based on transcriptional regulation and RNA regulation. (A) N-IMPLY gate (truth table on the right) based on transcriptional regulation. The expressions of LacI, TetR, λ CI and reporter GFP are controlled by P_T, P_{L2}, P_λ promoters respectively, and these promoters are repressed by TetR, LacI and λ CI, correspondingly. In a *lacI*$^-$ strain, this device induces the expression of GFP only in the absence of IPTG and in the presence of aTc. Adapted from Elowitz *et al.* (2002). (B) The AND gate based on RNA regulation. Each synthetic ribozyme containing actuator, sensor and transmitter modules functions as a single-input buffer that converts the input signal to increased gene expression output. Only in the presence of both inputs A and B, these two ribozymes are inactive, allowing production from the gene upstream of the device. Adapted from Win and Smolke (2008).

The logic gates described above depend on continuous presence of inputs to maintain their correct outputs, which have no memory of inputs. The development of genetic logic gates may require cellular memory, and this was realised in some recently constructed devices. The recombinase-based logic gates can logically convert transient chemical inputs into permanent changes in DNA architecture, implementing devices capable of memory (Bonnet *et al*., 2013; Siuti *et al*., 2013). The recombinase-based logic gates perform their function by expressing two orthogonal recombinases from different inducible promoters and inverting the DNA sequence between two pairs of recombination sites. More details of the recombinases-based logic gates are introduced below.

Siuti *et al*. (2013) engineered two-input logic gates using Bxb1 integrase and ϕC31 integrase to act on DNA segments containing gene regulatory elements (promoters and terminators) and output genes. For example, an AND gate shown in Figure 1.6A consists of an inverted promoter and an inverted *rfp* gene surrounded by ϕC31 and Bxb1 recognition sites, respectively. RFP expression can occur only after the expression of ϕC31 and Bxb1 are activated by both input signals, and the inversion of both segments.

Independently, Bonnet *et al*. (Bonnet *et al*., 2013) assembled recombinase-based logic gates, where using Bxb1 and TP901-1 integrases to control the gene regulatory elements, the majority of which were asymmetric terminators that disrupt RNA polymerase flow in only one of two possible orientations. For example, in the AND gate shown in Figure 1.6B, the logic-gate operation consists of two unidirectional terminators flanked by either Bxb1 or TP901-1 recognition sites, and located between a promoter and a reporter gene. GFP expression can occur only after the expression of Bxb1 and TP901-1 have been activated by both input signals, causing the inversion of both terminators. Logic gates constructed in this study were designed as three terminal devices that decoupled the logic-gate operation from both input (promoter upstream of the logic-gate operation) and output (reporter gene or other output gene), enabling simple tuning and reutilization (Bonnet *et al*., 2013). Overall, in both logic gate systems, the chemical input signals were processed logically into the permanent change of DNA state, leading to a stable change in gene expression.

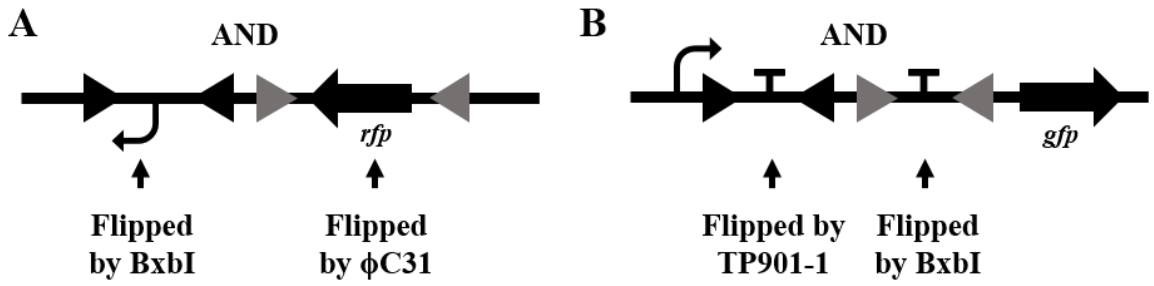


Figure 1. 6 Synthetic logic gates based on recombinases. (A) The AND gate based on flipping promoter and gene. Integrase Bxb1 and integrase ΦC31 expressions are under the control of two different regulated promoters, which are activated by two different input signals, respectively. After both input signals have been present, the expressed integrases recognise and recombine orthogonal sites and flip the promoter and gene to the same orientation, resulting in the expression of reporter protein RFP. Adapted from Siuti *et al.* (2013). (B) The AND gate based on flipping asymmetric terminators. Integrase BxbI and integrase TP901-1 expressions are under the control of two different regulated promoters, which are activated by two different input signals, respectively. After both input signals have been present, the expressed integrases recognise and recombine orthogonal sites and flip the terminators, allowing transcription to flow through and expressing the reporter protein GFP. Adapted from Bonnet *et al.* (2013)

As mentioned at the beginning of this section, these multiple-signal regulated devices would be useful in some practical applications. One example is a dual-promoter integrator constructed by Nissim and Bar-Ziv, which acts as a AND gate to more precisely and efficiently discriminate and kill cancer cells than using a single-promoter approach (Nissim and Bar-Ziv, 2010). This synthetic circuit uses two cancer-related promoters to control the expression of two fusion proteins. The fusion proteins form a transcription complex to activate the expression of an output gene (reporter gene or killer gene) regulated by its own synthetic dual-promoter. Thus, cell death is executed only when the combined activity of two cancer related promoters is high.

1.2.4 Engineering genetic counting systems and their development

Natural systems utilise counting mechanism to measure the age of biological materials, such as the shortening of telomere would ultimately lead to cell death (Counter *et al*., 1992). Construction of genetic circuits that can count would enable a variety of biotechnology applications, such as executing cell death after a specified number of cell divisions, monitoring of aging, and recording number of environmental events (Smolke, 2009). However, engineering such device has had limited success, and only two synthetic counting systems have been constructed until now (Friedland *et al*., 2009).

One of these synthetic counters exploited delays in a transcriptional cascade with additional translational regulation (Fig. 1.7A). The transcription of each gene in a downstream cascade is stimulated by a bacterial phage RNA polymerase (RNAP) expressed from a gene (in the upstream cascade). For the translational regulation, a cis-repressor sequence between the RBS and transcription start codon is complementary to the RBS sequence and therefore represses translation. A transactivating noncoding RNA (taRNA) binds to the cis repressor, relieving RBS repression, allowing translation. The production of taRNA is driven by the arabinose inducible P_{BAD} promoter. The first short pulse of arabinose activates the translation of T7 RNAP and stimulates promoter P_{T7}, resulting in transcription of T3 RNAP. After a specific interval, the second pulse of arabinose activates the translation of T3 RNAP and stimulates transcription of *gfp* from promoter P_{T3}. A significant increase of GFP signal can be detected only after the third pulse of arabinose is delivered (Friedland *et al.*, 2009).

The other synthetic counter was constructed using recombinases Flp_e and Cre as information recording tools (Fig. 1.7B). Phage P1 Cre recombinase (Austin *et al.*, 1981) and Flp_e recombinase, an improved mutation of yeast Flp recombinase (Buchholz *et al.*, 1998), catalyse reversible site-specific recombination between two identical sequences, *loxP* sites for Cre and *FRT* sites for Flp_e. The inducible P_{BAD} promoter, flp_e gene and a terminator were placed between inverted repeat recognition sites for Flp_e recombinase, forming a Single Invertase Memory Module (SIMM). This SIMM is arrayed with a Cre SIMM in tandem so that the expression of the upstream recombinase induced by an input pulse of arabinose flips the entire SIMM and inverts the promoter into the correct orientation for the next recombinase. In this way, the SIMM is set from one state to the other, storing one-bit information in the circuit. After adding two arabinose pulses, the promoter upstream of *gfp* gene is placed in a correct orientation to drive the expression of GFP, and when the third arabinose pulse is delivered, a significant increase of GFP signal can be detected. If the inducible promoters are different, this circuit can be used to record three different input events in a specific order (Friedland *et al.*, 2009).

Both counting circuits perform counting functions based on specific lengths of input pulse. The transcriptional-based counter had a faster response to pulses (efficient response occurred with pulse lengths of ~20 to 30 min and pulse intervals of 10 to 40 min), whereas inversion by recombinases took more time (efficient response occurred with pulse lengths and intervals range from 2 to 12 hours). They can be applied to count different events depending on the time scales and frequencies of events.

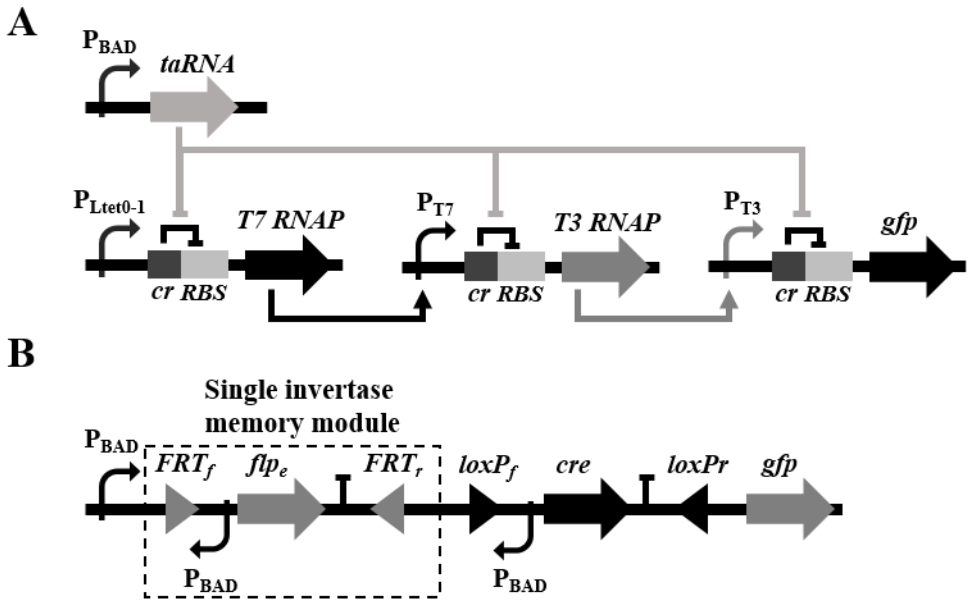


Figure 1. 7 Synthetic counters. (A) The counter based on transcriptional and translational regulation. The transcriptional regulatory cascade of this device has three nodes. The transcription of T7RNAP gene is driven by a constitutive promoter P_{Ltet0}-1, and its gene product drives transcription of the gene encoding T3 RNAP in the second node. T3 RNAP ultimately drives transcription of *gfp* gene in the third node. All transcripts are cis-repressed with the same riboregulator sequence. The taRNA, driven by arabinose inducible P_{BAD} promoter, binds to cis repressor to relieve its repression. An increase of GFP signal can be detected after three pulses of arabinose have been delivered. (B) The counter based on recombinases Flp_e and Cre. This device contains two Single Invertase Memory Modules (SIMMs) that are arrayed in a cascade. Each module consists of inverted repeat recombination sites flanking its own recombinase encoding gene and an inverted inducible promoter P_{BAD}. The first time the device receives a short pulse of input signal, the Flp_e recombinase in the upstream module is expressed, which can flip the DNA sequence between sites FRT_f and the FRT_r. This flipping places the promoter in the correct orientation for expressing Cre recombinase gene in the next module, the flp_e gene is no longer expressed and the cascade advances by one step. The second pulse of input signal drives expression of Cre recombinase and flips the DNA sequence between $loxP_f$ and $loxP_r$. This places the promoter in the correct orientation for driving expression of GFP and stops expression of Cre. After three input pulses, an increase of GFP signal can be detected. Both figures are adapted from Friedland *et al.* (2009).

In comparing the transcriptional-based and the recombinase-based counters, the information encoded into the DNA sequence by site-specific recombination is not susceptible to perturbations, and the state can be maintained in DNA sequence even after cell death. The recombinase-based counting system would be a promising direction as long as this technology is proved to be reliable under more circumstances.

Despite the advantages of the recombinase-based counting system discussed above, several limitations with this system need to be addressed. First, the state of each recombinase-based bit can only be set but not reset. Thus, the number it can count is linearly proportional to the number of recombinases used. Second, the uncontrolled directionality of reaction catalysed by recombinases in this system leads to continual flipping and influence on stability. Third, only transient signals with specific time scale can be recorded and only the final state can be

reported by this circuit, limiting its use in recording events with different time scales and frequencies.

In consideration of these limitations, two different synthetic circuits have been proposed and constructed, which have the potential to be developed into counters. One such example is a push-on push-off switch, which can generate two distinct outputs in response to the same input signal (Fig. 1.8, Lou *et al*., 2010). This device rationally coupled a bistable switch memory module and a NOR gate. The memory module consists of two mutually repressed repressors λ CI and CI434 (encoded by the *cI434* gene from phage 434). The transcription of *cI* and *cI434* genes are controlled by P_{RM} and P_R promoters, which are repressed by CI434 and λ CI in turn. Reporter proteins RFP and GFP are expressed from promoters P_{RM} and P_R, respectively. In the NOR gate, the transcription of cI_{ind-} (a mutant of *cI* whose gene product cannot be degraded by RecA protease) is controlled by the regulated promoter P_{NOR}, which is repressed by both LacI and LexA repressors. If either LacI or LexA is in an ON state, the output CI_{ind-} from P_{NOR} is OFF. UV irradiation activates RecA protease to specifically degrade CI434, CI, and LexA, but not LacI and CI_{ind-} (Lou *et al*., 2010).

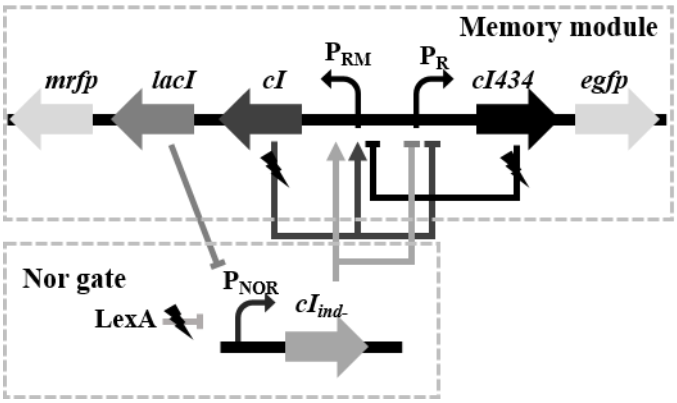


Figure 1. 8 A synthetic push-on push-off switch. This device consists of two modules: the memory module and the NOR gate module. In the memory module, two promoters (P_{RM} and P_R) drive the expression of their mutual repressors (CI and CI434). In the NOR gate, cI_{ind-} is under the control of a regulated promoter, which can be repressed by either LexA or LacI. LexA is constitutively expressed in the cell. The LacI expressed from the memory module and the CI_{ind-} expressed from the gate are two interconnecting parts that connect these two modules. The UV irradiation activates RecA to specifically degrade CI434, CI, and LexA, but with no effect on LacI and CI_{ind-} (represented by the lightning). Adapted from Lou *et al.* (2010).

When the initial state of this device is "OFF" (RFP OFF), only P_R promoter is activated and this device expresses GFP and CI434. With the first UV pulse, LexA and CI434 are degraded by RecA and the expression of CI_{ind-} is activated from the NOR gate, resulting in the activation of P_{RM} and inhibition of P_R, and pushing the memory module to "ON" (RFP ON)

state. During this state, LacI and CI are expressed from P_{RM}, and the expression of CI_{ind-} is repressed. After the second UV pulse, the CI is degraded by RecA and its repression on promoter P_R is removed, restoring the memory to "OFF" state.

This study proposed that by using N distinct proteases an N-bit binary counter can be constructed that counts from 0 to 2^N-1, with advantage of storage capacity over linear counters. However, the switching efficiency of this device is less than 30% after one operation cycle, and continuous protein expression is necessary for state maintenance.

Given the reality that genetically encoded counting systems have developed slowly. The design for a toggle flip-flop was proposed (Fig. 1.9, Subsoontorn and Endy, 2012). This toggle flip-flop consists of a pair of orthogonally-operating set-reset latches, "A-latch" (grey parts in Fig. 1.9) and "I-latch" (black parts in Fig. 1.9). Each latch performs like a rewritable recombinase addressable data module (Bonnet *et al*., 2012) described in section 1.2.3. The A-latch controlled outputs (Ii, or Ii and Ei) execute the set and reset operation of I-latch, while the I-latch controlled outputs (Ia, or Ia and Ea) execute the set and reset operation of A-latch. The overall output of the system is connected to A-latch by placing a constitutive promoter between recombination sites for integrase Ia. An input signal triggers the expression of activator A and repressor R, which activates and represses the promoters flanked by recombination sites in A-latch and I-latch, respectively.

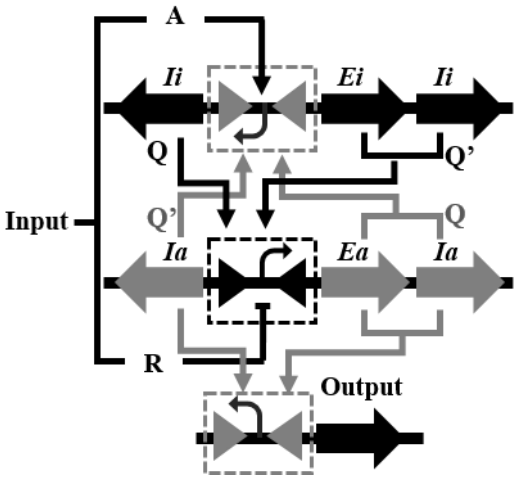


Figure 1. 9 A synthetic toggle flip-flop. This device consists of a set of orthogonal recombinase-based set-reset latches, "A-latch" (light grey part) and "I-latch" (dark grey part). I-latch outputs (Q and Q') are connected to A-latch inputs (set and reset operation), respectively. A-latch outputs (Q and Q') are connected to I-latch inputs (reset and set operation), respectively. The single input signal for the whole device triggers the expression of activator A and repressor R, which activates and represses the promoters flanked by recombination sites in A-latch and I-latch, respectively. The output of the device is determined by the state of A-latch. The *Ia* and *Ii* represent integrase genes, and the *Ea* and *Ei* represent Xis/RDF genes. Adapted from Subsoontorn and Endy (2012).

Given a flip-flop toggle with both A and I latches in “OFF” state (as shown in Fig. 1.9), during the first input pulse, the Ii expression is activated to set the I-latch from “OFF” to “ON”, while the A-latch holds “OFF” state. After removing the first input pulse, the Ia expression is activated to set the A-latch from “OFF” to “ON”, switching on the output expression, while the I-latch holds state. During the second input pulse, the Ii and Ei expressions are activated to reset I-latch from “ON” to “OFF”, while A-latch holds state. After removing the input pulse, the Ia and Ea expressions are activated to reset A-latch from “ON” to “OFF”, switching off the output expression, while the I-latch holds state. The whole system is restored to the initial state for a new cycle of operation.

In summary, the mechanisms described above allow a toggle flip-flop to switch back and forth between two states using the same input pulse. A combinatorial N bits binary counter, which counts from 0 to 2^N-1, could be assembled from N toggle flip-flops by using the output of the $(N-1)^{th}$ bit as an input for the N^{th} bit. To construct such counter, 2N sets of orthogonal integrase-excisionase are needed since each toggle flip-flop is based on two sets of integrase-Xis/RDF. This design based on the highly directional integrases has the potential to overcome unstable switching with the previous recombinase-based counters (Friedland *et al*., 2009), and it can be used to detect and record events with a large time scale. However, this design is complex and to record one bit information needs two sets of integrase-Xis/RDF, limiting the scaling up and information storage capacity.

The purpose of this study is to construct a simple and efficient binary counting module based on highly directional serine integrases which belong to the serine site-specific recombinase family. Before presenting details of this study, the mechanisms of site-specific recombination are first described in section 1.3.

1.3 Introduction to genetic recombination

Genetic recombination is an important process that brings about DNA modification and rearrangement. Three main types of recombination, homologous recombination, site-specific recombination, and transpositional recombination, have been identified in prokaryotes and eukaryotes. In this section, homologous recombination and transpositional recombination are briefly introduced, while site-specific recombination (the subject of this study) is described in more details.

1.3.1 Homologous recombination

Homologous recombination is an exchange between two similar or identical DNA sequences, resulting in new sequence combinations. A series of proteins is involved in the process of homologous recombination in bacterial strains, including RecA, RecBCD, RecE, and RecF (Kowalczykowski *et al*., 1994). With the help of these proteins a key crossed-strand intermediate, the Holliday junction is formed (Holliday, 1964). This model possesses three important features: 1) single strand break in both of the DNA homologs; 2) formation of a heteroduplex region through symmetric exchange of the DNA strands; 3) cleavage and rejoining of the strands to yield recombinant progeny (Kowalczykowski *et al*., 1994). Many advanced variations of the Holliday model were developed to explain particular details of recombinant production, such as Meselson-Radding Model (Meselson and Radding, 1975) and double-strand break repair model (Dudás and Chovanec, 2004). Homologous recombination plays an essential role in repairing damage DNA, generating genetic diversity, and its application as biotechnological tools for gene targeting, protein engineering, and cancer therapy.

1.3.2 Site-specific recombination

In contrast to homologous recombination, site-specific recombinases catalyse the exchange between DNA strands, without a requirement for significant sequence homology (Stark *et al*., 1992). The recombination minimally involves, two specific DNA sequences (recombination sites), a specialised recombinase protein for recognising, breaking and re-joining the DNA sites, and a mechanism for DNA breakage and reunion with conservation of energy (Grindley *et al*., 2006). In some systems a number of accessory proteins and accessory sites are needed, for example, integration host factor (IHF) is necessary for λ integration (Landy, 1989).

Site-specific recombination systems can be highly specific and efficient in a variety of heterologous cells including bacteria and higher eukaryotes (Hirano *et al*., 2011). Depending on the initial arrangement of the recombination sites, site-specific recombination generates one of three possible outcomes: integration, resolution, or inversion (Fig. 1.10, Gellert and Nash, 1987). Integration results from recombination between sites on separate circular DNA molecules. For sites on the same molecule, resolution results from recombination between

two sites that are in the same orientation (direct repeat), whereas inversion results from recombination between two sites that are in an opposite orientation (inverted repeat).

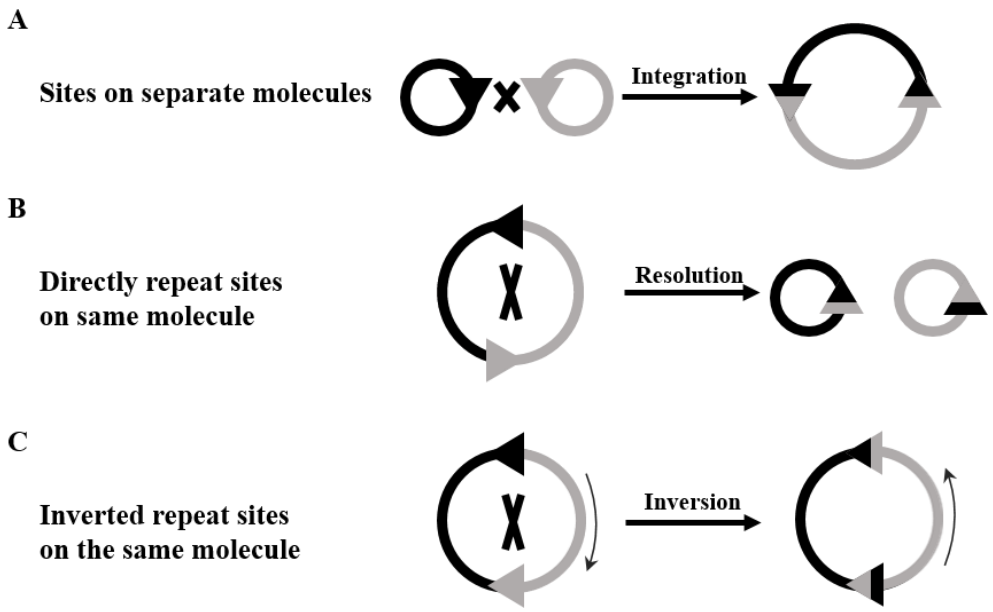


Figure 1. 10 Outcomes from site-specific recombination in circular molecules. (A) Integration. (B) Resolution. (C) Inversion. Recombination sites are represented by black and grey triangles.

With these characteristics, site-specific recombinases are employed in number of applications such as: genome manipulation (Brown *et al*., 2011; Sauer, 1998; Thyagarajan *et al*., 2001), metabolic pathway assembly (Colloms *et al*., 2013), and their functions in genetic devices (Bonnet *et al*., 2012, 2013; Friedland *et al*., 2009; Ham *et al*., 2008; Siuti *et al*., 2013). Site-specific recombination is reviewed in more detail in a later section (section 1.4).

1.3.3 Transpositional recombination (Transposition)

Another kind of DNA rearrangement, transpositional recombination (transposition), is usually grouped together with site-specific recombination, since it also promotes rearrangements between DNA segments that lack extensive homology (Craig, 1988). However, its genetic consequences and reaction mechanisms are dramatically different from the conservative site-specific recombination introduced above (section 1.3.2), therefore, transposition is introduced separately and the definition of site-specific recombination in this study excludes transposition events.

Transposition is a process in which a mobile genetic element moves from one site to another in DNA. These mobile genetic element belongs to one of two classes, according to the

transposition intermediate: class I transposable elements (or retrotransposons) move via an RNA intermediate and class II transposable elements transpose directly at the DNA level (Finnegan, 1992). The class I transposition functions by a “copy and paste” mechanism that leaves behind the original copy and generates a new copy to be inserted elsewhere. The class II mobile elements are named transposons and the process is catalysed by enzymes called transposases. This type of transposons generally moves with a “cut and paste” mechanism, in which the transposon in the donor DNA is excised and then inserted into the target DNA by a transposase (Muñoz-López and García-Pérez, 2010).

Transposons have been used as genetic tools to create tagged mutations and deliver transgenes to a wide variety of organisms. One common group of transposable elements in bacteria is called "insertion sequences" (ISs), containing terminal inverted repeat sequences (IRs) and a gene encoding a transposase. Two ISs can mobilise a segment of DNA between them, forming a composite transposon. Transposon ISY100, as a member of IS630/Tc1/mariner family of transposons, has been demonstrated to undergo double strands cleavage at the transposon ends, and to transpose to new TA targets in *E.coli* on induction (Urasaki *et al*., 2002). This transposon was used as a genetic tool to deliver a single copy DNA sequence into the *E. coli* chromosome in this study, and the detailed methods are described in chapter 5.

1.4 Classification of site-specific recombinases: tyrosine vs. serine recombinases

The overall process of site-specific recombination can be divided into four general steps: 1) a recombinase recognises and binds to recombination sites; 2) the recombinase brings sites together to form a synaptic complex; 3) the recombinase mediates strands cleavage, exchange and the rejoining of the DNA; 4) the synaptic complex dissociates to release the recombinant products (Stark *et al*., 1992; Urasaki *et al*., 2002). Details of this process are different between the two distinct families of site-specific recombinase: tyrosine recombinases and serine recombinases. These two families have distinct protein sequences and recombination mechanisms, and were named after the nucleophilic amino acid residue that links the recombinase and DNA phosphodiester backbone covalently in the reaction intermediate (Landy, 1989; Stark *et al*., 1992; Nash, 1996). The mechanisms of both families and their representatives are described in this section.

1.4.1 Tyrosine recombinases

The mechanism of site-specific recombination mediated by tyrosine-type recombinases is shown in Figure 1.11. The tyrosine recombinase recognises a specific sequence in DNA and forms a synaptic complex containing two DNA duplexes bound by a tyrosine recombinase tetramer. A diagonally opposite pair of recombinases cleaves one strand of each duplex at a time, creating covalent DNA-protein phosphotyrosine linkages at the 3' ends of DNA. The free 5' ends attack the 3' phosphotyrosines of the other strands to form a Holliday junction intermediate (Van Duyne, 2002). After the isomerization of the Holliday junction intermediate the process is repeated with exchange of the second pair of strands catalysed by the other diagonally opposite pair of recombinases (Grindley *et al*., 2006).

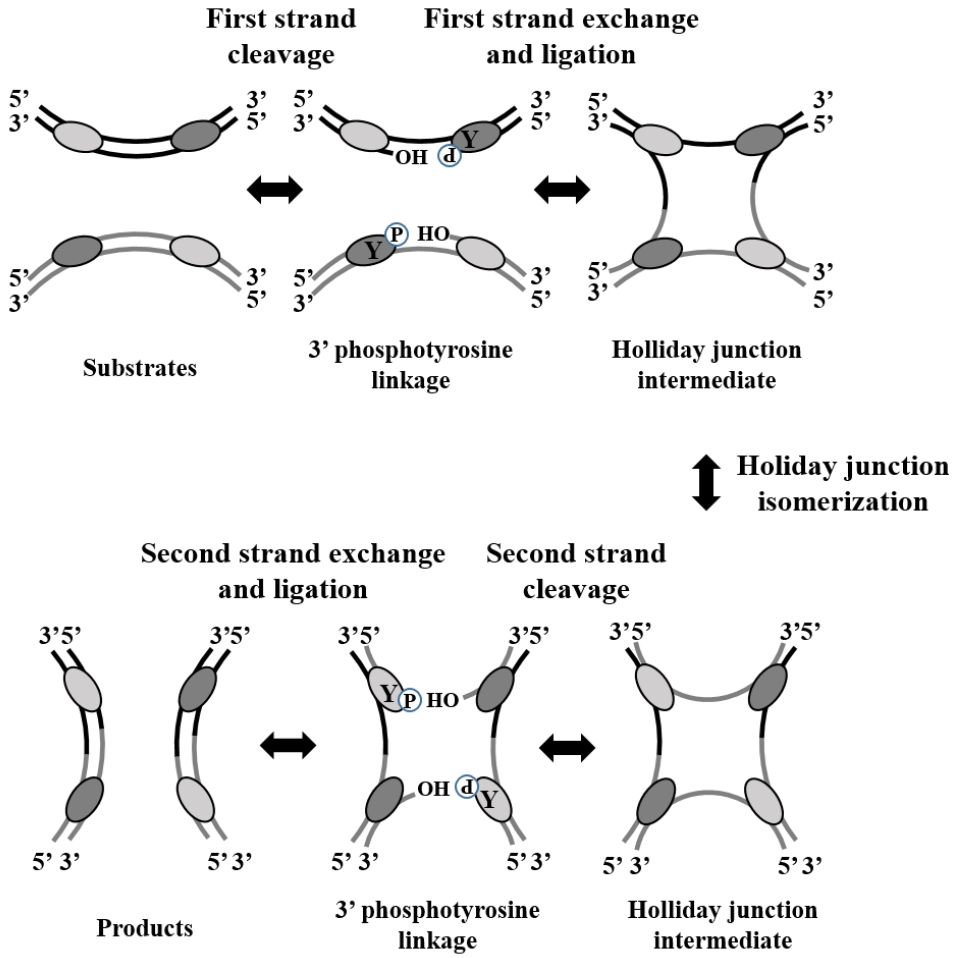


Figure 1. 11 Mechanism of recombination by tyrosine recombinases. The scheme illustrates the strand cleavage, exchange and ligation catalysed by tyrosine recombinases. The synaptic complex comprises two DNA duplexes bound by four recombinase monomers. Dark grey monomers are active for cleavage in the top half of the pathway, and light grey monomers are active for cleavage in the bottom half of the pathway. Initially, one strand in each duplex is cleaved by a nucleophilic tyrosine resulting a covalent 3' DNA-protein phosphotyrosine linkage and a free 5' DNA hydroxyl. Then, the free 5' DNA hydroxyl ends attack the 3' phosphotyrosine on the opposing DNA duplex to form a Holliday junction structure. The complex then isomerizes to alternate the catalytic activity between the two pairs of monomers, and the process is repeated with the other strand. Last, the protein is released and the Holliday junction-like structure is resolved. Y represents active tyrosine residues. Adapted from Van Duyne (2002).

1.4.2 Serine recombinases

Serine recombinases recognise specific sequences in DNA and form synaptic complexes containing two crossover sites bound by four recombinase subunits forming a tetramer (Fig. 1.12). In contrast to tyrosine recombinases, all DNA strands are cut by the serine recombinase in advance of strand exchange (Stark *et al*., 1992). The cleaved intermediate contains covalent DNA-protein phosphoserine linkages at the 5' ends of DNA. The strand exchange mechanism involves a motion of 180 ºrotation of one half of the complex relative to the other half, and then the 5' phosphates are religated to the free 3' hydroxyl groups of the unrotated DNAs (Smith *et al*., 2010).

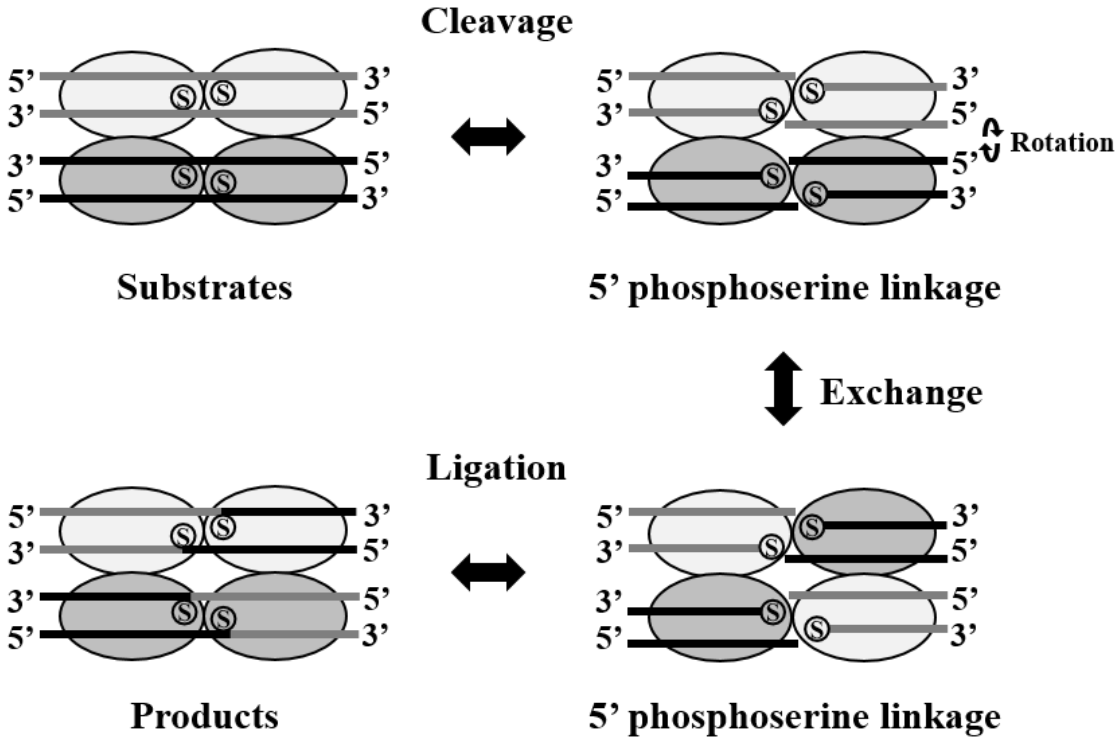


Figure 1. 12 Mechanism of recombination by serine recombinases. The scheme illustrates the strand cleavage, exchange and ligation catalysed by serine recombinases. A tetramer is formed by two serine recombinase dimers binding recombination sites and coming together. Then, the four serine recombinase monomers are activated and make two double strand breaks. One half of the complex rotates relative to the other half, and the free DNA 3' hydroxyls attack the 5'phosphoseryl linkages to rejoin the sites in the recombinant configuration. S represents active serine residues. Adapted from Grindley *et al.* (2006).

1.4.3 Biological function of site-specific recombinases

Site-specific recombination reactions are commonly used in nature for phage integration and excision, regulation of gene expression by DNA inversion, and cointegrate resolution (Grindley *et al*., 2006). Both tyrosine and serine recombinases are structurally and

functionally diverse and include resolvases, invertases, and integrases (Smith and Thorpe, 2002). Resolvases normally impose a topological filter to promote the excision reaction and prevent inversions and intermolecular recombination. Invertases only recombine sites in inverted repeat on the same molecule. Both the resolvases and invertases catalyse recombination between identical sites. In contrast, the phage integrases usually catalyse recombination between sites with different sequences, and recombination mediated by phage integrase is not limited to a specific topology, and can be integration, resolution, or inversion. Examples of the two families of site-specific recombinases and their biological functions are listed in Table 1.1.

Table 1. 1 The classification and function of recombinases mentioned in this study

Recombinase	**Biological function**
Tyrosine recombinase family	
λ Int and many other phage integrases	Integration and excision of phage genomes
Cre	Excision: dimer reduction of phage P1 plasmids
XerC/D	Excision: dimer reduction in the bacterial chromosome and plasmids
FimB, FimE	Inversion: alternation of gene expression
Flp	Inversion: for amplification of yeast 2-μm plasmid
Serine recombinase family	
Tn3 resolvase	Excision: transposon co-integrate resolution
Sin of Staphylococcus aureus	Excision: dimer reduction in straphylococcal plasmids
Hin	Inversion: alternation of gene expression in Salmonella
Gin, Cin	Inversion: alternation of gene expression in phage Mu and P1
ϕC31 Int and other large serine integrase	Integration and excision of phage genomes

1.5 Integrases and their applications

The tyrosine recombinases are further subdivided into two groups based on recombination directionality: bidirectional simple recombinases and unidirectional tyrosine-type phage integrases. The serine recombinases are subdivided into two groups due to protein size: small serine-type resolvase/ invertases and large serine-type phage integrases (Nafissi and Slavcev, 2014). The synthetic gene circuits constructed in this study were based on a serine integrase, one kind of phage integrase. Here, more details about phage integrases are introduced.

Phage integrases catalyse integration and excision of a prophage into and out of the host chromosome (Fig. 1.13). Integration is mediated by the phage encoded integrase in the absence of any other phage-encoded factors through a site-specific recombination reaction between the phage attachment site, *attP*, and the bacterial attachment site, *attB*, to give the hybrid sites, *attL* and *attR*. In the presence of a phage-encoded accessory protein known as Xis or RDF, the cognate integrase mediates the reverse reaction between the *attL* and *attR*, leading to excision (Smith *et al*., 2010). *attB* and *attP* have different sequences, therefore, the recombination products, *attR* and *attL* also have distinct sequences, allowing the recombinase to distinguish between the forward integration (*attP*×*attB*) and reverse excision (*attR*×*attL*) reactions. The mechanisms and applications of the two different groups of integrase (tyrosine-type integrases and large serine integrases) are introduced below.

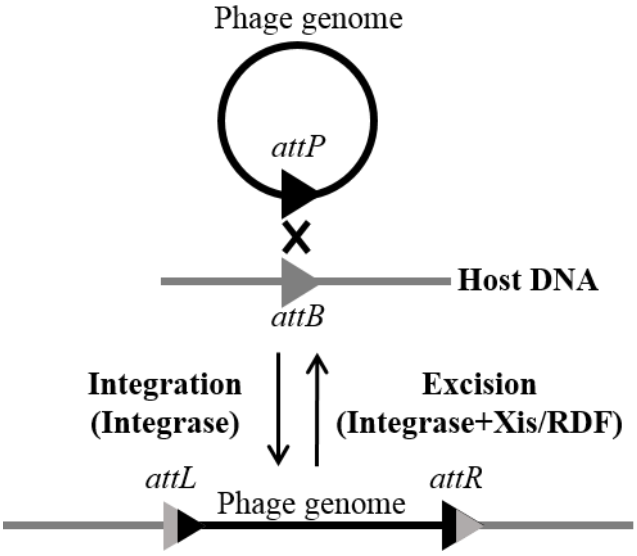


Figure 1. 13 Mechanism of integration and excision catalysed by phage integrase. Integrase alone catalyses integration between *attP* and *attB*, generating *attR* and *attL* and resulting in integration. In the presence of Xis/RDF, integrase catalyses excision between *attL* and *attR*, regenerating *attP* and *attB* and resulting in excision. The phage genome is represented by a black circle and its recombination site is represented by a black triangle. The host genome is represented by grey line and its recombination site is represented by a grey triangle.

1.5.1 Tyrosine integrases

The recombination mechanisms used by Cre from phage P1, and tyrosine integrase, λ Int, are representatives of almost all tyrosine recombinases (Grindley *et al*., 2006). The P1 phage encoded Cre mediates recombination between two *loxP* sites, which can circularise the P1 DNA or break down dimer P1 DNA molecules (Abremski and Hoess, 1984). Since the

substrate and product *loxP* sites are identical, Cre cannot distinguish between forward and reverse reactions (e.g. integrase and excision). Recombination mediated by Cre does not require accessory factors, nor is it biased to any specific conformation of DNA substrate or product.

The paradigm of the tyrosine integrases is λ phage integrase, which mediates integration and excision of the bacteriophage λ genome into and out of the *E. coli* chromosome. Unlike Cre and its simple recombination sites *loxP*, λ Int catalyses recombination on four different sites (*attP, attB, attL,* and *attR*), and accessory proteins are necessary to drive the reaction to desired direction. For the integration reaction, the synaptic complex containing a tetramer of integrases bound to *attB* and *attP* is formed with the help of a host-encoded protein integration host factor (IHF). Then, the sequential DNA cleavage and exchange steps occur, generating *attL* and *attR*. For the excision reaction, integrase, IHF and additional accessory proteins, a phage-encoded Xis and a host-encoded factor for inversion stimulation (Fis), are required (Abremski and Gottesman, 1982; Ball and Johnson, 1991). The excision reaction proceeds through a complex whose geometry differs from the complex formed during integration. Strand cleavage and exchange reactions then occur, resulting in *attP* and *attB* sites.

1.5.2 Applications of Tyrosine integrases

The most widely used tyrosine integrase recombination system in biotechnology is Gateway cloning, which is based on the *in vitro* site-specific recombination of λ integrase (Hartley *et al*., 2000). This system exploits the ability of the integrase to rearrange DNA that is flanked by two *att* sites. These two sites must contain 7 bp of homologous core sequence, but this 7 bp core sequence can be mutated to generate orthogonal variant pairs for multipart cloning.

The use of tyrosine integrase in the Gateway cloning process overcomes the problems associated with traditional cloning, for example, reliance on the restriction sites, large time-consuming and multiple purification steps. This method is also proven to facilitate multipart assembly and the assembled constructs can be further modified (Cheo *et al*., 2004). However, there are several drawbacks of tyrosine integrase based DNA assembly. First, the tyrosine integrase-mediated recombination requires host-encoded accessory proteins which limits the potential for using this system in heterologous hosts. Second, the scar sequences (recombination sites) generated during complex DNA assembly by tyrosine integrase might lead to unexpected effects on the expression condition of nearby genes. Third, the efficiency

of Gateway reaction decreases with increasing size of the DNA parts involved, and the reactions have been shown to be inefficient when transferring DNA parts over 3 kb (Hartley *et al*., 2000; Marsischky and LaBaer, 2004).

In addition to Gateway cloning, λ integrase, as well as several other tyrosine integrases, have been used in chromosome integration and cloning (Mart ń *et al*., 1991; Hoang *et al*., 2000). Integrase-mediated chromosomal integration systems rely on either delivering a single copy suicide plasmid encoding Int, *attP* and a selectable marker gene into the target cell carrying an endogenous *attB* site on the chromosome, or introducing a suicide plasmid containing *attP* into target cells that already contain a separate integrase expression plasmid (Fogg *et al*., 2014). The recently developed "clonetegration" technique, combining cloning and site-specific integration into a single step, decreases the time and effort needed for chromosome integration. As reported, up to four expression cassettes can be integrated into the chromosome in separate rounds of clonetegration and two cassettes can be integrated into the chromosome in the same round (St-Pierre *et al*., 2013). The tyrosine integrases have also been demonstrated to be able to integrate large DNA fragments into a specific genomic location, for example, the intB31 λ-like integrase has been shown to introduce cosmid and BAC substrates (up to 75 kbp), into the chromosome of *Pseudomonas putida* (Miyazaki and van der Meer, 2013).

1.5.3 Serine integrases

Serine integrases belong to the large serine recombinase family (Smith and Thorpe, 2002). Currently, the best characterised serine integrases are those of the *Streptomyces* phage, ϕC31 and ϕBT1, the mycobacteriophage Bxb1 and the actinophage TG1 (Ghosh *et al*., 2003; Gregory and Smith, 2003; Morita *et al*., 2009; Smith *et al*., 2010). Each integrase efficiently catalyses recombination between short (< 50 bp) recombination sites *attP* and *attB*, generating *attR* and *attL* which are no longer substrates for integrase alone. In the presence of the phage encoded Xis or RDF, integrase mediates recombination between *attR* and *attL*, and usually recombination between *attP* and *attB* is inhibited in this situation (Ghosh *et al*., 2006; Khaleel *et al*., 2011).

Serine integrases consist of two conserved regions: a ~150 amino acid N-terminal catalytic domain (NTD) and a 300-550 amino acid C-terminal DNA binding domain (CTD). The two domains are linked by an α-helix (Van Duyne and Rutherford, 2013). The NTD contains the

catalytic serine residue and is involved in synapsis formation, strand cleavage, and exchange. The CTD is known to be required for *att* site recognition and binding, and control of the recombination directionality (Smith *et al*., 2010).

The recently reported structure of the CTD from the LI integrase bound to *attP* half-site provided clues about how integrase regulates recombination directionality (Rutherford *et al*., 2013). As shown in Figure 1.14, the CTD is comprised of three domains: a recombinase domain (RD), a zinc-ribbon domain (ZD), and an extended coiled-coil motif (CC) which is embedded in the ZD. The RD interacts with a stretch of 13 bp adjacent to the central dinucleotide of *attP*, and the ZD forms interactions with a 9-bp "ZD motif" at the distal end of the site. With the knowledge of how integrase binds to *attP*, a plausible model for integrase binding to *attB* was proposed. In this model, the RD binds to *attB* in a similar position to the one seen in the Int-*attP* complex, whereas, the ZD-binding motif is shifted 5-bp towards the central dinucleotide relative to its position in *attP* (Rutherford *et al*., 2013).

In the CTD, the CC motif was reported to be involved in both promoting *attP*×*attB* recombination and inhibiting *attL*×*attR* recombination (Rowley *et al*., 2008), and its function was suggested by using the structural model of Int-*attP* and Int-*attB* complexes (Rutherford *et al*., 2013). The CC motifs from the P and B half-sites or from the P' and B' half-sites are well positioned to interact due to the proper ZD-ZD distance, bringing *attP* and *attB* together. After strand cleavage, rotation and ligation, the CC motifs from the P and B' or from the B and P' are close enough to form autoinhibitory interactions, blocking further recombination (Fig. 1.14).

However, in the presence of a RDF, *attP*×*attB* recombination is inhibited and *attL*×*attR* recombination is stimulated. DNA binding experiments indicated that RDFs of BxbI and ϕC31 do not bind to attachment sites, instead, the RDFs bind to the integrase proteins and alter the recombination direction. These RDFs have also been reported to inhibit the formation of *attP*×*attB* synapsis while stimulating the formation of *attR*×*attL* synapsis (Ghosh *et al*., 2006; Khaleel *et al*., 2011). From the crystal structure of the LI DNA binding domain bound to *attP* site, a plausible explanation for the mechanism of the RDF has been proposed. The RDFs are suggested to influence the interactions of CC motifs, inhibiting the

formation of the *attB*×*attP* synaptic complex and providing compensatory interactions for the formation of the *attL*×*attR* synaptic complex (Rutherford and Van Duyne, 2014).

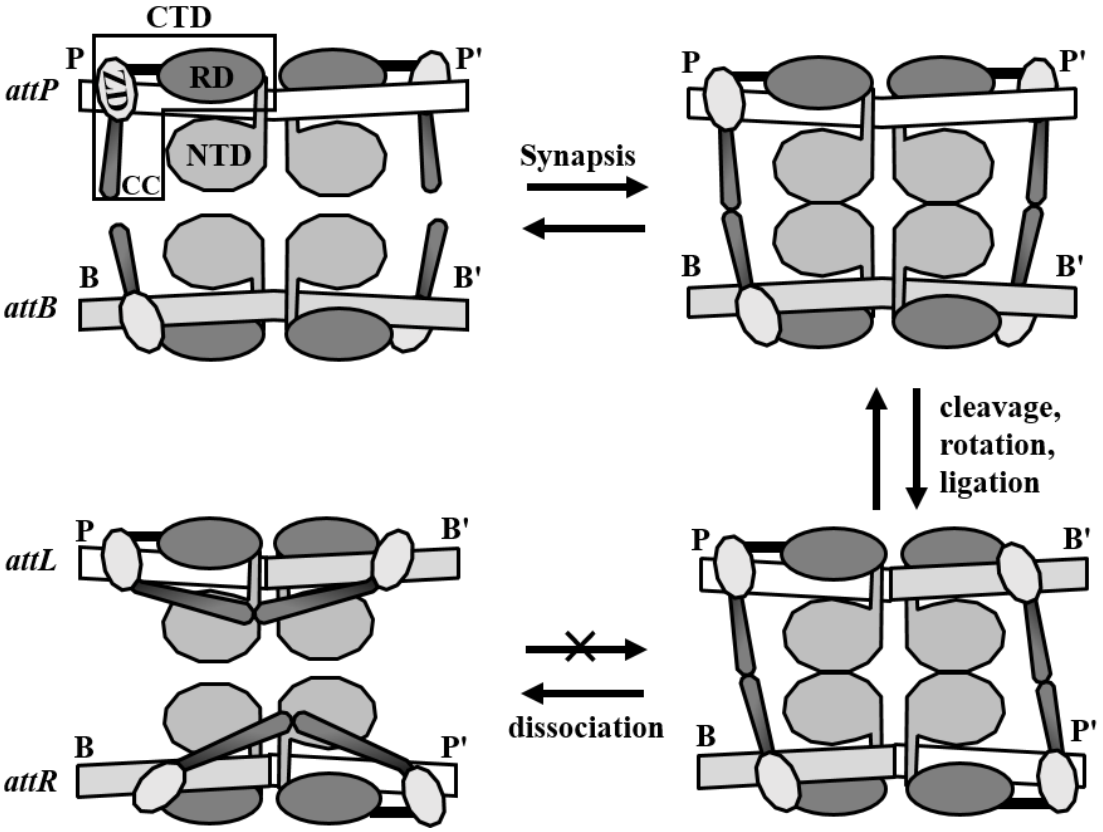


Figure 1. 14 Recently proposed model of the integration reaction catalysed by serine integrases. Integrase dimers bind to specific *attP* and *attB* sequences to form Int-*attP* and Int-*attB* complexes. These complexes are conformationally distinct due to the different positioning of the zinc-ribbon domain (ZD). Int-*attP* and Int-*attB* associate to form a tetramer complex that is stabilised by interactions between the coiled-coil (CC) motifs. Then the strand cleavage, exchange and ligation are catalysed by a serine integrase in the same manner as other serine recombinases, resulting in formation of two new attachment sites, *attL* and *attR*. Finally, the arrangement of ZDs in the *attL* and *attR* sites allows the CC motifs to form interactions that prevents the reverse reaction. NTD: N-terminal catalytic domain; RD: recombinase domain; ZD: zinc ribbon domain; CC: coiled-coil motif. Adapted from Rutherford *et al.* (2013).

1.5.4 Applications of serine integrases

Serine integrases possess a number of advantages (short and simple substrate sites, high specificity, lack of requirement for DNA supercoiling, and strict control of recombination direction) over tyrosine integrases, making them powerful tools for genome engineering and synthetic biology.

ϕC31integrase, one of the most widely used serine integrases, was originally applied in the development of an integration vector for use in *Streptomyces* containing an innate *attB* site for integration (Kuhstoss *et al*., 1991). ϕC31 integrase was also proven to have integration ability in the mammalian cell using endogenous pseudo-*attP* sites (Thyagarajan *et al*., 2001). Integration into pseudo-*attP* sites has been used in several medicinally relevant applications, for example, production of human Coagulation Factor VIII in mice (Chavez *et al*., 2012) and expression of transcription factors to reprogram mouse embryonic fibroblasts and human amniotic fluid cells into induced pluripotent stem cell (Ye *et al*., 2010). To better use the high efficiency of site-specific recombination in heterologous hosts, a docking or landing site containing a cognate attachment site has been introduced into the chromosome of various organisms for genome modification, such as Drosophila (Venken *et al*., 2006), zebrafish (Hu *et al*., 2011), and mammalian cell lines (Thyagarajan *et al*., 2001).

In addition to chromosome integration, serine integrases have been widely applied in synthetic biology, mainly including metabolic pathway assembly (Colloms *et al*., 2013) and synthetic genetic devices (Bonnet *et al*., 2012, 2013; Siuti *et al*., 2013; Yang *et al*., 2014). Integrase-based DNA assembly exploits the ability of the integrase to recognise and recombine a pair of *attB* and *attP* sites with complementary central dinucleotides accurately and efficiently. By changing the central dinucleotides, six pairs of orthogonal target sites are available for use in the assembly system. Using this strategy, five-gene zeaxanthin pathway was assembled based on ϕC31 integrase. The productivity of pathways can be optimised with combinatorial assembly by varying gene orders and by varying RBS strengths. This system demonstrated that a chosen DNA sequence can be replaced by *attR*×*attL* recombination when provided with RDF (Colloms *et al*., 2013). Applications of serine integrases in construction of synthetic genetic devices have been introduced in section 1.2.

1.6 Synthetic biological devices based on recombinases

Synthetic biological devices and their performances have been described in the previous section (section 1.2). Here, the recombinase-based devices are summarised. The information processing devices were first built on invertases, FimB/FimE and Hin (Ham and Lee, 2006; Ham *et al*., 2008), and simple tyrosine recombinases Cre and Flp (Friedland *et al*., 2009). However the efficiencies and capabilities of these devices are limited because of the poor recombination directionality of these recombinases. Phage integrases mediate highly directional recombination, making them good candidates for recording information in DNA

sequences. Though both tyrosine and serine integrases have controllable reaction directionality, serine integrases are preferred due to their recognition of simple short sites and lack of requirement for any accessory host factor proteins for recombination.

Among the three outcomes (integration, resolution, and inversion) of site-specific recombination introduced in section 1.3.2, inversion recombination is commonly used in synthetic biological systems. The orientation of genetic parts between inverted repeat recombination sites can be flipped by a serine integrase alone, and its cognate Xis/RDF can stimulate the integrase to reverse the flipping. The practicality of this system has been proven in a rewritable recombinase addressable data module based on Bxb1 integrase and its RDF gp47 (Bonnet *et al*., 2012). Using more sets of integrase-RDFs and their corresponding recombination sites to increase the information storage capacity can scale up the complexity of these devices.

By combining two integrase controlled units, genetic logic gates were constructed (Bonnet *et al*., 2013; Siuti *et al*., 2013). These systems were programmed to logically respond to different combinations of two independent signals. More recently, a memory array which can record 1.375 bytes of information was constructed (Yang *et al*., 2014). This system consists of 11 integrase based switches that are orthogonal to each other, and demonstrates that recombinase-based modules can be layered without any interference with each other.

Apart from the well characterised and widely used RDFs Gp47 of Bxb1 (Ghosh *et al*., 2006) and Gp3 of ϕC31 (Khaleel *et al*., 2011), four other RDFs associated with corresponding bacteriophage integrases have been identified, which are Xis of ϕRv1 (Bibb *et al*., 2005), Orf7 of TP901-1 (Breüner *et al*., 1999), Gp3 of ϕBT1, and Gp25 of TG1 (Zhang *et al*., 2013). Without doubt, more sets of integrase-RDF are yet to be discovered, and they would be potentially ideal for expanding the data storage capacity of integrases-based devices.

1.7 Aims of the study and thesis outline

A binary counter can be constructed from modules (such as the flip-flops in electronic circuits) by taking the output of a preceding module as the input of a subsequent one. Each module is required to be able to switch between two states (“0” and “1” of a binary digit) in response to a single input signal, and to send an input signal to the next module when it

restores state to “0”. Construction of a binary counting system in living cells is expected to promote the development of cellular computers for a variety of biological applications.

Synthetic switches introduced in section 1.2.1 provide different design principles for construction of two-state devices, whose states are commonly represented as either a “0” or a “1”. Among these designs, the rewritable memory module based on serine integrase from Bonnet *et al*. (2012) has substantial advantages over others, owing to its ability to record transient input signal into permanent DNA change, and highly directional recombination to switch DNA states back and forth. The rewritable manner of this module realised increased data storage capacity, which is limited in the single write module in the recombinase-based counter from Friedland *et al*. (2009). However, this rewritable module needs two different input signals to trigger its switch, and it cannot be used to record multiple occurrences of a single event.

This study aims to construct an efficient integrase-based binary counting module, which can alternate between two output states (“0” and “1”) in response to a single repeating input pulse. The bidirectional recombination system of large serine integrase was selected to achieve this aim. In the rewritable memory module (Bonnet *et al*., 2012), expressions of proteins (integrase and integrase plus Xis/RDF) for both directional reactions are controlled by two different input signals. In contrast, in this study the expression of integrase is controlled by only one input signal, while the expression of RDF is controlled by the binary state (e.g. when the state is “0”, RDF expression is off). If the initial state of this device is “0”, an integrase pulse stimulated by a pulse of input signal will change the device state from “0” to “1” through $attB \times attP \rightarrow attR \times attL$ recombination, allowing the expression of RDF. The RDF can alter the recombination directionality of integrase, and the next integrase pulse will change the device state from “1” to “0” through $attR \times attL \rightarrow attB \times attP$ recombination.

A multiple-digit counter can be constructed to count a large number through connecting multiple modules based on orthogonal sets of integrase-RDF (Fig. 1.16). The state of each module can be indicated by coupling its state with reporter protein expression (OFF=0 and ON=1). Each module is expected to generate a new input signal to trigger the state change of the next module only when it switches from state “1” to “0”. This combinatorial counter

acts like a frequency divider that takes an input signal of a frequency f_{in}, and generates an output (or a new input signal) of a frequency: $f = f_{in}/2^N$ (N is the number of modules).

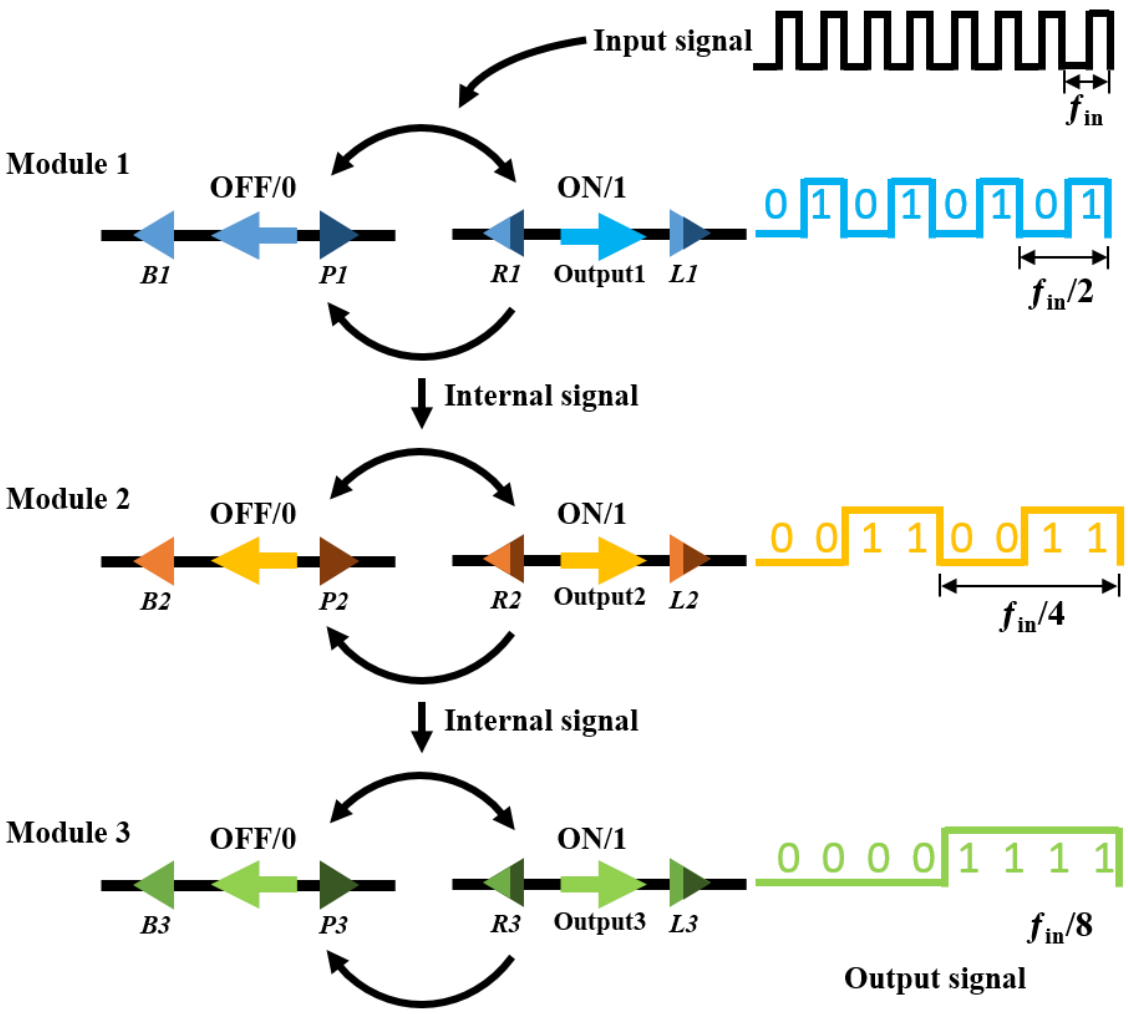


Figure 1. 15 A binary counter based on DNA inversion. A 3-bit binary counter consisting of three binary counting modules is shown. Each module contains an orthogonal set of recombination sites (B1-3 and P1-3) flanking a functional fragment to indicate the state change and generate a signal to control the subsequent module. The expression of the recombinase for Module 1 is controlled by an input signal. Module 1 is designed to flip back and forth between "0" and "1" with repetition of a single input signal. When a given module restores its state from "1" to "0", a new signal will be produced and sent to the subsequent module. When using each module to control the production of a distinct reporter signal, the recorded number can be indicated by the different combinations of reporter output signals. The frequency of output signal (or input signal for subsequent module) from each module is represented by square wave figure with binary digit number.

To construct the integrase-based genetic module mentioned above, the first stage of this study was to characterise the recombination properties of the integrase and its RDF chosen in this study, which will be described in Chapter 3.

Next, a series of plasmid-based gene circuits (substrate sequence which can be recombined was placed on the plasmid) with memory function was constructed. The controlled bidirectional inversion recombination was first tested in a set-reset latch, in which integrase and integrase plus RDF produced from two independent inducible systems mediated the state change of the latch. Then, a state-based latch was constructed, in which the RDF expression was connected to the state of the latch. On this basis, a binary counting module was finally developed that cell populations containing this binary counting module switched repeatedly between two distinct states in response to a single repeating input signal. All these devices will be described in Chapter 4.

Finally, to further improve the state consistency of individual cells, the invertible substrate sequence was delivered to the chromosome of a bacterial host cell as a single copy, resulting in the chromosomal state-based latch and the chromosomal binary counting module. These will be described in Chapter 5.

2 Materials and Methods

2.1 Bacterial Strains

All bacterial strains used in this study are listed in the Table 2.1 below.

Table 2. 1 Bacterial Strains

Name	Genotype	Source and reference
AB1157	*thr-1, araC14, leuB6(Am), Δ(gpt-proA)62, lacY1, tsx-33, qsr'-0, glnV44(AS), galK2(Oc), λ⁻ , Rac-0, hisG4(Oc), rfbC1, mgl-51, rpoS396(Am), rpsL31(str^R), kdgK51, xylA5, mtl-1, argE3(Oc), thiE1*	Adelberg, 1962
DS941	AB1157 *recF lacIq lacZ*ΔM15.	Summers and Sherratt, 1988
MG1655Z1	F⁻ λ⁻ *ilvG⁻ rfb⁻50 rph⁻1* Z1(*Laci^q TetR Sp^R*)	Bohn *et al*., 2004
DS941Z1	DS941 Z1(*Laci^q TetR* Sp^R)	This study
DS957	DS941 *pepA::Tn5*	McCulloch *et al*., 1994
DH5	F⁻ Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (r_K^-, m_K^+) *phoA supE44* λ⁻ *thi*-1 *gyrA96 relA1*	F. Olorunniji
DH5α	DH5 ϕ80*lacZ*ΔM15	Invitrogen
DH5αZ1	DH5α Z1(*Laci^q TetR Sp^R*)	Lutz and Bujard, 1997
TOP10	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (*Str^R*) *end*A1 *nup*G	Invitrogen
Π3	F⁻ *supE hsdΔ5 thi* Δ(*lac-proAB*) Δ*thyA::(erm-pir116)[EmR]*	Demarre *et al*., 2005
MFDpir	MG1655 RP4-2-Tc*::[ΔMu1::aac(3)IV-ΔaphA-Δnic35-ΔMu2::zeo] ΔdapA::(erm-pir) ΔrecA*	Ferrières *et al*., 2010

2.2 Chemicals and buffer solutions

2.2.1 Chemicals

Sources of chemicals generally used are listed in Table 2.2 below.

Table 2. 2 Chemicals

Chemicals	Source
General chemicals, biochemicals, organic solvents	Sigma/Aldrich, BDH, May and Baker
Medium	Difco and Oxoid
Agarose, acrylamide	FMC, Biorad, and Flowgen
Restriction enzyme buffers and ligase buffer	NEB
PCR buffer	NEB, Promega, and Stratagene

2.2.2 Buffer solutions

Standard solutions are listed in Table 2.3 below.

Table 2. 3 Buffer Solutions

Buffer solutions	Composition
Agarose gel loading dye	30% glycerol, 3 mM Tris, 0.3 mM EDT A, 0.5% SDS and 0.25% bromophenol blue
Ammonium persulphate (APS)	10% ammonium persulphate in water made fresh before each use.
Commercial enzyme buffers	As supplied with enzymes
Digest buffer Z4BSA	100 mM KCl, 20 mM $MgCl_2$, 0.12 mg/ml BSA, and 2 mM DTT
E buffer	2 M Tris base, 1.25 M acetic acid, 0.05 M EDTA (pH 8.0), and 1 M sodium acetate
Formamide loading buffer	80% deionised formamide, 10 mM EDTA (pH 8.0), 1 mg/ml xylene cyanol, 1 mg/ml bromophenol blue
Integrase dilution buffer (IDB)	25 mM Tris-HCl (pH 7.5), 1 mM DTT, 1 M NaCl, and 50% (v/v) glycerol
Integrase reaction buffer (1.43× IRB5)	71.5 mM Tris-HCl (pH 7.5), 0.143 mM EDTA (pH 8.0), 7.15 Mm Spermidine, 0.143 mg/ml BSA, and deionized water
In vitro recombination reaction gel loading buffer SDS/ protease K/ EDTA (SKE)	25mM Tris-HCl (pH 8.2), 20% (w/v) Ficoll, 0.5% sodium dodecyl sulphate, 5 mg/ml protease K, 0.25 mg/ml bromophenol blue, and 70 mM EDTA
L-Broth (Luria broth)	1 % (w/v) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 0.5% (w/v) NaCl, adjust to pH 7.5 with NaOH.
Phosphate-buffered saline (PBS)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, adjust to pH 7.4 with HCl
SOC	2% (w/v) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, and 20 mM glucose.
TAE (50x)	2 M Tris base, 1 M acetic acid, 0.05 M EDTA (pH 8.0)
TBE (10x)	0.89 M Tris base, 0.89 M boric acid, and 20 mM EDTA
TE	10 mM Tris-HCl (pH 8.0) and 1 mM EDTA (pH 8.0)
$TE_{0.1}$	10 mM Tris-HCl (pH 8.0) and 0.1 mM EDTA (pH 8.0)

All buffer solutions were made up in deionized or double distilled water unless otherwise stated.

2.3 Antibiotics

Antibiotics were used in liquid or solid medium for selective growth. The stock and used concentrations are given in Table 2.4.

Table 2. 4 Antibiotics

Antibiotics	Stock concentration	Final concentration in medium
Kanamycin (Kan)	25 mg/ml in H_2O	25 μg/ml
Ampicillin (Amp)	100 mg/ml in H_2O	100 μg/ml
Chloramphenicol (Cm)	25 mg/ml in Ethanol	25 μg/ml
Spectinomycin (Spec)	25 mg/ml in H_2O	50 μg/ml
Streptomycin (Strep)	10 mg/ml in Ethanol	50 μg/ml

2.4 Repressors and Inducers

Rrepressors and inducers used to control the regulated promoters are listed in Table 2.5.

Table 2. 5 Inducers and repressors

Inducers	Stock concentration	Final concentration in medium	Reference
glucose	20% (w/v)	0.2% (w/v)	Guzman *et al*., 1995
arabinose	10% (w/v)	0.2% (w/v)	Guzman *et al*., 1995
anhydrotetracycline (aTc)	50 μg/ml	100 ng/ml	Lutz and Bujard, 1997

2.5 Oligonucleotides

All oligonucleotides used for sequencing and plasmid construction were synthesised commercially by Eurofins MWG Operon. All of the oligonucleotides were dissolved in TE buffer to a concentration of 100 pmol/μl and stored at -20 ℃. Details (name, length, sequence, and purpose) of each oligonucleotide are listed in Table 2.6.

Table 2. 6 Oligonucleotides designed in this study

Oligo name	Length	Sequence (5'→3')	Purpose
SacI-RBS-NdeI-phiC31-F	43 bp	AAAGAGCTCAGGAGTAACATATGGACACGTACGCGGGTGCTTA	Forward primer used to amplify ϕC31 *int* from pARM050
SphI-XbaI-SpeI-phiC31-R	51 bp	AAAGGATCCGCATGCTCTAGATTATTAACTAGTCGCCGCTACGTCT TCCGT	Reverse primer used to amplify ϕC31 *int* from pARM050
SacI-PstI-RBS-gp3-F	49 bp	AAAGAGCTCCTGCAGAGGAGGATTACAAAATGGCGAAGCGTTCG ATCTG	Forward primer used to amplify ϕC31 *gp3* from pEY301
MluI-SpeI-gp3-R	41 bp	AAAACGCGTTTATTAACTAGTGTCGGCAATCGCGTCGTTGT	Reverse primer used to amplify ϕC31 *gp3* from pEY301
MluI-gpf-F	37 bp	AAAACGCGTTACTAGAGTCACACAGGAAAGTACTAGA	Forward primer used to amplify *gfp* from bio-brick K145205
EcoI-SalI-gpf-R	42 bp	AAAGAATTCGTCGACTTATTAAGCTACTAAAGCGTAGTTTTC	Reverse primer used to amplify *gfp* from bio-brick K145205
MluI-rfp-F	32 bp	AAAACGCGTGTGAGCGGATAACAATTTCACAC	Forward primer used to amplify *rfp* from bio-brick J04450
SalI-rfp-R	36 bp	AAAGTCGACTTATTAGCTAGCACCGGTGGAGTGACG	Reverse primer used to amplify *rfp* from bio-brick J04450
ATG-to-GTG-Int-Top	31 bp	GCTCAGGAGTAACATGTGGACACGTACGCGG	Top oligo used to site directed mutate ATG to GTG for Int
ATG-to-GTG-Int-Bot	31 bp	CCGCGTACGTGTCCACATGTTACTCCTGAGC	Bottom oligo used to site directed mutate ATG to GTG for Int
RBS-Int-Mut1-Top	20 bp	CCAGCTCGAGATGATTAACA	Top oligo used to replace the RBS for Int
RBS-Int-Mut1-Bot	26 bp	TATGTTAATCATCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
RBS-Int-Mut2-Top	20 bp	CCAGCTCGAGATTATTAACA	Top oligo used to replace the RBS for Int
RBS-Int-Mut2-Bot	26 bp	TATGTTAATAATCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
RBS-Int-Mut3-Top	20 bp	CCAGCTCGAGCTTCTTAACA	Top oligo used to replace the RBS for Int
RBS-Int-Mut3-Bot	26 bp	TATGTTAAGAAGCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
RBS-Int-Mut4-Top	20 bp	CCAGCTCGAGAAGAGAAACA	Top oligo used to replace the RBS for Int
RBS-Int-Mut4-Bot	26 bp	TATGTTTCTCTTCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int

RBS-Int-Mut5-Top	20 bp	CCAGCTCGAGAGGAGAAACA	Top oligo used to replace the RBS for Int
RBS-Int-Mut5-Bot	26 bp	TATGTTTCTCCTCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
RBS-Int-Mut6-Top	20 bp	CCAGCTCGAGAGAAGGAACA	Top oligo used to replace the RBS for Int
RBS-Int-Mut6-Bot	26 bp	TATGTTCCTTCTCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
RBS-Int-Mut7-Top	20 bp	CCAGCTCGAGAAGAAGAACA	Top oligo used to replace the RBS for Int
RBS-Int-Mut7-Bot	26 bp	TATGTTCTTCTTCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
RBS-Int-Mut8-Top	20 bp	CCAGCTCGAGGAGGTGGCCA	Top oligo used to replace the RBS for Int
RBS-Int-Mut8-Bot	26 bp	TATGGCCACCTCCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
RBS-Int-Mut9-Top	20 bp	CCAGCTCGAGGAGGTAATCA	Top oligo used to replace the RBS for Int
RBS-Int-Mut9-Bot	26 bp	TATGATTACCTCCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
RBS-Int-Mut10-Top	20 bp	CCAGCTCGAGAGGAGGTACA	Top oligo used to replace the RBS for Int
RBS-Int-Mut10-Bot	26 bp	TATGTACCTCCTCTCGAGCTGGAGCT	Bottom oligo used to replace the RBS for Int
Ptet-mcs-top	46 bp	AGCTTCTCGAGCTGCAGGGATCCTCTAGAACGCGTGGTACCATGCA	Top oligo used to insert multiple cloning sites into cloning vector pDV
Ptet-mcs-bot	38 bp	TGGTACCACGCGTTCTAGAGGATCCCTGCAGCTCGAGA	Bottom oligo used to insert multiple cloning sites into cloning vector pDV
EcoI-ScaI-NotI-RBS-F	81 bp	AAAGAATTCAGTACTGCGGCCGCTTTCCCTCTAGATTTAACTTTAAGAAGGAGATATAAATATGGCGAAGCGTTCGATCTG	Foward primer used to amplify ϕC31 *gp3* from pEY301 to replace the optimized *gp3* on pSWITCH1
MluI-SpeI-gp3-R	41 bp	AAAACGCGTTTATTAACTAGTGTCGGCAATCGCGTCGTTGT	Reverse primer used to amplify ϕ C31 *gp3* from pEY301
Gfp-LAA-Top	57 bp	CTATAAGAGGCCTGCAGCAAACGACGAAAACTACGCTTTAGCAGCTTAATAAGGTAC	Top oligo used to add LAA degradation tag to Gfp
Gfp-LAA-Bot	57 bp	CTTATTAAGCTGCTAAAGCGTAGTTTTCGTCGTTTGCTGCAGGCCTCTTATAGAGCT	Bottom oligo used to add LAA degradation tag to Gfp

Gfp-AAV-Top	57 bp	CTATAAGAGGCCTGCAGCAAACGACGAAAACTACGCTGCAGCAG TTTAATAAGGTAC	Top oligo used to add AAV degradation tag to Gfp
Gfp-AAV-Bot	57 bp	CTTATTAAACTGCTGCAGCGTAGTTTTCGTCGTTTGCTGCAGGCCT CTTATAGAGCT	Bottom oligo used to add AAV degradation tag to Gfp
Int-LAA-Top	61 bp	CTAGTAGGCCTGCAGCAAACGACGAAAACTACGCTTTAGCAGCTT AATAATCTAGACTGCA	Top oligo used to add LAA degradation tag to Int
Int-LAA-Bot	63 bp	CTAGTAGGCCTGCAGCAAACGACGAAAACTACGCTTTAGCAGCTT AATAATCTAGACTGCA	Bottom oligo used to add LAA degradation tag to Int
Int-AAV-Top	61 bp	CTAGTAGGCCTGCAGCAAACGACGAAAACTACGCTTTAGCAGCTT AATAATCTAGACTGCA	Top oligo used to add AAV degradation tag to Int
Int-AAV-Bot	53 bp	CTAGTAGGCCTGCAGCAAACGACGAAAACTACGCTTTAGCAGCTT AATAATCTAGACTGCA	Bottom oligo used to add AAV degradation tag to Int
gfp-3'-out	21 bp	CTAGTAGGCCTGCAGCAAACGACGAAAACTACGCTTTAGCAGCTT AATAATCTAGACTGCA	Sequencing primer
gp3-out-seq	20 bp	TCTTGGACGCTCTTCAAGGT	Sequencing primer
T-P-attL-Top1	54 bp	CCGCAAAAAACCCCGCTTCGGCGGGGTTTTTTCGCGGGCCCTTGA CAGCTAGAT	Oligo with terminator, promoter and *attL* site to be inserted into pZJ50on
T-P-attL-Top2	84 bp	CAGTCCTAGGTATAATGCTAGAAGATCTCCGCGGTGCGGGTGCCA GGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGCGTAGGGG	As above
T-P-attL-Bot1	58 bp	CTAGCCCCTACGCCCCCAACTGAGAGAACTCAAGGGCACGCCCTG GCACCCGCACCGC	As above
T-P-attL-Bot2	88 bp	GGAGATCTTCTAGCATTATACCTAGGACTGATCTAGCTGTCAAGGG CCCGCGAAAAAACCCCGCCGAAGCGGGGTTTTTTGCGGCATG	As above
gfp-no-prom	43 bp	GTTGACATCGCCATCCAGTTCAACCAGGATCGGTACTACACCC	Top oligo for site directed mutagenesis of *gfp* in pZJ52off to remove promoters

gfp-no-prom-bott	43 bp	GGGTGTAGTACCGATCCTGGTTGAACTGGATGGCGATGTCAAC	Bottom oligo for site directed mutagenesis of *gfp*
gfp mut2 top	37 bp	GACGCTTTGGGTGGACAGGTAATGGTTATCCGGCAGT	Top oligo for site directed mutagenesis of *gfp* in pZJ52off to remove promoters
gfp mut2 bot	37 bp	ACTGCCGGATAACCATTACCTGTCCACCCAAAGCGTC	Bottom oligo for site directed mutagenesis of *gfp*
gfp mut3 top	38 bp	CCAGTTTATGCCCCAGAATGTTGCCATCTTCCTTAAAG	Top oligo for site directed mutagenesis of *gfp* in pZJ52off to remove promoters
gfp mut3 bot	38 bp	CTTTAAGGAAGATGGCAACATTCTGGGGCATAAACTGG	Bottom oligo for site directed mutagenesis of *gfp*
pZJ53-P1	18 bp	CGAAGTCAAAGCCGTCGC	Sequencing and PCR primer
pZJ53-P2	22 bp	CGCTTTGGGTGGACAGATAATG	As above
F-gp3-RBS-mut	50 bp	GCCTCTAGATTTAACTTTAAGNRRGNGRTATAAATATGGCGAAGCGTTCG	Forward primers with random RBS sequences used to amplify *gp3*.
R-gp3-RBS-mut	22 bp	GTTCGGCGCGTCGTACTCCTTC	Reverse primers used to amplify *gp3*.
F-gp47-RBS-mut	73 bp	GCGCCTCTAGATTCAACTTTAARNRGGNGRTATAAATATGACTCAGCGTATCGTC	Forward primers with random RBS sequences used to amplify *gp3*.
R-gp47-RBS-mut	20 bp	ATGAGGCCAGTCTTGTGCTC	Reverse primers used to amplify *gp3*.
Pswitch1	20 bp	TCCACTGAAGCTGCCATTTT	Sequencing primer
Pswitch2	20 bp	TTTACCATTCGTCGCGTCAC	Sequencing primer
Pswitch3	20 bp	GATGAATGACTGTCCACGAC	Sequencing primer
Pswitch4	20 bp	GGTTCGTTCTCATGGCTCAC	Sequencing primer

2.6 Custom DNA synthesis

In order to easily manipulate the regularly used sequence, synthetic DNA were made and cloned into a vector commercially by GeneArt (Life Technology), resulting in plasmids pSWITCH1 and pSWITCH2 (Fig. 2.1). About 5 µg plasmid with synthetic sequence was dissolved in 50 µl of distilled water and stored at -20 ℃. In synthetic plasmid pSWITCH1, the superfolder *gfp* and *gp3* genes are placed between inverted repeat recombination sites *attB* and *attP*. The $P_{Ltet0-1}$ promoter in an opposite orientation to *gfp* and *gp3* genes is placed outside of the recombination sites. The whole invertible sequence is flanked by two terminators (rnpB and part B0011) to inhibit transcription from surrounding sequence. In synthetic pSWITCH2, superfolder *gfp* gene is placed between inverted repeat recombination sites *attB* and *attP*. The constitutive promoter part J23104 in an opposite orientation to *gfp* gene is placed outside of the recombination sites. The functional fragments are separated by restriction enzyme sites. These sites make it easy to change each fragment in the sequence.

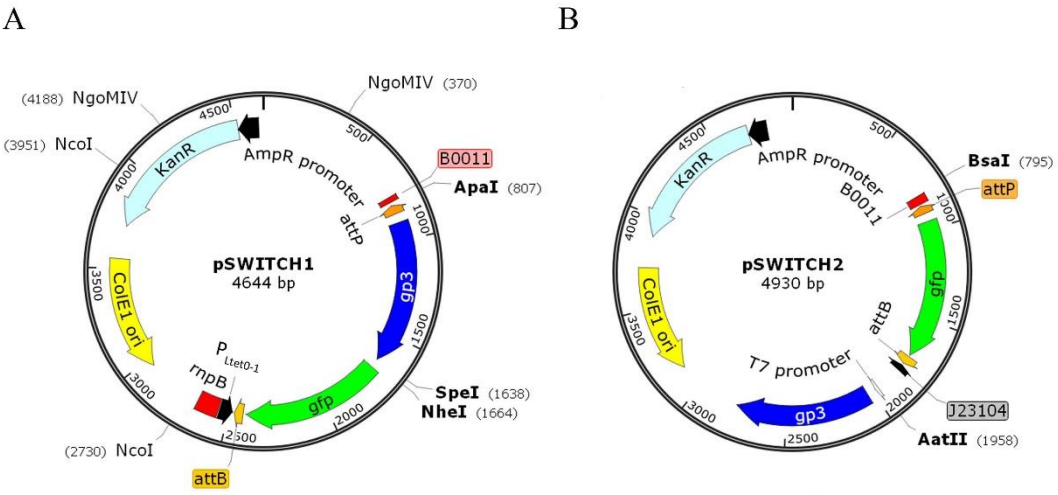


Figure 2. 1 Maps of synthetic plasmids. (A) Plasmid pSWITCH1 containing synthetic sequence used to construct state-based latch. (B) Plasmid pSWITCH2 containing synthetic sequence used to construct set-reset latch. Restriction sites used for the later construction are labelled on maps.

2.7 Plasmids

Plasmids used and constructed in this work are listed in Table 2.7. More details about the construction processes and characteristics of some plasmids will be introduced in the result chapters

Table 2. 7 Plasmids used and constructed in this study

Name	Antibiotic	Description (main characteristics and construction process)	Source
pSDC343	Amp	Plasmid containing *attP* site	S. Colloms
pSDC362	Amp	Plasmid containing *attB* site	S. Colloms
pEY301	Kan	Plasmid containing ϕC31 *gp3* encoding gene	E.Younge
pARM050	Amp	Plasmid containing ϕC31 *integrase* encoding gene	A. McEwan
pBAD33	Cm	Plasmid containing *araC* gene, P_{BAD} promoter, p15a origin of replication, derived from pACYC184	Guzman *et al*., 1995
pSDC404	Amp	Plasmid containing $P_{Ltet0-1}$ promoter, ColE1 origin of replication	S. Colloms
K145205	Cm	Biological part encoding GFP-LVA on BioBrick vector pSB1A2	IGEM 2010 kit plate 2,well 4D
J04450	Cm	Biological part encoding RFP on BioBrick vector pSB1C3	IGEM 2010 kit plate 1,well 3A
pCR2.1-TOPO	Amp, Kan	TOPO-TA cloning vector, pUC origin	Invitrogen Life Technologies
pDV	Kan	Plasmid containing $P_{Ltet0-1}$ promoter, pSC101 origin of replication	D.Houston
pCER19	Amp	Plasmid containing *cer* site of ColE1, pMB1 origin of replication, derived from pUC19	Caryl *et al*., 2004
pMS183	Kan	Cloning vector containing pSC101 origin of replication	M. Stark
pUC19	Amp	Cloning vector containing pMB1 origin of replication	Yanisch-Perron *et al*., 1985
pFM16	Amp, Kan	Recombination substrate plasmid containing *attB* and *attP* sites in direct repeat flanking a *kan*^{r} gene	F. Olorunniji
pSWITCH1	Kan	Plasmid containing inverted repeat *attB, attP* sites, *gfp*, *gp3* and $P_{Ltet0-1}$,ColE1 origin of replication	Synthesied by GeneArt® , more details see section 2.6

pSWITCH2	Kan	Plasmid containing *attB* and *attP* sites in inverted repeat, *gfp*-LAA, *gp3* and T7 promoter, ColE1 origin of replication	Synthesied by GeneArt®, more details see section 2.6
pGP47-INV-RL	Amp	Plasmid containing Bxb1 *attR* and *attL* sites flanking a *gp47* gene and promoter J23119	S.Colloms
pFC31	Kan	Plasmid containing synthetic insertion element (sites ISY100 *IRR* and *IRL* flanking *Kan* resistance gene), ColE1 origin of replication	S.Colloms
pSW23T	Cm	Plasmid containing $oriV_{R6K\gamma}$ origin of replication and oriT transfer origin	Demarre *et al*., 2005
pGD01	Kan	Plasmid containing *attB* and *attP* sites in direct repeat flanking a *galK* gene, based on pMS183	K. Poop
pGD02	Kan	Plasmid containing *attR* and *attL* sites in direct repeat flanking a *galK* gene, based on pMS183	K. Poop
pZJ1	Kan, Amp	PCR2.1-TOPO vector + ϕC31 *integrase* gene	Chapter 3
pZJ2	Kan, Amp	PCR2.1-TOPO vector + ϕC31 *gp3* gene	Chapter 3
pZJ3	Kan, Amp	PCR2.1-TOPO vector + *gfp*-LVA gene	Chapter 3
pZJ7	Cm	Vector pBAD33+ ϕC31 *integrase* gene from pZJ1 (SacI, XbaI)	Chapter 3
pZJ7m	Cm	Plasmid containing ϕC31 *integrase* gene with GTG start codon, derived from pZJ7 (NdeI site is eliminated)	Chapter 3
pZJ20	Cm	Plasmid containing a predicted weaker RBS for ϕC31 *integrase*, derived from pZJ7 (RBS: ATGATTAA)	Chapter 3
pZJ21	Cm	Plasmid containing a predicted weaker RBS for ϕC31 *integrase*, derived from pZJ7 (RBS:ATTATTAA)	Chapter 3
pZJ22	Cm	Plasmid containing a predicted weaker RBS for ϕC31 *integrase*, derived from pZJ7 (RBS:CTTCTTAA)	Chapter 3
pDVZJ	Kan	Vector pDV+ Oligo Ptet-mcs (HindIII, PstI)	Chapter 3
pZJ14	Kan	Vector pDVZJ + ϕC31 *gp3* (MluI, PstI)	Chapter 3
pZJ24	Amp	Vector pUC19+ ϕC31 *gp3* (PstI, KpnI)	Chapter 3
pZJ123	Amp	Vector pSDC404 + ϕC31 *integrase* gene + ϕC31 *gp3* (*integrase* and *gp3* controlled by $P_{Ltet0-1}$ promoter)	Chapter 3

pZJ10off	Amp	*attP*, pMB1 ori from pSDC343 (AlwNI,SacI)+ *attB* from pSDC362 (AlwNI, EcoRI)+ *gp3* from pZJ2 (EcoRI, MluI)+ *gfp*-LVA from pZJ3 (MluI, SacI)	Chapter 3
pZJ12off	Amp	Derived from pZJ10off by deleting the *gp3* gene (MluI, PstI)	Chapter 3
pZJ15off	Amp	Derived from pZJ10off by changing *gfp*-LVA to *rfp* (SalI, MluI)	Chapter 3
pZJ16off	Amp	Derived from pZJ15off by deleting the *gp3* gene (MluI, PstI)	Chapter 3
pZJ16on	Amp	Derived from pZJ16off by recombination	Chapter 3
pZJ18off	Amp	Vector pCER19+ invertible fragment from pZJ16off (BamHI, NheI)	Chapter 3
pZJ18on	Amp	Derived from pZJ18off by recombination	Chapter 3
pZJ19off	Kan	Vector pMS183+ invertible fragment from pZJ18off (BglII, NdeI)	Chapter 3
pZJ19on	Kan	Derived from pZJ19off by recombination	Chapter 3
pZJ36off	Kan	Derived from pSWITCH1 by changing the codon optimized *gp3* to natural *gp3*	Chapter 4
pZJ38off	Kan	Vector pMS183+invertible fragment from pZJ36off (NcoI, NgoMI)	Chapter 4
pZJ38on	Kan	Derived from pZJ38off by recombination	Chapter 4
pZJ31off	Kan	Derived from pSWITCH2 by removing the degradation tag of GFP	Chapter 5
pZJ46	Cm	Plasmid containing ϕC31 *integrase* with LAA degradation tag, derived from pZJ7	Chapter 5
pZJ47	Cm	Plasmid containing ϕC31 *integrase* with AAV degradation tag, derived from pZJ7	Chapter 5
pZJ48	Cm	Plasmid containing ϕC31 *integrase* with LAA degradation tag, derived from pZJ7m	Chapter 5
pZJ49	Cm	Plasmid containing ϕC31 *integrase* with AAV degradation tag, derived from pZJ7m	Chapter 5
pZJ91	Cm	Plasmid containing a predicted stronger RBS for ϕC31 *integrase*, derived from pZJ46 (RBS: AGGAGGTA)	Chapter 5
pZJ94	Cm	Plasmid containing a predicted stronger RBS for ϕC31 *integrase*, derived from pZJ47 (RBS: AGGAGGTA)	Chapter 5
pZJ50on	Kan	Derived from pZJ38on by deleting *attL* site	Chapter 4
pZJ51on	Kan	Plasmid containing B1002+J23119+*attL* between *gp3* and *gfp*, derived from pZJ50on	Chapter 4

pZJ52on	Kan	Derived from pZJ51on by deleting part BBa_B1002	Chapter 4
pZJ53on	Kan	Plasmid containing mutant *gfp* (mut1), derived from pZJ52on	Chapter 4
pZJ56off	Kan	Plasmid containing invertible fragment from pZJ53off, based on pSWITCH1 (BsaI,AatII)	Chapter 3
pZJ56on	Kan	Plasmid containing invertible fragment from pZJ53on, based on pSWITCH1	Chapter 3
pZJ53off	Kan	Derived from pZJ53on by recombination	Chapter 4
pZJ61off	Kan	Derived from pZJ53off by changing the *gp3* to *gp47*, and recombination sites to the ones for Bxb1 integrase (EagI, NheI)	Chapter 4
pZJ68off	Kan	Derived from pZJ53off by inserting *tetR* gene, derived from pZJ53off (AatII, NcoI)	Chapter 4
pZJ72	Kan	Derived from pFC31 by inserting *Tnp* gene (KpnI, NdeI)	Chapter 5
pZJ74	Kan, Cm	Plasmid containing insertion element (sites *IRR*, *IRL*, and *Kan*^{r}) and *Tnp* from pZJ72, based on pSW23T	Chapter 5
pZJ76off	Kan	Plasmid containing invertible fragment from pZJ53off, based on pFC31 (AatII, BsaI)	Chapter 5
pZJ76on	Kan	Plasmid containing invertible fragment from pZJ53on, based on pFC31 (AatII, BsaI)	Chapter 5
pZJ78off	Kan, Cm	Derived from pZJ74 by replacing its insertion element using the one from pZJ76off (SalI, BamHI)	Chapter 5
pZJ79off	Kan	Plasmid containing a predicted weaker RBS for Gp3, derived from pZJ68off (RBS: AAGTGAGCGGTATAAAT)	Chapter 4
pZJ80off	Kan	Plasmid containing a predicted weaker RBS for Gp3, derived from pZJ53off (RBS: AAGTAAGTGGTATAAAT)	Chapter 4
pZJ81off	Kan	Inserting oligo containing restriction site SphI into pZJ78off. (AatII, SalI)	Chapter 5
pZJ82off	Kan	Plasmid containing *tetR* gene, derived from pZJ81off (SphI, AatII)	Chapter 5
pZJ118off	Kan	Plasmid containing invertible fragment from pZJ31off, based on the pMS183 vector (AatII, BsaI)	Chapter 3
pZJ118on	Kan	Derived from pZJ118off by recombination	Chapter 3

2.8 Bacterial growth medium

E.coli strains were routinely cultured in liquid L-Broth or on solid L-Broth agar from the MVLS Stores of University of Glasgow. Other specialized medium used are: X-gal agar and MacConkey galactose agar. X-gal agar plates were made by spreading 40 µl of X-gal (20 mg/ml in dimethyl formamide) and 40 µl of L-Broth on top of a normal L-Broth agar plate. MacConkey galactose agar was made by boiling 4 g premix carbohydrate-free MacConkey agar base (Difco) in 100 ml of deionized water and addition of 5 ml 20% galactose solution.

2.9 Bacterial growth conditions

E. coli strains in liquid L-Broth containing appropriate antibiotics were grown at 37 ℃ with shaking at 225 rpm (New Brunswick Scientific Excella E24 Incubator Shaker). For activation of promoters in strains, appropriate inducers were added into the liquid L-Broth. *E. coli* strains on solid L-Broth agar plates with appropriate antibiotics were incubated at 37 ℃ overnight. The X-gal agar was used for blue-white screening. The MacConkey galactose agar plates was used for *in vivo* recombination assays.

2.10 Generalised transduction in *E.coli* by bacteriophage P1

The MG1655ZI *E.coli* strain carries a module with two copies of Lac Repressor and Tet Repressor encoding genes driven by the constitutive promoters P_{laciq} and P_{N25}, respectively as well as the spectinomycin resistance gene at the bacteriophage λ *attB* locus. This module was transferred into the strain DS941 by P1 transduction to construct DS941Z1.

2.10.1 Preparation of P1 lysates

Donor cells were cultured overnight. 250 µl of the overnight culture was inoculate in 10 ml of fresh L-Broth and grown at 37 ℃ with shaking (225 rpm) to mid-log phase. 100 µl of culture was centrifuged and the resulting cell pellet was resuspended in 100 µl fresh L-Broth with 100 µl $CaCl_2$ (50 mM) and 100 µl $MgCl_2$ (100 mM). Sufficient P1 was added to give an m.o.i (multiplicity of infection) approximately 0.001. The mixture was then incubated at 37 ℃ for 30 min, mixed with 2.5 ml precooled (45 ℃) molten soft L- Broth agar (0.75% agar w/v) and plated onto a fresh, undried L-Broth agar plate. After overnight incubation at 37 ℃, 2.5 ml of phage buffer was added and left at room temperature for 15 min. The buffer

and soft agar was scraped off into a 30 ml centrifuge tube with about 300 µl chloroform and vortexed for 30 sec. The tube was left at room temperature for 30 min with intermittent vortexing and then centrifuged (10,000 rpm, 10 min) to collect the supernatant P1 lysate. The P1 lysate was then titred against *E. coli* strain to check its concentration.

2.10.2 P1 transduction

Recipient cells were cultured overnight. 250 µl of the overnight culture was inoculate in 10 ml of fresh L-Broth and grown at 37 ℃ with shaking (225 rpm) to mid-log phase. 500 µl of the cells were centrifuged (8,000 rpm, 2 min), resuspended in 500 µl fresh L-Broth, 500 µl $CaCl_2$ (50 mM) and 500 µl $MgCl_2$ (100nM). 300 µl of the mixture was transferred into a 1.5ml Eppendorf tube and 10 µl P1 lysate (10^8-10^{10} pfu/ml) was added. Mixtures without recipient cells or P1 lysate was prepared as controls. After 20 min incubation at 37 ℃, phage infection was stopped by adding 200 µl of 1 M sodium citrate. 500 µl of L-Broth was added to each tube and incubated at 37 ℃ for 1 h. The cells were then centrifuged (8,000 rpm, 2 min) and resuspended in 300 µl L-Broth and 70 µl sodium citrate. 200 µl of cells were plated on suitable selective plates. All transductants were streaked out to single colonies at least twice to remove any contaminating P1.

2.11 Preparation of competent cells and transformation.

Plasmid DNA was routinely transformed into *E.coli* by chemical transformation. Electroporation was used when very high transformation efficiency was needed.

2.11.1 Chemical transformation

250 µl of overnight *E. coli* culture was inoculated in 10 ml of fresh L-Broth containing appropriate antibiotics and grown at 37 ℃ with shaking (225 rpm) until OD_{600} of 0.4-0.5 was reached. The cells were harvested in a pre-cooled 30 ml centrifuge tube (8,000 rpm, 2 min, 4 ℃) and all supernatant was removed. The cell pellet was gently resuspended in 10 ml of ice-cold 50 mM $CaCl_2$, and the centrifugation was repeated after cells were left in cold $CaCl_2$ for 1 h. The cell pellet was gently resuspended in 1 ml of ice-cold 50 mM $CaCl_2$. The competent cells thus prepared were then left on ice until transformation or stored at -70 ℃ with about 17% glycerol. One Shot TOP10 Chemically Competent *E.coli* from Invitrogen was also used in the chemical transformation.

For each transformation, 0.5~2 µl of plasmid or ligation product (10 pg to 100 ng DNA) was added to 100 µl of $CaCl_2$ competent cells and incubated on ice for 20 min. The cells were heat shocked in 37 ℃ water bath for 5 min and returned to ice for 5 min. 100 µl of L-Broth was added to each transformation. The cells were then incubated at 37 ℃ for 60-90 min to allow expression of antibiotic resistance genes. 50 µl of liquid cultures for plasmid transformation or 150 µl of liquid cultures for ligation product transformation were spread on selective agar plates, which were then incubated at 37 ℃ overnight. For the commercial competent cells, transformation was carried out according to the manufacturer's instructions.

2.11.2 Electroporation

2.5 ml of overnight *E. coli* culture was inoculated in 200 ml of fresh L-Broth containing appropriate antibiotics and grown at 37 ℃ with shaking (225 rpm) until OD_{600} of 0.4-0.5 was reached. Cells were then cooled down on ice and harvested in a pre-cooled 250 ml centrifuge tube (Beckman Coulter JA-14, 6,000 rpm, 15 min, 4 ℃) and all supernatant was removed. The cell pellet was resuspended in 200 ml of ice-cold 10% glycerol and centrifugation was repeated. The cell pellet was resuspended in 150 ml of ice-cold 10% glycerol and centrifugation was repeated. The cell pellet was then gently resuspended in 20 ml of filter sterilised ice-cold 10% glycerol, transferred to a 30 ml centrifuge tube and centrifuged (Beckman Coulter J-20, 6,000 rpm, 15 min, 4 ℃). The final pellet was resuspended in a final volume of 500 µl of ice-cold 10% glycerol. The electro-competent cells thus prepared were either used for transformation straight away or stored at -70 ℃.

For each transformation, about 10 ng of de-salted DNA was added to 30 µl of electrocompetent cells on ice, mixed well and transferred to an ice-cold electroporation cuvette. The cuvette was placed into the slide of the Biorad micropulser which was set at EC2 according to the cell type and cuvette size. The electrical pulse was delivered and 1 ml of S.O.B was immediately added to the cuvette. The resulting liquid culture was transferred into a 15 ml centrifuge tubes and incubated at 37 ℃ with shaking (250 rpm) for more than 60 min. Then, cell cultures were centrifuged (8,000 rpm, 1 min) in a 1.5 ml Eppendorf tube, resuspended with 200 µl fresh L-Broth and spread on selective agar plates, which were then incubated at 37 ℃ overnight.

2.12 Conjugation

To transfer suicide plasmids into recipient strains which cannot sustain the replication of the plasmid, conjugation was used to obtain a higher transfer efficiency. The donor cells and recipient cells were cultured separately in 5 ml of L-Broth without antibiotic overnight. 250 µl of each culture was added onto the center of a well-dried agar plate and mixed all over the surface gently by a glass spreader. The plate was incubated at 37 ℃ overnight. After overnight incubation, cells were scraped by a toothpick and streaked on suitable selective plates to select the successfully conjugated recipients. The obtained cells were further confirmed by streaking on fresh selective plates.

2.13 Transposition

A suicide plasmid containing oriVR6Kγ origin, transposase, *kanamycin* resistance gene (*kan*) and invertible sequence for recombination was constructed and transformed into a pir^+ strain π3 which can sustain the replication of the plasmid. The plasmid was than transformed into a ΔdapA::pir donor strain MFDpir and conjugated into the pir^- strain DS941 or DS941Z1. The *Kan* gene and invertible sequence flanked by terminal inverted repeats *IRR* and *IRL* was transposed onto the chromosome of the recipient cells. Only the successfully conjugated and transposed recipients can grow on L-Broth agar plate with kanamycin antibiotic without diaminopimelic acid (DAP). More details of this process are introduced in chapter 5.

2.14 Plasmid DNA preparation

Plasmid DNA was prepared using QIAprep Spin Miniprep Kit (Qiagen). For each column, 3-4.5 ml of overnight cell culture was used routinely. Large volume (10 ml~20 ml) cells cultures from the early time course points were used because of the low cell concentrations. Plasmid DNA was isolated according to the manufacturer's instructions. The purified plasmid DNA was usually eluted in 50 µl of the supplied elution buffer.

2.15 Gel Electrophoresis

For routine separation of DNA (100 bp~10 kbp) resulting from plasmid minipreps or restriction digests, agarose gels were used. For purification of commercially synthesised oligonucleotides (20~60 bp) for site-directed mutagenesis, denaturing polyacrylamide gels were used.

2.15.1 Agarose gel electrophoresis

Small scale agarose gels (40 ml or 80 ml) were made of 1×TAE buffer and 1 % agarose. 2 µl of miniprep products or 20 µl of restriction digest products were mixed with 5 µl of loading buffer before loading onto a gel. Then the samples were run in 1×TAE buffer on a horizontal electrophoresis kit for 1.5 h at 80 V (80 ml gel) or 1 h at 100 V (40 ml gel). Large scale gels (250 ml) made of 1×E buffer and 1.2 % agarose were used to run DNA from the *in vivo* or *in vitro* site-specific recombination reactions. 2 µl of miniprep products or 20 µl of restriction digest products were mixed with 5 µl of loading buffer before loading onto a gel. Then, the samples were run in 1×E buffer on a horizontal electrophoresis kit for 16 h at 40 V or 20 V depending on gel length. Agarose gels were stained in their corresponding buffers by the addition of 0.5 µg/ml of Ethidium Bromide for more than 30 min and destained for 10 min, then, removed into a Bio-Rad UV Transilluminator and the Quantity One software was used for gel imaging and analyzing. The 1 Kb Plus DNA Ladder (Invitrogen) was for all agarose gel electrophoresis in this study.

2.15.2 Denaturing polyacrylamide gel electrophoresis

Denaturing 8% urea polyacrylamide gels were prepared by making 30 ml solution (6 ml of 40% acrylamide solution, 3 ml 10× TBE buffer, 14.4 g urea, distilled water). Then, 360 µl of 10% ammonium persulphate (APS) and 18 µl of tetramethylethylenediamine (TEMED) were added into the thoroughly mixed solution. The freshly made gel mixture was poured immediately between glass plates that were clamped together with a 1 mm spacer and a well-forming comb was inserted after pouring the gel. The gel was allowed to polymerise for 1 h and then pre-run at 400 V for 30 min in 1×TBE buffer in a vertical electrophoresis kit. Samples containing 1 nmol oligonucleotide solution and equal volume of formamide loading buffer were heated to 80 ℃ for 5 min prior to loading onto a gel. Then, the samples were run at 400 V for 1.5 h. The gel was stained in 70 ml dH_2O, 20 ml isopropanol, 10 ml 'StainsAll'(1-Ethyl-2-[3-(1-ethylnaphtho[1,2-d]thiazolin-2-ylidene-2-methylpropenyl]naphtho[1,2-d]thiazolium bromide, supplied by Sigma Aldrich) for 5 min and destained with distilled water for 5 min. Samples on the gel were visible by naked eye.

2.16 Extraction of DNA from gels

DNA fragments either on agarose gels or on polyacrylamide gels were extracted according to the methods below.

2.16.1 Extraction of DNA from agarose gels

Following electrophoresis and staining agarose gels, DNA fragment bands of the correct size were cut out and QIAquick Gel Extraction Kit (Qiagen) was used to isolate DNA from the gel according to the manufacturer's instructions. Extraction products were eluted in 30 µl of the supplied elution buffer. For isolation of monomeric plasmids, each single band containing the monomeric plasmids was cut out, put into a 0.45 µm Costar Spin-X Extraction column and centrifuged (10,000 rpm, 10 min) to collect the plasmid-containing flow-through. The plasmid DNA was then retransformed to obtain cells containing monomeric plasmids.

2.16.2 Extraction of DNA from polyacrylamide gels

Fragment bands of the correct size for the corresponding oligonucleotides were cut out from polyacrylamide gels following electrophoresis and staining. The gel chip was transferred into a 1.5 ml Eppendorf tube and crushed. For each tube, 0.5 ml of TE was added and incubated at 37 ℃ with shaking (1150 rpm) overnight (Eppendorf Thermomixer Compact). Then, the tube was centrifuged (10,000 rpm, 1min) and the supernatant was transferred to 0.22 µm Costar Spin-X Filter. The pellet in the tube was resuspended with 100 µl of TE, centrifuged and the supernatant was transferred to the same filter as previous supernatant. The filter was centrifuged (3,000 rpm, 1min) and the collection was vacuum dried to a final volume of 200 µl in an Eppendorf tube. The DNA samples were then purified by ethanol precipitation as described in section 2.17.

2.17 Ethanol precipitation of DNA

Oligonucleotides extracted from polyacrylamide gels or ligation products to be used for electroporation were further purified by ethanol precipitation. For each DNA sample, 0.22 volume of 5 M NH_4OAc and 2.5 volumes of cooled 100% ethanol were added (0.5 µl of 1 mg/ml tRNA was added if the DNA pellet was not obvious). The sample was thoroughly mixed and incubated at -20 ℃ for more than 1 h. The precipitated DNA was centrifuged (13,000 rpm, 1 h, 4 ℃) and the supernatant was removed. The pellet was washed with 1 ml of 70 % ethanol (80% ethanol for fragments smaller than 20 bp), centrifuged (13,000 rpm, 5 min, 4 ℃) and the supernatant was removed carefully. Then, the pellet was air dried and dissolved in 10 µl ddH_2O or TE 0.1.

2.18 Concentration measurement

The concentrations of cell cultures and oligonucleotides were measured using spectrophotometers (Thermo Biomate 3 UV-Vis Spectrometer or Perkin-Elmer Lambda 45 UV/Visible Spectrophotometer). The relative cell concentrations were determined from the optical density of the suspension at 600 nm wavelength. The oligonucleotide concentrations were calculated from the absorbance of the solution at 260 nm wavelength (absorbance of 1 at 260 nm contains 50 μg/ml double-stranded DNA).

2.19 Restriction endonuclease digestion of DNA

Restriction digest was carried out by restriction enzymes from NEB. 0.2-1 μg DNA was digested by addition of 10 units of enzyme in a 20 μl reaction according to the manufacturer's instructions. The digestion was incubated in a water bath at 37 °C (unless the manufacturer's instructions suggested otherwise) for more than 1 hour.

2.20 Ligation

Ligations were set up using T4 DNA ligase from NEB. About 1 μg of DNA fragments (the molar ratio of vector to insert was typically 1:3) were ligated with 200-400 unit of ligase in a 10 μl ligation reaction according to the manufacturer's instructions. The ligation was incubated at room temperature overnight and then used to transform competent *E. coli* cells.

2.21 Oligonucleotide insertion

Oligonucleotides were dissolved in $TE_{0.1}$ buffer. 10 μl top oligonucleotide (100 μM) was then mixed with 10 μl bottom oligonucleotide (100 μM) and 80 μl ddH_2O and the samples were incubated at 85 ℃ for 5 min and cooled down slowly to anneal the two strands. Products were diluted to different concentrations (typically 0.1 μM), then, ligated with the vector in a 10 μl ligation system.

2.22 Polymerase chain reaction (PCR)

PCRs were carried out with different polymerases in accordance with the manufacturer's instructions. Phusion High-Fidelity DNA Polymerase (NEB) was used in high fidelity DNA amplification. GoTaq DNA Polymerase (Promega) was used to add 3' deoxyadenosine to

blunt-ended DNA fragments for TA cloning and for detection of the size or orientation of the inserted sequence in the vector. PfuTurbo DNA Polymerase (Stratagene) was used in site mutagenesis and polishing reactions.

For general PCR reaction, a 50 µl reaction containing 0.2 mM dNTPs, an appropriate amount of template plasmid DNA (< 5 ng) or bacterial cells (take from colony using toothpick or 1 µl from overnight culture), and 0.5 µM of each primer was carried out with 1-2 units of polymerase in the 1 ×buffer supplied. For PCR reactions using Taq polymerase, $MgCl_2$ was added to a final concentration of 2.5 mM. A typical PCR reaction was denatured at 98 ℃ for 1 min and went through 30 cycles of 20 s at 98 ℃, 30 s at (Tm-10) ℃ and 1 min at 72 ℃. Finally, the reaction was incubated at 72 ℃ for 10 min. Reactions were run in a DNA Thermal Cycler (Eppendorf Mastercycler 5332). PCR products were run on a gel and purified for further usage.

Pfu turbo was used to make blunt ends from 5' or 3' overhangs generated by restriction enzyme. A 20 µl reaction containing 0.2 mM dNTPs and 10 µl of PCR DNA (10-500 ng) was carried out with 2.5 units of PfuTurbo DNA Polymerase in the 1 ×buffer supplied. The reaction was incubated at 72 ℃ for 30 min, and then, removed to ice. The blunt ended DNA fragments can be used directly in ligation reactions. PCR reactions used in TOPO TA cloning and Site-Directed mutagenesis are introduced below independently.

2.23 TOPO TA cloning

Some of the PCR amplified sequences were first put into the pCR2.1-TOPO vector by TOPO TA cloning for easily restriction digest or for sequencing. This was carried out using TOPO TA Cloning Kit (Life technologies). DNA was first amplified using Phusion polymerase (section 2.22). The PCR product was then treated with GoTaq DNA Polymerase (Promega) to add a single deoxyadenosine (dA) at the 3' end. 20 µl GoTaq reactions containing 0.2 mM dNTPs, 10 µl of purified PCR product (~ 200 ng DNA) and 2.5 µM $MgCl_2$ were carried out with 1-2 units of GoTaq polymerase in 1 × Colorless GoTaq Reaction buffer at 72 ℃ for 10 min. 4 µl of the product was mixed with the supplied 0.8 µl of PCR2.1 TOPO vector and 1 µl of salt solution and incubated at room temperature for more than 15 min, then, put on ice for 10 min. The TOPO Cloning Reactions were then transformed into the One Shot Top10 competent cells. Cells were grown on X-gal agar plates and white colonies were selected and analysed for insert.

2.24 Site-Directed mutagenesis

Site directed mutagenesis was carried out according to the Stratagene QuikChange site-directed protocol. The final concentration of purified oligonucleotides was measured using spectrophotometer. 50 µl reactions containing 0.2 mM dNTPs, 10 ng of template DNA and 125 ng of each primer were carried out with 2 units of PfuTurbo polymerase in 1×buffer. The reactions were denatured at 95 ℃ for 30 s and went through 14 cycles of 30 s at 95 ℃, 1 min at 55 ℃ and 1 min/kb of plasmid length at 68 ℃. 10 µl of the resulting PCR products was digested in a 20 µl digestion reaction containing DpnI buffer, 20 units DpnI and deionized water at 37 °C for more than 2 h and then 1 µl of the digest product was transformed into One Shot competent cells.

2.25 Sequencing

DNA samples were sequenced commercially by Eurofins MWG Operon. The samples were prepared according to the guide on MWG website.

2.26 *In vitro* recombination reactions

For the purpose of *in vitro* recombination, proteins (ϕC31 integrase and gp3) were purified as previously described (Khaleel *et al*., 2011; Smith *et al*., 2004), and provided by Dr Femi Olorunniji. Concentration and purity of the proteins were estimated by SDS-PAGE and measurement of the absorbance at 280 nm. Integrase and gp3 were diluted in integrase dilution buffer [25 mM Tris·HCl (pH 7.5), 1 mM DTT, 1 M NaCl, and 50% (v/v) glycerol] and kept at -20 ℃. Substrate plasmids used for *in vitro* assay were chemically transformed into strain *E. coli* DH5. Supercoiled plasmid DNA was purified from transformed DH5 cells, using a plasmid miniprep kit (Qiagen). DNA concentrations were estimated by measuring the absorbance at 260 nm.

All recombination reactions were performed in 1×IRB5 buffer. A typical reaction contained IRB5 buffer, 10 nM of substrate DNA and premixed proteins in various concentrations according to requirement. Reactions were carried out at 30 ℃ with reaction times ranging from 1 min to 3 h. Reactions were terminated by heating the sample to 80 ℃ and holding it at this temperature for 10 min. 8 µl of products were digested by adding 12 µl Z4BSA and appropriate restriction enzymes. The digested samples and the untreated samples were

analysed by gel electrophoresis. 5 µl of SKE loading buffer was added to each sample and incubated at 37 ℃ for 30 min prior to loading onto an agarose gel.

2.27 Bacterial induction conditions

2.27.1 Overnight induction

A single transformant colony was inoculated, or 5 µl overnight liquid culture was diluted 1:1000 in L-Broth (5 ml) containing appropriate antibiotics and inducers, and cultured at 37 ℃ with shaking (225 rpm) overnight (18 h, Fig. 2.2). This overnight culture could be diluted into fresh medium with appropriate inducers and cultured overnight to continue the induction process.

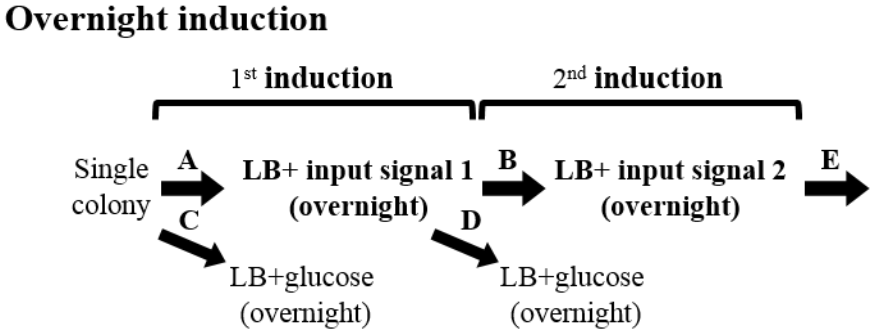


Figure 2. 2 Flow diagram showing the overnight induction process. For the first induction, a single colony was divided and added into medium medium with input signal 1 or with repressor (glucose), and cultured overnight (A, with input signal; C, with repressor). After overnight culturing, the second induction was carried out by diluting culture (1/1000) into medium with input signal 2 or with repressor, and cultured overnight (B, without input signal; D, with input signal). More inductions can be carried out using the same process (E).

2.27.2 Pulsed induction

A single transformant colony was inoculated in L-Broth (5 ml) containing appropriate antibiotics, and cultured at 37 ℃ with shaking (225 rpm) overnight (18 h). The overnight culture was diluted 1:40 in fresh L-Broth and precultured for 90 min or 120 min before adding inducers. After a specific induction time, cells were either frozen straightway or diluted 1:1000 in L-Broth without inducer and cultured at 37 ℃ with shaking (225 rpm) overnight (18 h, Fig. 2.3). The induction could be continued by diluting the overnight culture into fresh medium, precultured for 90 min or 120 min, and induced with the same or different inducers as the previous induction.

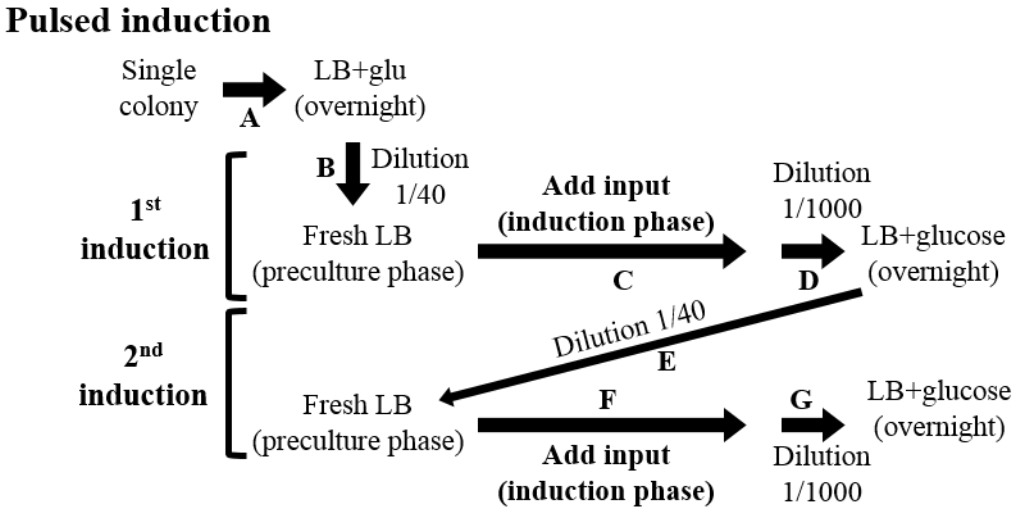


Figure 2. 3 Flow diagram showing the short pulse induction process. For the first induction time steps 1-4 were carried out. (A) A single colony was inoculated in medium with repressor (glucose) and cultured overnight. (B) The overnight culture was diluted (1/40) into fresh medium without input signal and precultured for 1.5 h or 2 h. (C) After the preculture phase, input signal was added into medium. (D) After a specific induction time, cell culture was diluted into fresh medium without input signal again and cultured overnight. For the second induction, the same steps (E, F, and G) as the first induction steps (B, C, and D) were carried out. The input signal for the second induction can be either the same as the first or different, depending on requirements.

2.27.3 Time course of induction recombination

2.27.3.1 Continuous induction time course

To detect the dynamic condition of recombination *in vivo*, cells were cultured and induced using the pulsed induction process (section 2.27.2). After the induction phase, cell cultures were centrifuged (10,000 rpm, 4 ℃) and frozen directly at different time points. Plasmids in the cells were extracted, digested, and checked by gel electrophoresis as previously described.

2.27.3.2 Short-pulse induction time course

To determine the optimal length of input pulse, cells were cultured and induced using the pulsed induction process (section 2.27.2). After the induction phase, cell cultures were diluted 1:1000 in L-Broth without inducer at different time points and cultured at 37 ℃ with shaking (225 rpm) overnight (18 h). Plasmids from the overnight culture were extracted, digested and checked by gel electrophoresis as previously described.

2.28 Cell fixation

1 ml of cells was centrifuged (8,000 rpm, 2 min) and washed twice with 1 ml 0.22 μm filtered PBS to remove the autofluorescent L-Broth. Cells were resuspended thoroughly in 900 μl PBS and 100 μl 40% (v/v) Formalin for 1 min at room temperature, then washed twice with 1 ml of PBS. The cells were finally resuspended in 1 ml of PBS and stored at 4 ℃ in the dark.

2.29 Cell fluorescence measurement

2.29.1 Typhoon Scanner

The concentrations of overnight cell cultures or fixed cells were checked by spectrophotometer. 100 μl of the same concentrated cell cultures or fixed cells were transferred into a 96 well plate and the fluorescence was measured by Typhoon FLA9500 laser scanner under fluorescence mode with method EGFP (473 nm laser and 510 nm long pass LPB filter). Pictures were taken under the condition of photomultiplier voltage 400 V, Pixel size 100 μm. The collected data was analysed by Typhoon fluorimager.

2.29.2 Blue light Transilluminator

Cell cultures were dropped on L-agar plates (10 μl), centrifuged in 1.5 ml Eppendorf tubes or streaked out on L-Broth agar plates. The tubes or plates were placed on a blue conversion screen for viewing the cells fluorescence by Bio-Rad blue light Transilluminator and the Quantity One software was used for imaging.

2.29.3 Flow cytometry

Single cell fluorescence was checked by CyAn ADP Analyzer. Fixed cells were diluted to about 10^6 cells/ml. 30000 events were acquired with about 500 e/s flow rate under 488 nm laser light. To place the cell population of interest on scale, acquisition parameters FSC, SSC and GFP was adjusted. The collected data was analysed by Flowjo program after gating cells by forward and side scatter.

2.29.4 Fluorescent microscopy

Fluorescent microscopy images were obtained by a Zeiss Axiovert 200M fluorescent microscope, using a 40x objective lens. 5 μl diluted (10x or 100x) or non-diluted fixed cell

culture was placed on glass slide with cover glass on top. GFP fluorescence of *E. coli* was visualized using filter set 10 (Zeiss) with excitation and emission filters of 450-490 nm and 515-565 nm, respectively. Images were analysed by Zeiss software.

2.30 MacConkey assay

To check the *in vivo* recombination efficiency, a "MacConkey recombination assay" was first developed by D. Blake (Blake, 1993). DS941 (*galK*⁻) *E. coli* cells were chemically transformed with substrate plasmids (containing *attB* and *attP* sites in direct repeat, flanking a *galK* gene) and integrase expression plasmids (containing the ϕC31 *integrase* gene under the control of P_{BAD} promoter). Cells were cultured in L-Broth overnight at 37 ℃ either with inducer or without. 1 ml of overnight cell culture was centrifuged, washed with phage buffer to get rid of the inducers, plated out onto selective MacConkey agar plate (approximate 100 colony forming unit), and incubated overnight at 37 ℃. After incubation the plates were photographed using a digital camera (Canon EOS D30). Colonies on the plate were washed off by addition of 1 ml of L-Broth, diluted 1:1000 by adding 5 μl into 5 ml L-Broth containing selective antibiotics and cultured overnight. Plasmid DNA was isolated from the overnight culture and analysed by agarose gel electrophoresis.

2.31 Production of substrate containing *attR* and *attL*

To create *attR*-*attL* substrate plasmids, substrate plasmids containing *attB* and *attP* constructed by routine cloning methods (DNA synthesis, PCR, restriction digest, and ligation) were transformed into cells containing integrase expression plasmids. After induction of the *attB*×*attP* recombination, plasmid DNA was isolated from cells, diluted 20 times and retransformed into competent cells. Transformants containing only the resistance marker from the substrate plasmid were selected. Alternatively, the recombination can be conducted *in vitro* with a purified integrase protein. Reaction was diluted 20 times and transformed into competent cells. DNA extracted from the selected transformants were further checked by restriction digest to ensure they were correct substrate plasmids. The same method can be applied to achieve *attB*-*attP* substrate plasmids from substrate plasmids containing *attR* and *attL*, but via recombination in the presence of integrase and RDF.

2.32 Random library construction and selection

Two random libraries of RBSs for Gp3 were constructed during this work. The details of construction and selection processes are introduced in section (4.3.3)

2.33 Quantification by Quantity One

Quantity One software was used to quantitate the gel bands gel images taken by Bio-Rad UV Transilluminator. A volume boundary that follows the outer edge of the object (band or pellet) was created by volume tool and the intensity data inside the boundary was compared with others using Volume Analysis Report with local Background Subtraction. The relative concentration of the DNA sample was achieved by dividing its intensity data by the size of DNA fragment in bp. The proportion of a DNA sample can be further calculated out. For example, a mixture of restriction digest product from plasmid A and B is run on an agarose gel and the bands were quantitated to collect their intensity data. Ai and As are used to represent the intensity data and the size of one band from A, and Bi and Bs are used to represent the intensity data and the size of one band from B. The proportion of DNA sample A among the total DNA can be calculated according to the equation 2.1

$$\%\ of\ A = \frac{\frac{Ai}{As}}{\frac{Ai}{As}+\frac{Bi}{Bs}} \times 100\% \qquad 2.1$$

3 Preliminary characterisation and testing of ϕC31 integrase inversion recombination system

3.1 Introduction

ϕC31 integrase (Int) belongs to the large serine integrase family of site-specific recombinases, and its normal role is to integrate the circular phage genome into its bacterial (*streptomyces lividans* and *streptomyces ambofaciens*) host genome through site-specific recombination (Kuhstoss and Rao, 1991). This requires recombination between *attP* (on the phage DNA) and *attB* (on the bacterial chromosome), producing two new attachment sites, *attL* and *attR* which are no longer substrates for Int alone. However, in the presence of ϕC31 gp3, a phage encoded accessory protein known as the Recombination Directionality Factor (RDF), Int will promote the excision reaction between *attL* and *attR*, to excise the prophage DNA and allow it to enter the lytic lifestyle (Fig. 3.1A, Azaro and Landy 2002; Khaleel *et al*., 2011).

As has been described in chapter 1, site-specific recombination results in integration, excision or inversion, depending on the location and orientation of recombination sites (Gellert and Nash, 1987). When the recombination sites *attB* and *attP* are placed on the same DNA molecule in inverted repeat, Int would be expected to mediate an inversion reaction, resulting in *attR* and *attL* sites and changing the orientation of the intervening sequence. Similarly, Int with Gp3 would be expected to catalyse an inversion reaction between *attR* and *attL* in inverted repeat, resulting in *attB* and *attP* and reversing the orientation of the intervening sequence (Fig. 3.1B). This inversion recombination system will be used as the basis for the synthetic circuits constructed in this study.

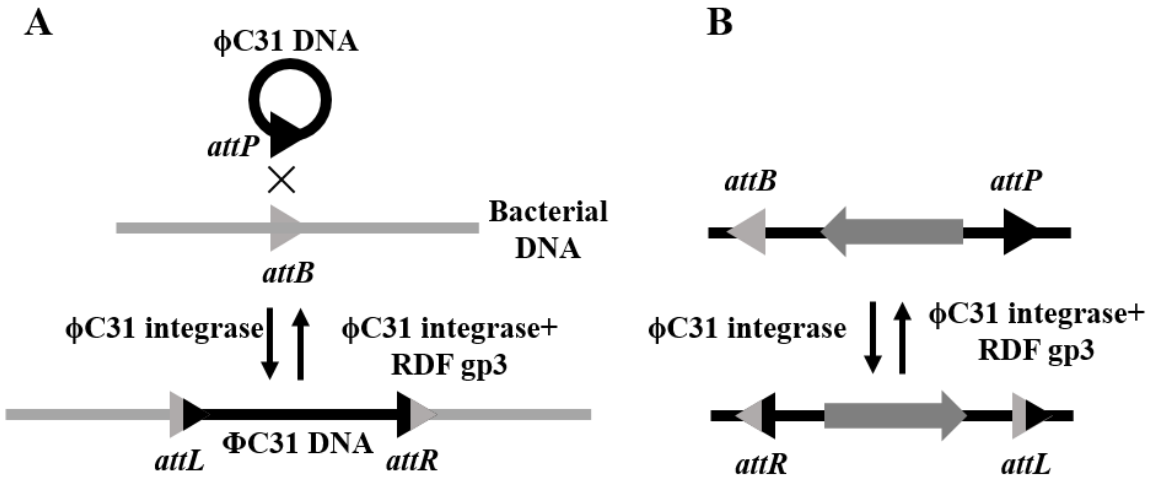


Figure 3. 1 Mechanism of ΦC31 integration, excision and inversion reaction. (A) ϕC31 Int-mediated integration and excision reaction in a natural system. (B) ϕC31 Int-mediated inversion reaction in an engineered system.

This chapter will discuss the experiments used to characterise and test the inversion reaction catalysed by ϕC31 integrase. In the first part of the chapter, *in vitro* assays using purified Int (or Int with Gp3) are presented. Later in the chapter, *in vivo* assays using regulated expression of Int (or Int with Gp3) in *E. coli* are reported.

3.2 Investigations of the inversion reaction mediated by ϕC31 integrase *in vitro*

Previously, *in vitro* integration and deletion recombination reaction mediated by ϕC31 integrase, and the effects of the RDF gp3 have been extensively studied in Smith's group (Khaleel *et al*., 2011; Rowley *et al*., 2008; Smith *et al*., 2004; Thorpe and Smith, 1998). However, inversion reactions have not been studied widely. To test if the inversion recombination reactions have similar efficiency to the integration and deletion reactions, inversion assays were carried out *in vitro* using purified ϕC31 integrase and RDF gp3, which were provided by Dr Femi Olorunniji. For the *attB*×*attP* reaction, the substrate plasmid pZJ56off (section 2.7) was constructed which had the following characteristics: *attB* and *attP* in inverted repeat flanking a constitutive promoter pointing away from the *gfp* gene outside of recombination sites, carried on a kanamycin resistant, high copy number plasmid with the ColE1 origin for easy plasmid preparation, and convenient restriction sites (EcoRI and KnpI) for easy analysis of recombination results. For the *attR*×*attL* reaction, the substrate plasmid pZJ56on (section 2.7) containing *attL* and *attR* sites in inverted repeat was made by *in vitro* recombination of pZJ56off (as described in section 2.31). The orientation of the

sequence (containing a promoter) flanked by the inverted repeat recombination sites will be flipped after the inversion reaction and the restriction digest pattern of the substrate plasmid will be changed (Fig. 3.2).

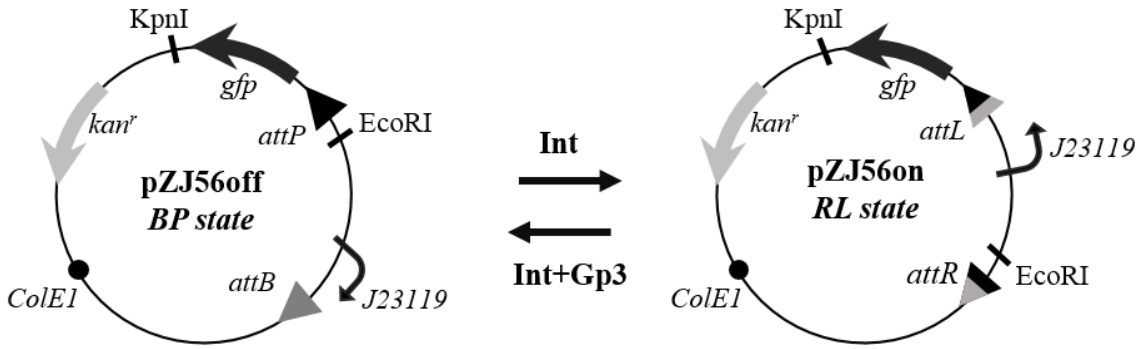


Figure 3. 2 Expected change of substrate plasmids upone inversion recombination. The substrate plasmid pZJ56off is recombined to pZJ56on by Int alone. The substrate plasmid pZJ56on is recombined to pZJ56off by Int in the presence of Gp3. Plasmid pZJ56off generates two fragments (size: 806 bp and 3792 bp) and pZJ56on generates two fragments (size: 1647 bp and 2951 bp) after restriction digestion with KpnI and EcoRI. The drawing is not to scale.

3.2.1 Effects of ϕC31 integrase and Gp3 concentrations on recombination efficiency *in vitro*

3.2.1.1 Reactions with substrate plasmid containing *attB* and *attP*

To study the effects of protein concentrations on *attB*×*attP* reaction efficiency, *in vitro* recombination reactions were carried out on substrate plasmid pZJ56off (Fig. 3.2) with varying concentrations of Int and Gp3 (Fig. 3.3). Each reaction was performed in a total reaction volume of 60 µl (as described in section 2.26) for 3 hours which was thought to be enough time for the reaction to reach steady state (Khaleel *et al*., 2011).

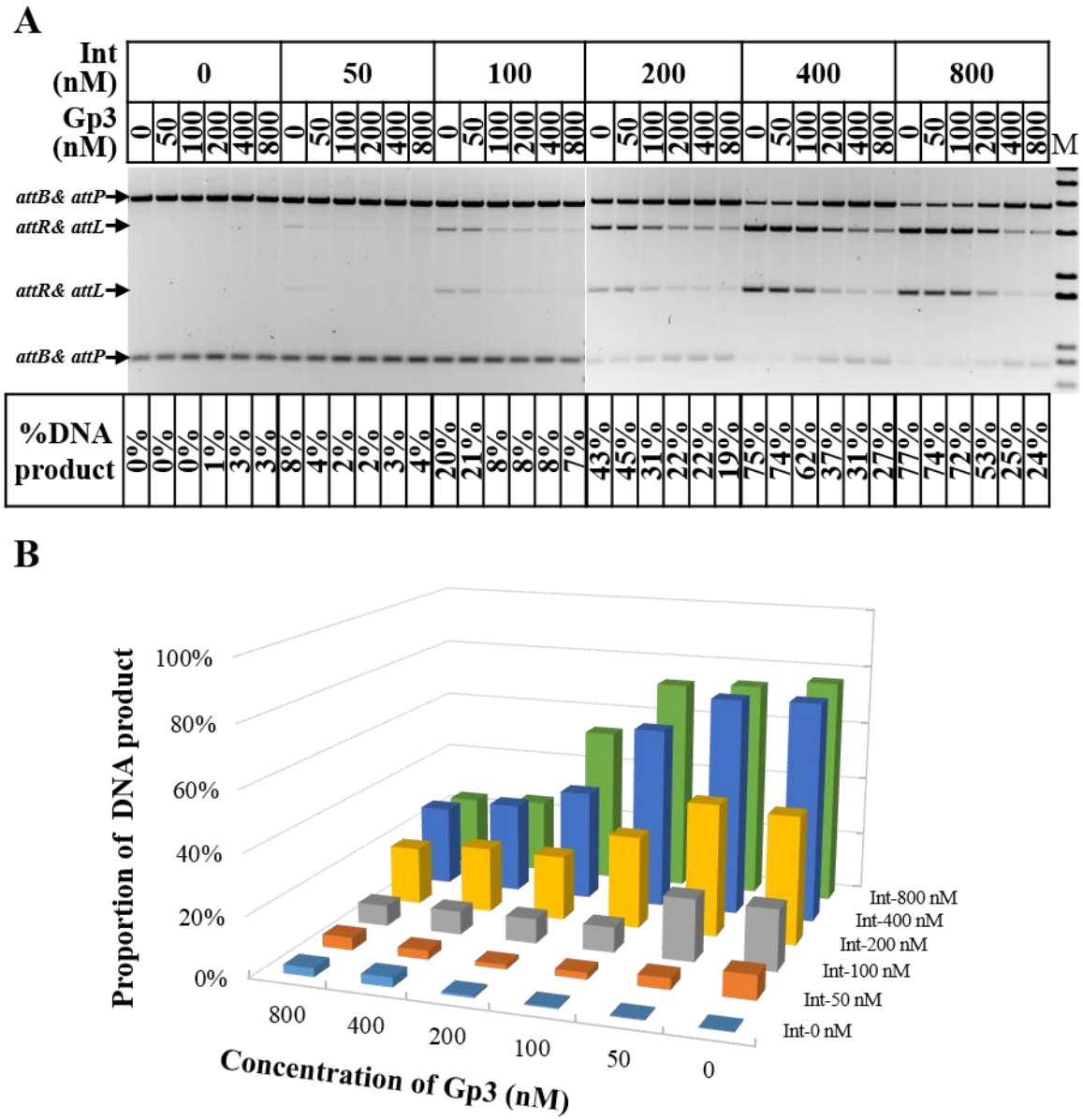


Figure 3. 3 attB×attP in vitro reaction. The substrate plasmid pZJ56off (*attB*, *attP*, ~10 nM) was incubated with different concentrations of Int and Gp3 *in vitro* for 3 hours. (A) Gel electrophoresis analysis of *in vitro* *attB*×*attP* reaction. The product pZJ56on (*attR*, *attL*) was separated from the substrate pZJ56off (*attB*, *attP*) by agarose gel electrophoresis after restriction digestion with KpnI and EcoRI. Band intensities were quantified using Quantity One, and the percentages of product (*attR* and *attL*) in total DNA were calculated (as section 2.33), and shown in table below the gel picture. (B) Graph showing the percentages of recombinant product under different protein concentrations.

3.2.1.2 Reactions with substrate plasmid containing *attR* and *attL* sites

To study the effects of protein concentrations on *attR*×*attL* reaction efficiency, *in vitro* recombination reactions were carried out on substrate plasmid pZJ56on (Fig. 3.2) with varying concentrations of Int and Gp3 (Fig. 3.4). Each reaction was performed in a total reaction volume of 60 µl (as described in section 2.26) for 3 hours.

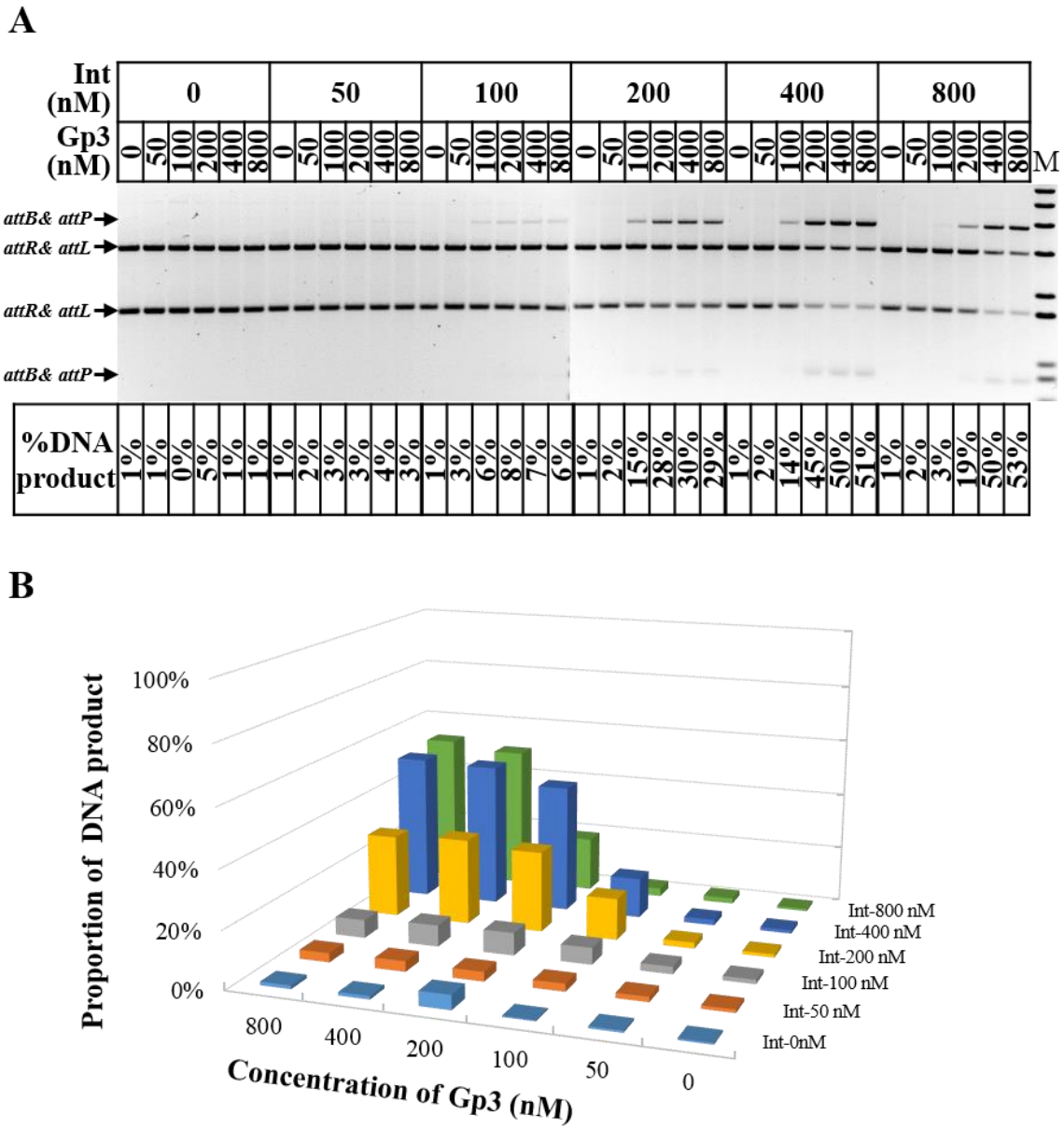


Figure 3. 4 attR×attL in vitro reaction. The substrate plasmid pZJ56on (*attR*, *attL*, ~10 nM) was incubated with different concentrations of Int and Gp3 *in vitro* for 3 hours. (A) Gel electrophoresis analysis of *in vitro* *attL*×*attR* reaction. The product pZJ56off (*attB*, *attP*) was separated from the substrate pZJ56on (*attR*, *attL*) by agarose gel electrophoresis after restriction with KpnI and EcoRI. Band intensities were quantified using Quantity One, and the percentages of product (*attB* and *attP*) in total DNA were calculated (as section 2.33), and shown in table below the gel picture. (B) Graph showing the percentages of recombinant product under different protein concentrations.

The *attB*×*attP* reaction only took place if Int was present (Fig. 3.3). In the absence of Gp3, the yield of recombination product increased as the Int concentration was increased, saturating at about 400 nM Int with an efficiency of 75 % (Fig. 3.3). However, when Gp3 was added, the recombination between *attB* and *attP* was inhibited, and this inhibition became particularly obvious when the ratio of Gp3 to Int exceeded 1:2 (e.g. the efficiency decreased to about 25 % when 800 nM Int and over 400 nM Gp3 were present, Fig. 3.3).

The *attR*×*attL* reaction was only observed in the presence of both Int and Gp3 (Fig. 3.4). The most efficient recombination (over 50% efficiency) was achieved only when sufficient Int (over 400 nM) and Gp3 (more than half of the Int concentration) were supplied (Fig. 3.4).

3.2.2 Effects of reaction times on recombination efficiency

The previous results presented the recombination efficiencies at the 3 hour time point with different protein concentrations, but were not informative about the kinetics of the recombination. To explore the reaction at early time points, reaction rates, and the time taken to reach steady state, time course reactions were performed in the following experiments.

The *in vitro* time course experiments were carried out in a total reaction volume of 240 µl (as described in section 2.26). Conditions (for *attB*×*attP* reaction: 800 nM Int, for *attR*×*attL* reaction: 800 nM Int and 800 nM Gp3) were chosen based on the results in section 3.3.1 which yielded the highest recombination efficiencies. Reactions were stopped at various time points by heating, digested with EcoRI and KpnI, and separated by agarose gel electrophoresis (Fig. 3.5)

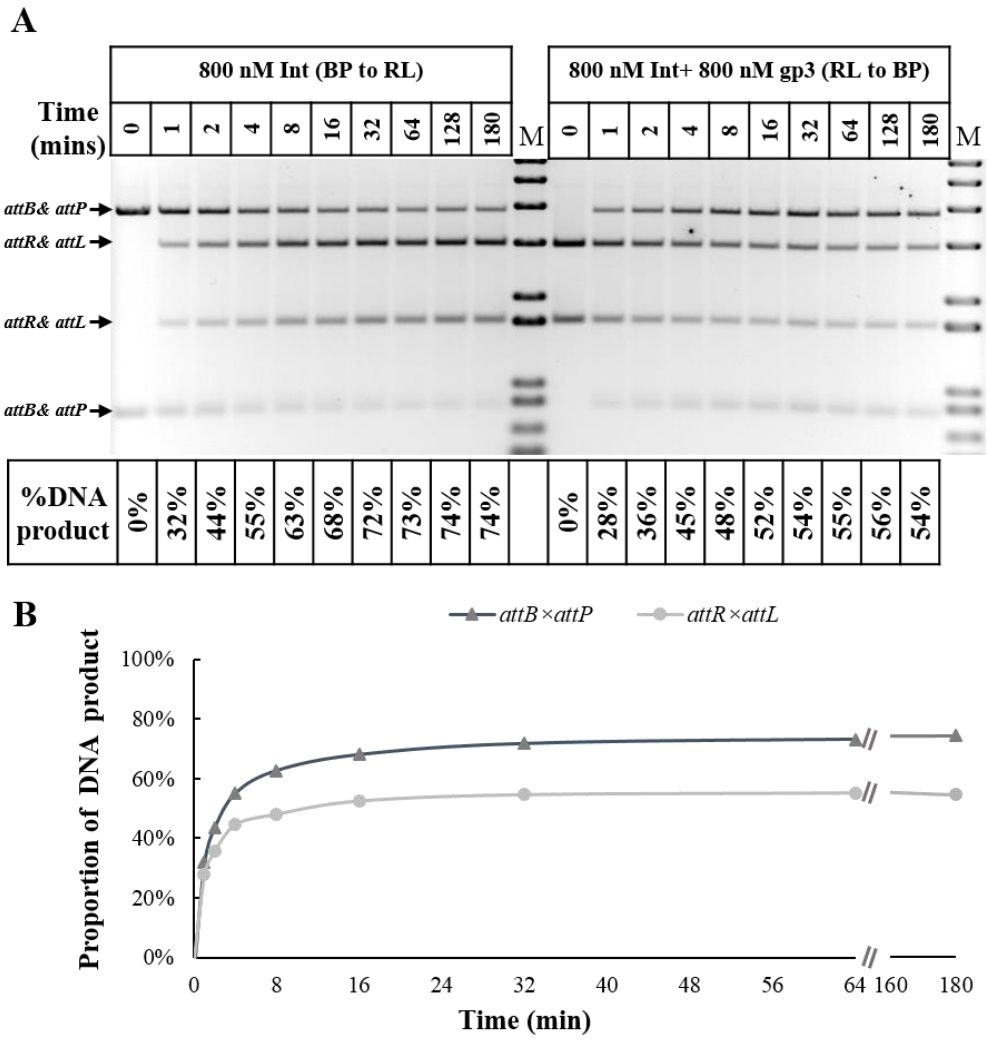


Figure 3. 5 Kinetics of the in vitro inversion reaction. In the *attB*×*attP* reaction time course, the initial substrate plasmid pZJ56off (*attB*, *attP*, ~10 nM) was recombined by Int (800 nM). In the *attR*×*attL* reaction time course, the initial substrate plasmid pZJ56on (*attR*, *attL*, ~10 nM) was recombined by Int (800 nM) and Gp3 (800 nM). For each time point, 20 µl of sample was heated at 80 ℃ for 10 min to stop the reaction. (A) Gel electrophoresis analysis of *in vitro* time course reaction. The plasmid containing *attB* and *attP* was separated from the plasmid containing *attR* and *attL* by agarose gel electrophoresis after restriction with KpnI and EcoRI (left half gel: *attB*×*attP* reaction; right half gel: *attR*×*attL* reaction). Band intensities were quantified using Quantity One, and the percentages of product in total DNA were calculated (as section 2.33), and shown in table below the gel picture. (B) Graph showing the percentages of recombinant products after different reaction times (dark grey line with triangle marker represents *attB*×*attP* time course reaction and light grey line with dot marker represents *attR*×*attL* time course reaction).

Recombination products were visible at 1 min and increased to more than half of the final reaction yield at 2 min for both *attB*×*attP* and *attR*×*attL* reactions (Fig. 3.5). After 30 min, about 70% of substrate had been recombined from *attB* and *attP* state (BP state) to *attR* and *attL* state (RL state), and about 50% of substrate had been recombined from RL state to BP state. The time course shows that both *attB*×*attP* and *attR*×*attL* inversion recombination *in*

vitro are very fast and they reach the same extent as reactions after 3 hour shown in section 3.2.1.

The work presented in section 3.2 provides information on the inversion recombination mediated by Int *in vitro*. Different concentrations of Int and Gp3 were applied into both *attB*×*attP* reaction and *attR*×*attL* reaction, and the kinetics of both reactions were tested under optimised condition. These *in vitro* data can also be used in mathematical modelling, which will be helpful for deeper understanding of the biological behaviours.

3.3 Investigations of the inversion recombination mediated by ϕC31 integrase *in vivo*

To explore if the inversion reaction is functional in living cells, recombination assays were carried out in *E. coli* as described in the following sections.

3.3.1 Construction and characteristics of a tightly regulated ϕC31 integrase expression system

To carry out recombination reactions *in vivo*, a tightly controlled Int expression system is crucial, and will be required for the operation of any genetic device according to the aims of study in chapter 1. It has been reported that the P_{BAD} promoter can be rapidly and efficiently repressed, quickly induced and can lead to high-level expression (Guzman *et al*., 1995). Therefore, the ϕC31 integrase gene was amplified by PCR and placed under the control of the P_{BAD} promoter in the plasmid pBAD33 (Guzman *et al*., 1995), resulting in the Int expression plasmid pZJ7 (medium copy-number plasmid with p15a origin of replication and chloramphenicol resistance gene, section 2.7). Transcription from the P_{BAD} promoter should be induced in the presence of arabinose, and repressed even further in the presence of glucose.

The properties of the Int expression plasmid were tested in an *in vivo* recombination assay. The substrate plasmid pGD01 (section 2.7), containing a *galK* gene flanked by directly repeated *attB* and *attP*, was used in this assay. Loss of the *galK* reporter gene was assayed using MacConkey galactose plates. By utilizing the Galactose available in the medium, $galK^+$ strain will produce acid, which lowers the pH of the agar below 6.8 and results in the appearance of pink colonies. While $galK^-$ strain will metabolize the amino acids, which raises the pH of the agar, and leads to the formation of white/colorless colonies on the plate. Detailed process is illustrated in Figure 3.6.

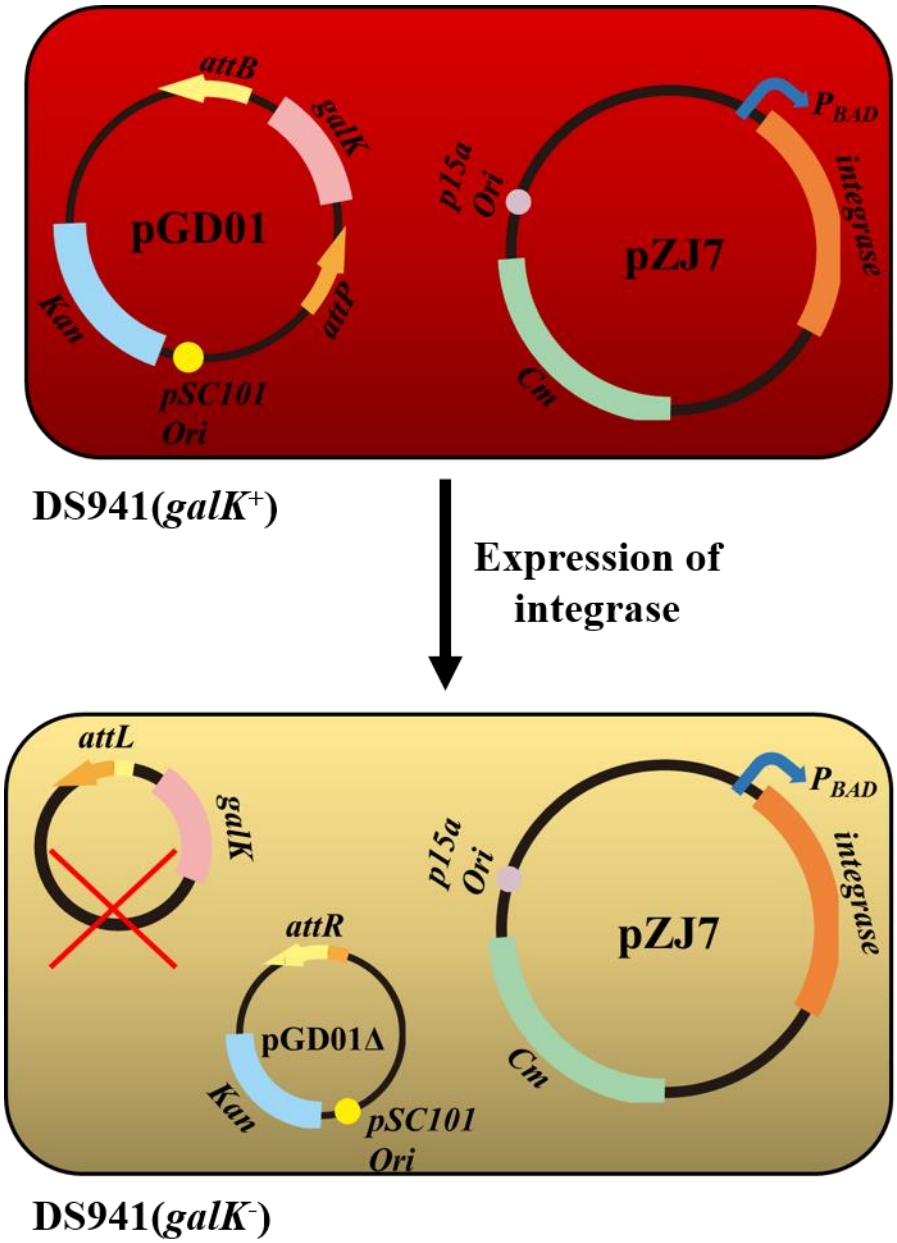


Figure 3. 6 The MacConkey recombination assay. *E. coli* DS941 was transformed with the substrate plasmid pGD01 (containing the *galK* gene flanked by directly repeat recombination sites *attB* and *attP*), and the Int expression plasmid pZJ7 (containing *Int* under the control of arabinose inducible P_{BAD} promoter). MacConkey agar contains a pH indicator 2-methyl-3-amino-6-dimethylaminoaphenazine, which is red at pH<6.8 and yellow at pH>8.0. Galactose was added as carbon source. DS941 strain is a *galK* mutant and therefore can only metabolize galactose when complemented by the *galK* gene from plasmid pGD01. If the expression of Int is controlled tightly, Int will not be expressed in the absence of arabinose. Cells can maintain the *galK* gene on the substrate plasmid, allowing them to metabolize galactose, produce acid, and lower the pH below 6.8, which results in red colonies. When the expression of Int is activated by arabinose, the pGD01 can be recombined into two DNA circles: one circle containing the resistance gene and the origin of replication (named pGD01Δ), and one small *galK*-containing circle, lacking a replication origin, which cannot replicate and will be lost after cell division. Cells losing the *galK* gene will be forced to metabolize the amino acids in the agar causing an increase in pH, resulting in white colonies. The drawing is not to scale.

To test the Int expression system, an experiment was performed by culturing *E. coli* strain DS491/pGD01+pZJ7 in L-Broth with glucose or arabinose overnight. Plasmid DNA was isolated from cells and checked by agarose gel electrophoresis (Fig. 3.7A). The overnight cultures were diluted and plated on MacConkey agar plates to check for loss of the *galK* gene (Fig. 3.7B).

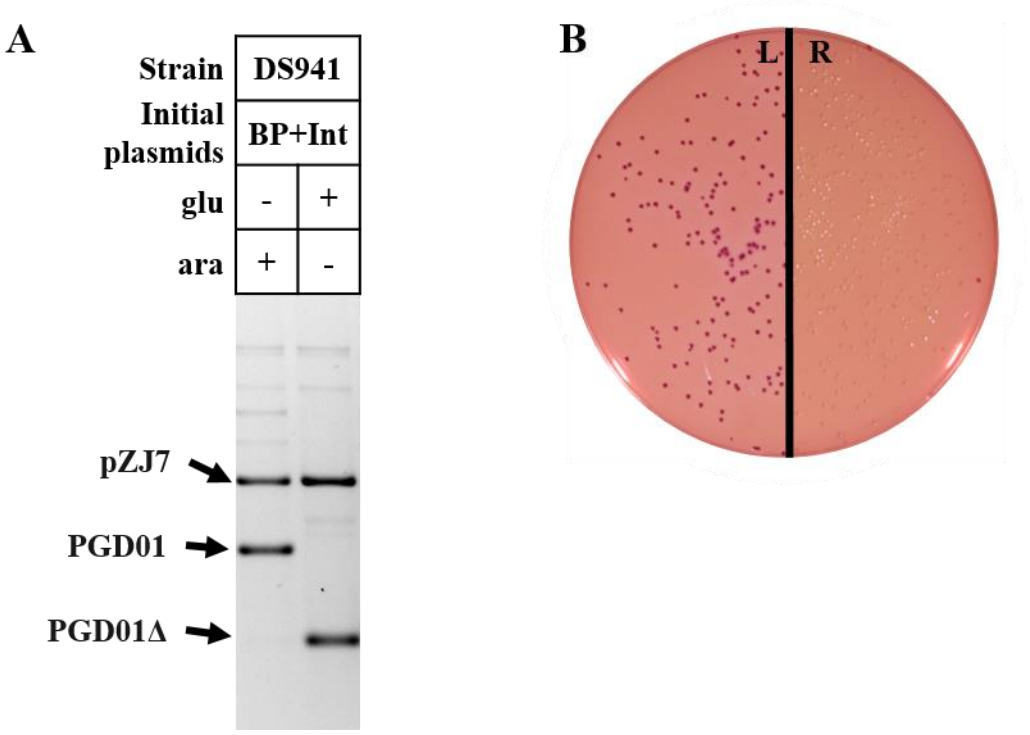


Figure 3. 7 MacConkey recombination assay used to test the ΦC31 integrase expression system. (A) Gel electrophoresis analysis of *in vivo* recombination. DS941 cells containing substrate plasmid pGD01 (BP) and Int expression plasmid pZJ7 (Int) were cultured in L-broth with glucose (glu) or arabinose (ara) overnight. Plasmids were extracted from cells and run uncut on an agarose gel. (B) Cells were centrifuged, washed and diluted before being spread on MacConkey agar plates to get single colonies. After overnight incubation, plates were photographed to compare the colour of colonies (the labels on the plate: L, cells that had been cultured with glucose; R, cells that had been cultured with arabinose).

The plasmid substrate did not recombined in cells cultured in medium containing glucose, whereas the plasmid substrate was recombined completely in cells cultured in medium containing arabinose, producing plasmid pGD01Δ which is smaller than the substrate plasmid due to the deletion of *galK* gene (Fig.3.7A). These results were also verified by the colour of colonies on MacConkey agar plate: the red colonies demonstrated that cells from glucose medium maintained the *galK* gene, however, the white colonies demonstrated that cells from arabinose medium lost *galK* gene due to the complete recombination (Fig. 3.7B)

The changes of both the DNA state in cells and the color of colonies on MacConkey plate proved that the expression of Int can be regulated tightly. In the presence of glucose, the transcription from P_{BAD} promoter is repressed tightly, no integrase is produced and no recombination products are observed. In the presence of arabinose, the transcription is turned on, integrase is expressed and efficient recombination takes place.

3.3.2 Inversion recombination reaction with high copy-number plasmid substrates *in vivo*

In view of the tight control of Int expression, the previously constructed expression system was further used to investigate the inversion reaction.

3.3.2.1 Investigations on the *attB×attP* inversion reaction with the substrate plasmid pZJ16off

To test the *attB×attP* inversion reaction *in vivo,* the substrate plasmid pZJ16off (section 2.7) containing *attP* and *attB* in inverted repeat was constructed. This plasmid has the following characteristics: high copy-number plasmid with pMB1 origin, ampicillin resistance gene, *rfp* reporter gene flanked by recombination sites in the opposite orientation to the P_{lac} promoter outside of the sites, and convenient restriction site (SacI) for easy recombination results analysis. It was predicted that the pZJ16off can be recombined by Int alone, resulting in pZJ16on. The substrate plasmid pZJ16off and the product plasmid pZJ16on were expected to give different restriction digest patterns (Fig. 3.8A) and RFP expression due to the opposite orientation of the sequence flanked by recombination sites. To test the inversion recombination with this high-copy substrate plasmid, the experiment was performed by culturing strain DS941/pZJ16off+pZJ7 in L-Broth with glucose or arabinose overnight and plasmid DNA was isolated from cells and checked by agarose gel electrophoresis (Fig. 3.8B). The cell pellets from overnight cultures were photographed (Fig. 3.8C).

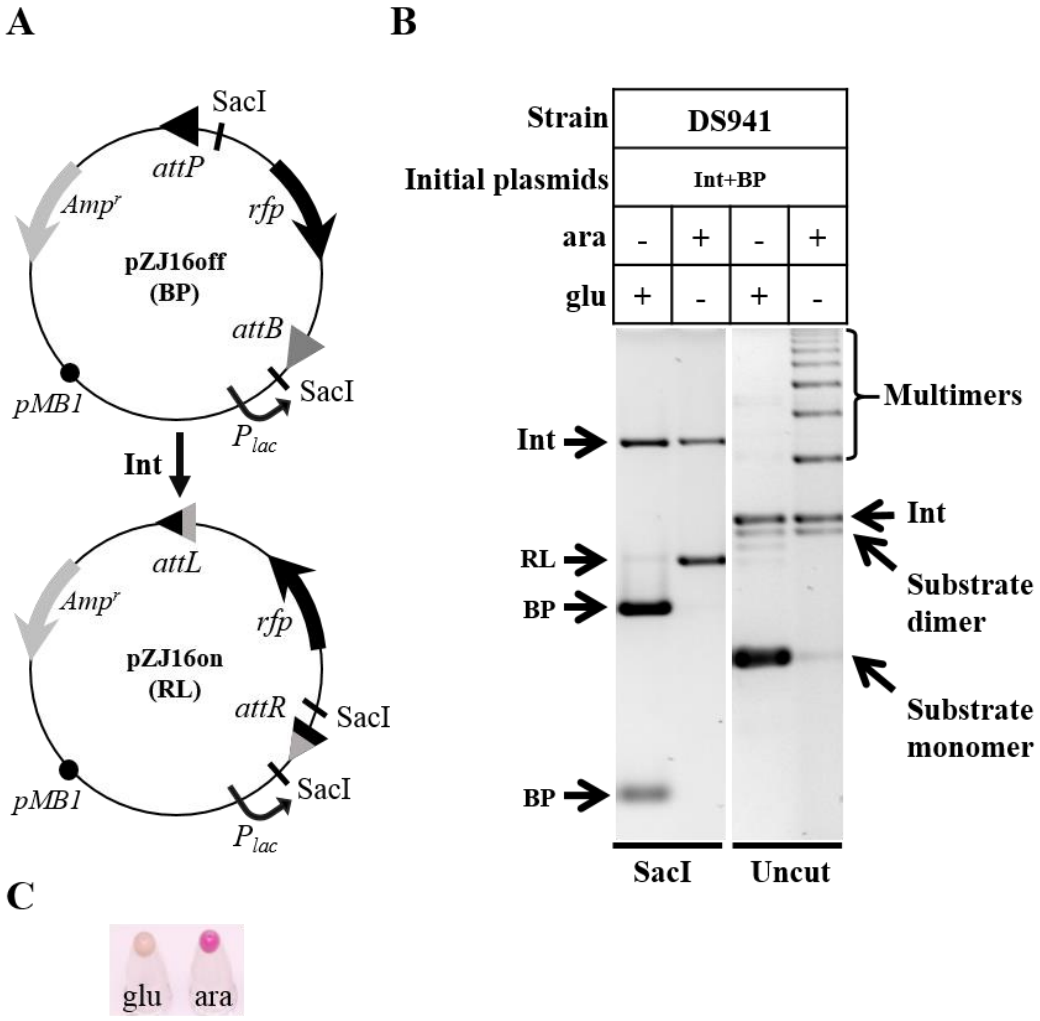


Figure 3. 8 Inversion reaction between attP and attB. (A) Expected change of substrate plasmids with *attP*×*attB* inversion recombination. Int alone is expected to recombine the substrate plasmid pZJ16off (BP) to pZJ16on (RL) and place the *rfp* gene to the same orientation as P_{lac} promoter. Plasmid pZJ16off gives two fragments (size: 813 bp and 2523 bp) and pZJ16on gives two fragments (size: 59 bp and 3277 bp) after restriction digestion with SacI (the drawing is not to scale). (B) Gel electrophoresis analysis of the *attB*×*attP* reaction. DS941/ pZJ16off +pZJ7 was cultured in L-Broth with glucose (glu) or arabinose (ara) overnight. Plasmid DNA was extracted from cells and run cut and uncut DNA on an agarose gel (bands smaller than 600 bp are not shown on the gel picture). (C) Image showing the colour of cells pellets. The overnight culture of DS941/ pZJ16off +pZJ7 in medium with glucose (glu) or arabinose (ara) were centrifuged in separate microcentrifuge tubes and photographed using a digital camera.

The inversion recombination mediated by Int was as efficient as the deletion recombination (section 3.3.1). Nearly all of the plasmid substrate was recombined when expression of Int was turned on by culturing cells in medium with arabinose, and almost all of the plasmid substrate was maintained unrecombined when the expression of Int was repressed by addition of glucose (Fig. 5.8B). This again demonstrated that the expression of Int can be tightly regulated and the concentration of Int was sufficient for recombining the high copy-number plasmid substrates completely in *E. coli* cells.

Cells cultured in medium with glucose gave a white coloured pellet and cells cultured in medium gave a red coloured pellet, indicating different RFP expression levels (Fig. 3.8C). The *rfp* gene flanked by inverted recombination sites was flipped to change the RFP expression. Thus, it is possible to visualize the recombination result through observation of the reporter gene expression.

However, the uncut plasmid DNA on the gel picture revealed that almost all of the monomeric plasmid substrate was gone and had been converted to multimers (Fig. 3.8B). The likely explanation for the production of multimers was that Int mediated both intramolecular and intermolecular recombination between *attB* and *attP* during the recombination.

As described in Figure 3.9 for a recombination reaction starting with pZJ16off, after restriction digestion with SacI, non-recombined plasmid containing *attB* and *attP* generates two fragments (size: 813 bp and 2523 bp), the intramolecular recombination product containing *attR* and *attL* sites gives two fragments (size: 59 bp and 3277 bp), and intermolecular recombination product contains all four sites and generates four fragments (size: 813 bp, 2523 bp, 59 bp, and 3277 bp). Similar patterns of bands are produced if further intermolecular recombination takes place producing higher and higher multimers. However, if recombination is complete, converting all *attB* and *attP* sites to *attR* and *attL* sites, the recombinant restriction pattern (59 bp and 3277 bp) will be obtained regardless of the plasmid conformation (monomeric or multimeric). The digest products on the gel picture showed that almost all of the substrate containing *attB* and *attP* was recombined to plasmid containing *attR* and *attL* by Int (Fig. 3.8B, band sizes are 59 bp and 3277 bp), though lots of multimers were formed during recombination. In consideration of the fact that multimerisation of substrate plasmids in cells may lead to a decrease in the plasmids heritable stability (Summers and Sherratt, 1984), the following work aimed to address this problem.

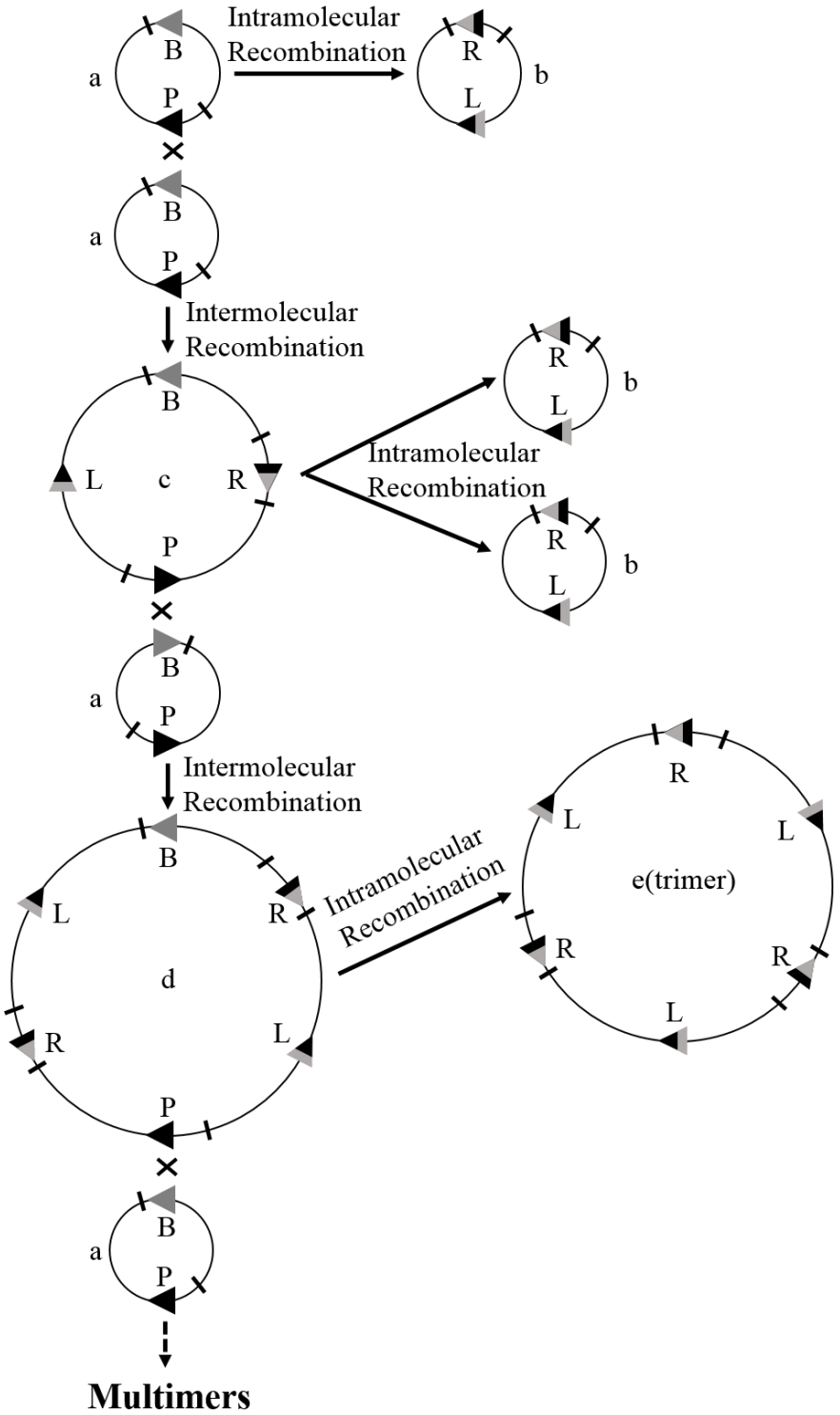


Figure 3. 9 Intramolecular and intermolecular recombination. The expected monomer product (b) can be achieved by intramolecular recombination between inverted repeat sites *attB* and *attP* of (a). Recombination between *attB* on one plasmid copy and *attP* on another plasmid copy will produce the dimer product (c). This dimer molecule (c) can be further recombined to form monomer product (b) or trimer product (d). The *attB* and *attP* on this trimer molecule (d) can be recombined by inversion recombination to form trimer (e) which cannot be recombined any more by Int alone. In this way, the trimer plasmid will be accumulated in cells. Using the same mechanism other multimers can be produced and accumulated in cells. The short lines on the plasmid circle represent the digest position of SacI (size of bands after restriction digestion with SacI, a: 813 bp and 2523 bp, b and e: 59 bp and 3277 bp, c and d: 813 bp, 2523 bp, 59 bp, and 3277 bp). The drawing is not to scale.

3.3.2.2 Investigations of *attB*×*attP* inversion reactions with substrate plasmid pZJ18off containing a *cer* recombination site

To solve the plasmid multimeriztion problem, a *cer* recombination site was introduced into the substrate plasmid. It has been reported that the natural plasmid ColE1 is maintained stably in *E.coli*. The *cer* site, which is recombined by the chromosomally encoded Xer recombination system to keep ColE1 monomeric, contributes to this stability (Summers and Sherratt, 1984). The recombinases XerC and XerD, and accessory proteins ArgR and PepA, are required for *cer* site-specific recombination (Alén *et al*., 1997; Blakely *et al*., 1993; Colloms *et al*., 1990; Stirling *et al*., 1988; Stirling *et al*., 1989). Recombination only happens when *cer* sites are in direct repeat on the same circular DNA molecule, which ensures that recombination resolves plasmid multimers to monomers, but does not create multimers from monomers (Colloms *et al*., 1996). Based on this theory, it was hypothesized that cloning a *cer* recombination site into the integrase inversion substrate plasmid could mitigate the multimerisation problem. Intermolecular recombination will produce multimers with at least two copies of *cer* on a single DNA molecule, and when two *cer* sites are in direct repeat they will be recombined by host-encoded Xer proteins to convert the multimers to monomers (Fig. 3.10).

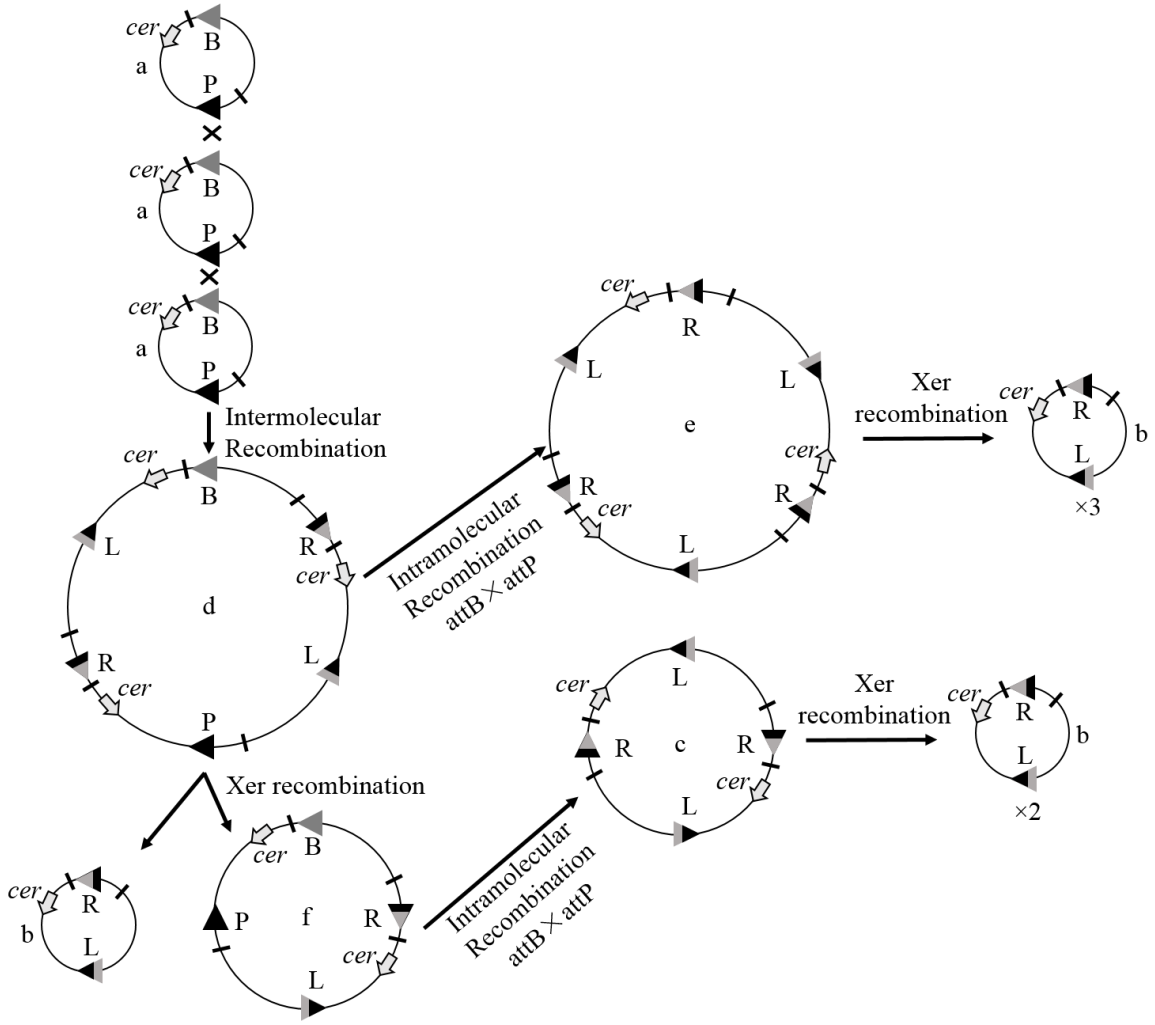


Figure 3. 10 Combination of integrase site-specific recombination and cer recombination to convert multimers to monomers. Multimers (such as c and e) in which all of the *attB* and *attP* have been recombined by Int will have all of the *cer* sites in direct repeat. These *cer* sites can be recombined to convert the multimers to monomers. In multimers (such as f) which containing unrecombined *attB* and *attP*, the *cer* sites which are in inverted repeat cannot be recombined to convert the multimers to monomers.

If recombination occurs as predicted both in this section and last section (3.3.2.1), the combination of complete integrase recombination and Xer recombination would finally convert the multimers to monomers.

To investigate if the Xer recombination can mitigate the multimerisation problem, the new substrate plasmid pZJ18off (section 2.7) containing *attB* and *attP* in inverted repeat was constructed, which has the following characteristics: high copy-number plasmid with pMB1 origin and *cer* recombination site, ampicillin resistance gene, *rfp* reporter gene flanked by recombination sites and in an opposite orientation to the P_{lac} promoter, and convenient restriction sites (AlwNI and NheI) for easy recombination results analysis. Strain DS941 (*PepA*$^{+}$), which can express all of the necessary proteins for Xer recombination, was used as

the test strain. Strain DS957 (*PepA*$^-$) which is a *pepA* mutant was used as the control strain that cannot catalyse recombination at *cer*. Recombination reactions were performed by culturing strain DS941/pZJ16off+pZJ7, DS941/pZJ18off+pZJ7, DS957/pZJ16off+pZJ7, and DS957/pZJ18off+pZJ7 in L-Broth with glucose or arabinose overnight and plasmid DNA was isolated from cells. Multimerisation was observed by gel electrophoresis of uncut plasmid DNA (Fig. 3.11A), while recombination efficiencies on substrate with and without *cer* were compared by checking the restriction digestion of DNA from strain DS941 (*PepA*$^+$) (Fig. 3.11B).

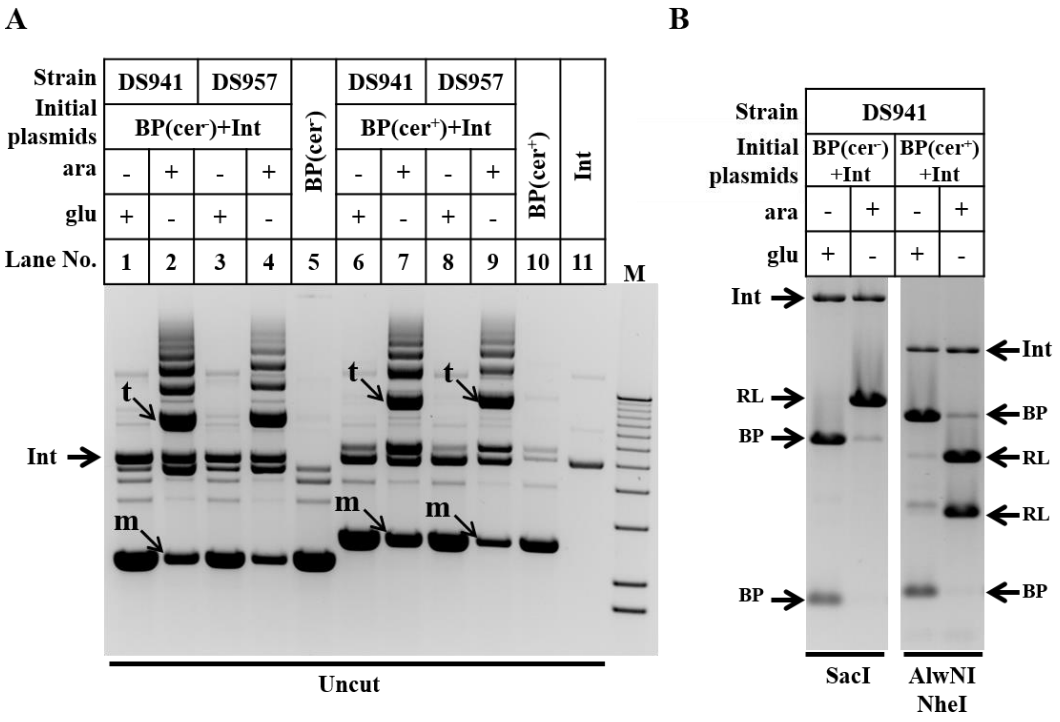


Figure 3. 11 Comparison of the recombination between substrates without and with cer recombination site. Strain DS941/ pZJ16off (BP, cer$^-$) + pZJ7 (Int), DS957/ pZJ16off (BP, cer$^-$) + pZJ7 (Int), DS941/ pZJ18off (BP, cer$^+$) + pZJ7 (Int) and DS957/ pZJ18off (BP, cer$^+$) + pZJ7 (Int) were cultured in medium with glucose (glu) or arabinose (ara) overnight. (A) Plasmid DNA was extracted from cells and run on an agarose gel to compare the uncut plasmid states (label 'm' represents monomers, label 't' represents trimers). (B) DNA extracted from cells was digested with SacI (pZJ16off, BP: 813 bp and 2523 bp, pZJ16on, RL: 3277 bp and 59 bp, pZJ7: 7175 bp) or with NheI and AlwNI (pZJ18off, BP: 896 bp and 3057 bp, pZJ18on, RL: 1610 bp and 2343 bp, pZJ7, 4844 bp, 1682 bp, 603 bp and 46 bp) and run on an agarose gel to check the recombination results in strain DS941 (bands smaller than 600 bp are not shown on the gel picture).

After overnight culture, most of the plasmids were maintained in monomeric form in cells which were cultured in medium with glucose (Fig. 3.11A, lanes 1, 3, 6 and 8). In contrast, extensive multimerisation was observed in DNA samples from cells which were cultured in medium with arabinose (Fig. 3.11A, lanes 2, 4, 7 and 9). These results revealed that the plasmid states in cells were changed in the presence of Int.

Comparing the results of pZJ16off (*cer*$^-$) with pZJ18off (*cer*$^+$) in the *PepA*$^+$ strain (Fig. 3.11A, lane 2 and 7), pZJ16off (*cer*$^-$) gave much higher ratio of trimer to monomer (lane 2) than pZJ18off (lane 7). Comparing the results of substrate plasmid (*cer*$^+$) in the *PepA*$^+$ with that in *PepA*$^-$ strains (Fig. 3.11A, lane 7 and 9), there were similar amounts of trimers in both DNA samples, but more monomer in the *PepA*$^+$ sample (lane 7). These comparisons indicated that more substrate was retained in a monomeric state if the substrate plasmid had a *cer* multimer resolution site and was propagated in cells containing a complete Xer site-specific recombination system.

The restriction digest samples showed that Int mediated *attB*×*attP* recombination on the substrate plasmid with *cer* was as efficient as recombination on the substrate plasmid without *cer* (Fig. 3.11B). Although the plasmid multimerisation did not appear to have an adverse effect on the integrase reaction, to ensure plasmid stable inheritance, substrate plasmid pZJ18off with *cer* was used for the following studies.

3.3.2.3 Investigations on the *attR* × *attL* inversion reaction with substrate plasmid pZJ18on containing a *cer* recombination site

The next thing was to test whether Int can mediate *attR* × *attL* inversion reaction in the presence of Gp3 *in vivo*. To do this, a Gp3 expression plasmid pZJ14 (section 2.7) was constructed. This plasmid has the following characteristics: low copy-number plasmid with pSC101 origin which was different from the origin of Int expression plasmid (p15a) or substrate plasmid (pMB1), kanamycin resistance gene, *gp3* gene is under the control of $P_{Ltet0-1}$ promoter. Promoter $P_{Ltet0-1}$ can be tightly repressed by the *Tet* repressor, and efficiently induced in the presence of aTc (Lutz and Bujard, 1997). Therefore, depending on whether the host strain can express the *Tet* repressor, this promoter can be used either as a constitutive or an inducible promoter.

Substrate plasmid pZJ18on (section 2.7, Fig. 3.12A) containing inverted repeat *attR* and *attL* was selected from the recombination product of pZJ18off (BP) *in vivo* (as described in section 2.31). Recombination was first tested in DS941 which does not express Tet repressor. In this strain Gp3 should be expressed constitutively from $P_{Ltet0-1}$ promoter. Strain DS941/pZJ18on+pZJ7+pZJ14 was cultured in medium with glucose or arabinose overnight to test the *attR*×*attL* reaction (Fig. 3.12B).

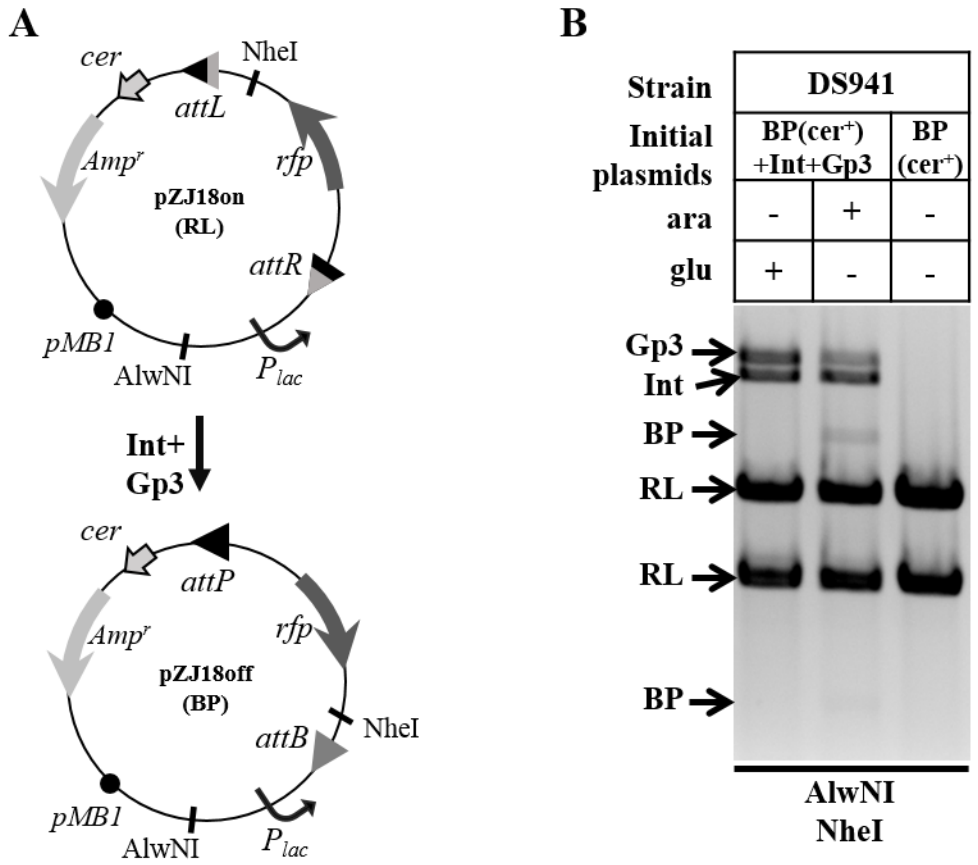


Figure 3. 12 Inversion reaction between attR and attL. (A) Expected change of substrate plasmids with *attR*×*attL* inversion recombination. Int and gp3 are expected to recombine the substrate plasmid pZJ18on (RL) to pZJ18off (BP) and place the *rfp* gene to the opposite direction as P_{lac} promoter. Plasmid pZJ18off gives two fragments (size: 896 bp and 3057 bp) and pZJ18on gives two fragments (size: 1610 bp and 2343 bp) after restriction digestion with AlwNI and NheI (the drawing is not to scale). (B) Gel electrophoresis analysis of the *attR*×*attL* reaction. Strain DS941/pZJ18on (BP, cer^+) + pZJ7 (Int) + pZJ14 (Gp3) was cultured in L-broth with glucose (glu) or arabinose (ara) overnight. DNA extracted from cells was digested with AlwNI and NheI (pZJ7: 4844 bp, 1682 bp, 603 bp, and 46 bp, pZJ14: 4278 bp, bands smaller than 600 bp are not shown on the gel picture).

There was no visible recombination product seen on the gel from cells cultured in medium with glucose (Fig. 3.12B). When the expression of Int was activated by culturing cells in medium with arabinose, only a small amount of substrate pZJ18on was recombined to pZJ18off (Fig. 3.12B).

The efficient *attB*×*attP* reaction observed with pZJ7 (Fig. 3.11B) suggests that there is enough Int expressed for recombining all of the *attB* and *attP* sites. The inefficient *attR*×*attL* (Fig. 3.12B) reaction suggests that there may not have been enough Gp3 for the amount of Int. Therefore, it was thought that decreasing the expression level of Int or increasing the expression level of Gp3 would improve the *attR*×*attL* reaction efficiency.

3.3.2.4 Effects of Int expression levels on inversion reaction efficiency

The results obtained previously (section 3.3.2) indicated two possible ways to improve the *attR*×*attL* reaction efficiency, which are increasing Gp3 and decreasing Int expression levels. To avoid extra metabolic load for cells, attempting to decrease the Int expression level rather than increasing the Gp3 expression level was selected. This was done by changing the ATG start codon of pZJ7 to GTG, which was suggested to be a weaker start codon (Bonnet *et al*., 2012; Reddy *et al*., 1985), or by changing the RBS for Int.

First, a new Int expression plasmid pZJ7m was made by site-directed mutagenesis to change the start codon of Int from ATG to GTG (as described in section 2.24). Next, a sequence library was designed to introduce random mutations into the original RBS sequence (GAGCTCAGGAGTAACATATG) of the int gene, simultaneously adding a new restriction site into the sequence (GAGCTCCAGCTCGAGARRARRAACATATG, R= G or A, XhoI site is underlined) for easy checking of the sequence change. There are 16 possible sequences in this library, and their translation initiation rates were calculated by an online software of Salis *et al*. (Salis *et al*., 2009), giving a range of predicted Translation Initiation Rates from 53 to 3379. Three RBS sequences with lower predicted initiation rate were selected, synthesised as oligonucleotides, and inserted into pZJ7 between restriction sites SacI and NdeI upstream of the int gene, resulting in three new Int expression plasmids (pZJ20, pZJ21, and pZJ22, section 2.7). This construction strategy is shown in Figure 3.13 and details of all these different Int expression plasmids are presented in Table 3.1.

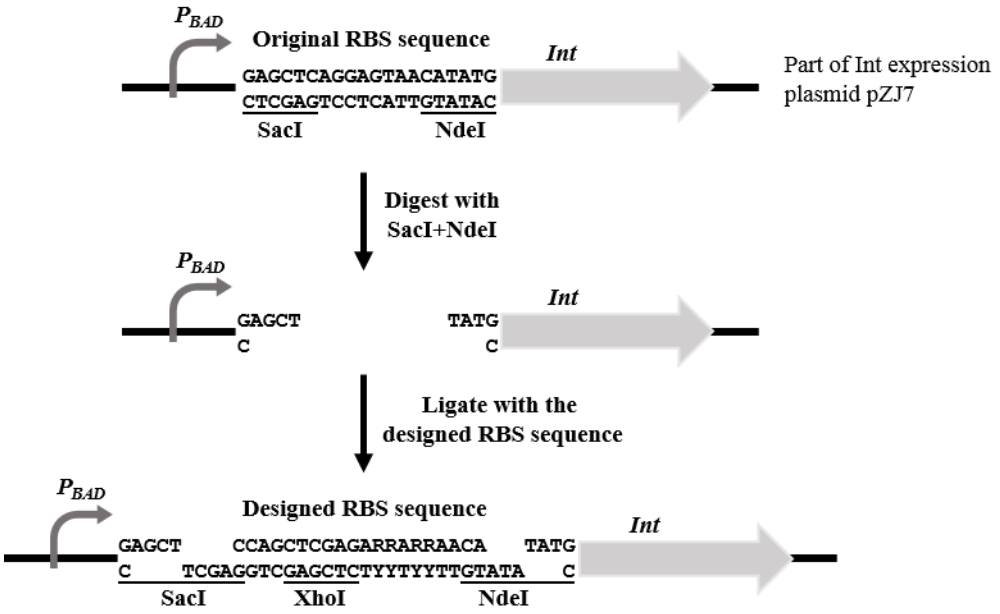


Figure 3. 13 Process of changing the RBS sequence for Int. The original Int expression plasmid pZJ7 was digested with SacI and NdeI. Designed RBS sequences were synthesised as oligonucleotides. Two oligonucleotides were annealed to form a double strand liner DNA and inserted into the plasmid pZJ7 to replace the original RBS sequence.

Table 3. 1 Int expression plasmids

Name	RBS sequence and Start codon	Predicted translation initiation rate (au)
pZJ7	GAGCTCAGGAGTAACAT**ATG**	227
pZJ7m	GAGCTCAGGAGTAACAT**GTG**	144
pZJ20	GAGCTCCAGCTCGAGATGATTAACAT**ATG**	126
pZJ21	GAGCTCCAGCTCGAGATTATTAACAT**ATG**	106
pZJ22	GAGCTCCAGCTCGAGCTTCTTAACAT**ATG**	56

Restriction sites are underlined in all RBSs, and the start codon in each sequence is shown in bold. SacI: GAGCTC, NdeI: CATATG, and XhoI: CTCGAG.

To investigate the effects of Int expression levels on recombination efficiency, the Int expression plasmids (Table 3.1) were used to mediate both *attB*×*attP* and *attR*×*attL* reactions in strain DS941. Strain DS941 containing substrate plasmid pZJ18off and one of the Int expression plasmids (pZJ7, pZJ20, pZJ21, and pZJ22) was cultured in medium with glucose or arabinose overnight to test the *attB*×*attP* reaction (Fig. 3.14A). Strain DS941 containing substrate plasmid pZJ18on, Gp3 expression plasmid pZJ14 and one of the Int expression

plasmids (pZJ7, pZJ20, pZJ21, and pZJ22) was cultured in medium with glucose or arabinose overnight to test the *attR*×*attL* reaction (Fig. 3.14B).

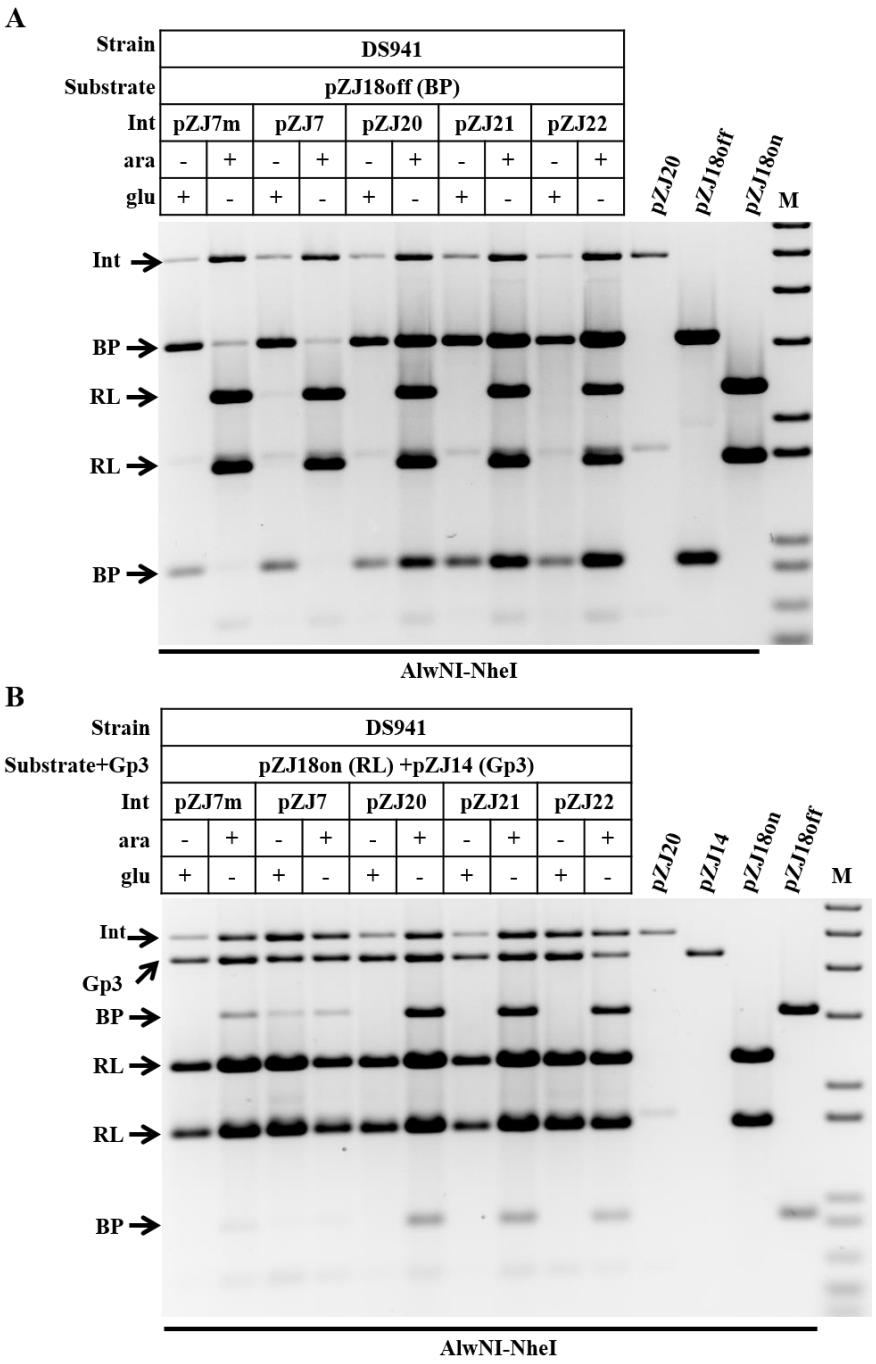


Figure 3. 14 Effects of different Int expression levels on attB×attP and attR×attL recombination on high copy number substrate plasmids. (A) Gel electrophoresis analysis of the *attB*×*attP* reaction. DS941 strain was transformed with substrate plasmid pZJ18off (BP) and one of the Int expression plasmids (pZJ7m, pZJ7, pZJ20, pZJ21 and pZJ22), and cultured in L-broth with glucose (glu) or arabinose (ara) overnight. DNA extracted from cells was digested with AlwNI and NheI (pZJ18off, BP: 896 bp and 3057 bp, pZJ18on, RL: 1610 bp and 2343 bp, pZJ7m, pZJ7, pZJ20, pZJ21 and pZJ22: 4844 bp, 1682 bp, 603 bp, and 46 bp) to check the *attB*×*attP* reactions (bands smaller than 600bp are not shown on the gel picture). (B) Gel electrophoresis analysis of the *attR*×*attL* reaction. DS941 strain was transformed with substrate plasmid pZJ18on (RL), one of Int expression plasmids (pZJ7m, pZJ7, pZJ20, pZJ21, and pZJ22) and the Gp3 expression plasmid pZJ14, and cultured in L-broth with glucose or arabinose overnight. DNA extracted from cells was cut with AlwNI and NheI (pZJ14:4278 bp) to check the *attR*×*attL* reactions (bands smaller than 600 bp are not shown on the gel picture).

Reactions had varying recombination efficiencies with different Int expression plasmids (Fig. 14). For both *attB*×*attP* and *attR*×*attL* reactions, Int expressed from pZJ7m (GTG start codon) gave the same recombination efficiency as Int expressed from pZJ7. Int expressed from pZJ20, pZJ21 and pZJ22 (weaker predicted RBSs) mediated lower efficient *attB*×*attP* reaction, but higher efficient *attR*×*attL* reaction than Int expressed from pZJ7 (stronger predicted RBS). These results suggest that the expression level of Int is influenced by its RBS sequence. Sufficient concentrations of Int can mediate nearly 100% *attB*×*attP* recombination *in vivo* and the efficiency can be decreased by lowering the Int expression level. Increasing the ratio of Gp3 to Int by decreasing the expression level of Int improved the *attR*×*attL* reaction efficiency from less than 5% to 30%, but still need to be further improved.

3.3.3 Inversion recombination reaction with low copy-number plasmid substrates *in vivo*

For easy observation of the recombination result, all of the experiments so far were carried out on high copy-number substrate plasmids. Therefore, lowering the Int expression level may lead to insufficient protein to completely recombine the recombination sites presented in a high concentration. Based on this hypothesis, the copy number of the substrate plasmid was decreased so that the relative protein to substrate (recombination sites) concentration was decreased. At the same time, the *gp3* gene was moved to a higher copy-number plasmid to increase the ratio of Gp3 to Int and be compatible with Int expression plasmid and substrate plasmid in the same cell.

3.3.3.1 Investigations on the inversion reaction mediated by Int with different expression levels

In order to address the issue of incomplete recombination on high copy-number plasmid substrates, the invertible sequence from pZJ18off was moved to a low copy-number vector to construct the substrate plasmid pZJ19off (section 2.7). This substrate plasmid has the following characteristics: low copy-number plasmid with pSC101 origin, kanamycin resistance gene, *rfp* reporter gene flanked by recombination sites and convenient restriction sites (XmnI and NheI) for easy recombination result analysis. Substrate plasmid pZJ19on (section 2.7) was made by recombining pZJ19off *in vivo* (as described in section 2.31). Meanwhile, the *gp3* gene with promoter $P_{Ltet0\text{-}1}$ was moved from pZJ14 (pSC101 origin) to pUC19 to construct the Gp3 expression plasmid pZJ24 (pMB1 origin, Amp^r, section 2.7), which was compatible with the substrate plasmid pZJ19off/on (pSC101 origin, Kan^r) and

the Int expression plasmid (p15a origin, Cm^r). Placing *gp3* on a higher copy-number plasmid had the advantage of increasing the ratio of Gp3 to Int, which was proved to be beneficial for *attR*×*attL* reaction in the *in vitro* assay (section 3.2.1.2). To investigate the effect of Int expression levels on recombination efficiency on low copy-number substrate plasmids, the family of integrase expression plasmids with different start codon and RBSs (table 3.1) were used for both *attB*×*attP* and *attR*×*attL* reactions in strain DS941. Strain DS941 with substrate plasmid pZJ19off and one of the Int expression plasmids (pZJ7, pZJ20, pZJ21, and pZJ22) was cultured in medium with glucose or arabinose overnight to test the *attB*×*attP* reaction (Fig. 3.15A). Strain DS941 containing substrate plasmid pZJ19on, Gp3 expression plasmid pZJ24 and one of the Int expression plasmid (pZJ7, pZJ20, pZJ21, and pZJ22) was cultured in medium with glucose or arabinose overnight to test the *attR*×*attL* reaction (Fig. 3.15B).

After decreasing the copy number of the substrate plasmids, the efficiency of *attB*×*attP* reaction was improved. All of the Int expression plasmids produced enough Int to catalyse efficient *attB*×*attP* recombination (greater than 95%) on pZJ19off, except the one containing the weakest predicted RBS (pZJ22) (Fig. 3.15). Reducing the substrate copy number also allowed efficient *attR*×*attL* reaction when the low levels of Int was expressed. Int from plasmids pZJ20, pZJ21, and pZJ22 showed about 70% *attR*×*attL* recombination efficiency, while Int from pZJ7 and pZJ7m showed less efficiency (50%).

These results suggest that 1) Reducing the copy number of recombination sites can increase the efficiency of *attB*×*attP* reaction, especially with Int expressed at low levels (pZJ21 on pZJ18off in Fig. 3.14 vs pZJ21 on pZJ19off in Fig. 3.15), 2) Increasing the ratio of Gp3 to Int by reducing Int expression levels in cells can improve the *attR*×*attL* reaction efficiency (pZJ21 and pZJ24 on pZJ19on vs pZJ7 and pZJ24 on pZJ19on in Fig. 3.15). Among all of the Int expression plasmids, Int expressed from pZJ21 generated the best *attB*×*attP* recombination result (nearly 100% efficiency) and *attR*×*attL* recombination result (nearly 70% efficiency). This might be because pZJ21 contains a predicted medium strength RBS for Int, which expresses enough Int for the *attB*×*attP* reaction and also provides an optimal ratio of Gp3 to Int for *attR*×*attL* recombination when Gp3 was expressed from pZJ24.

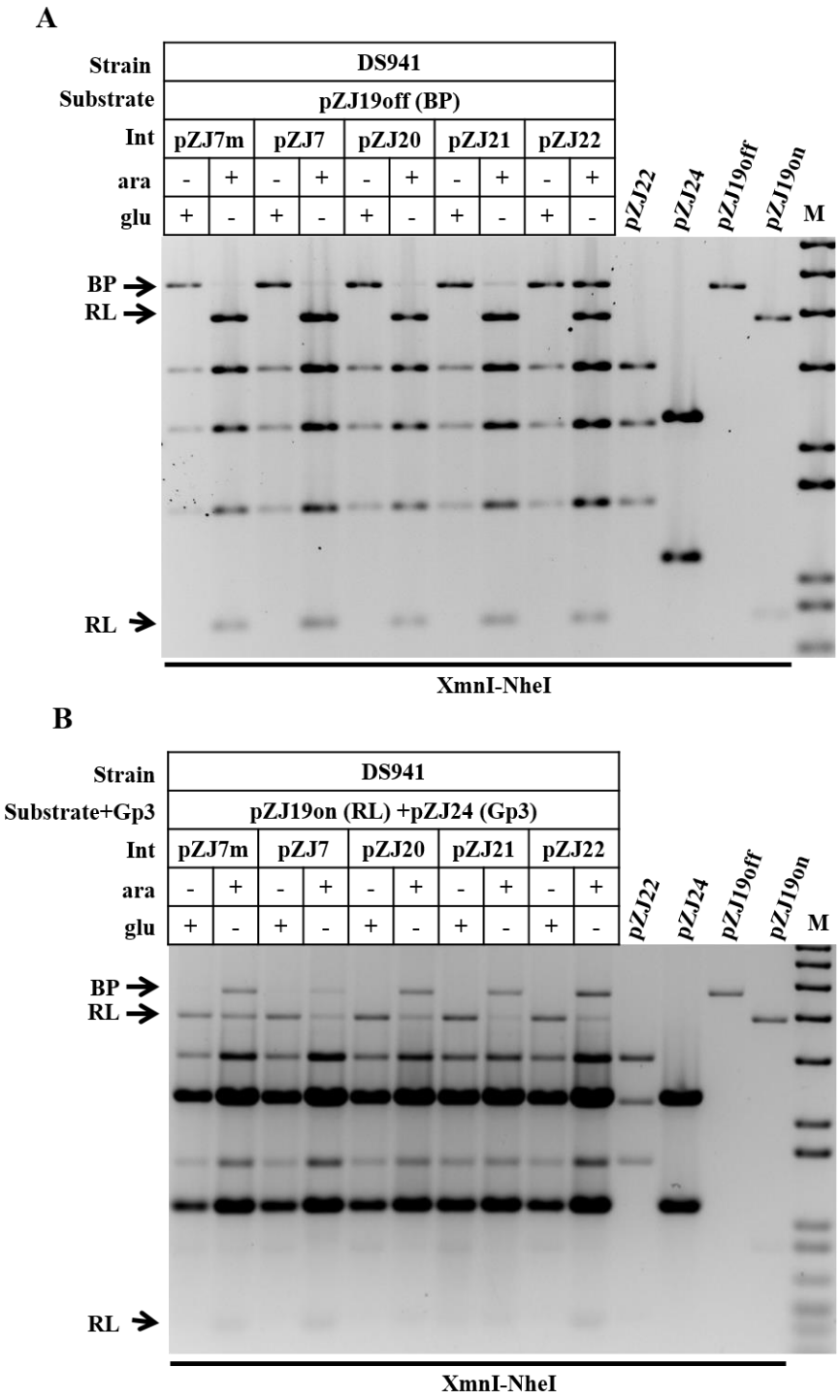


Figure 3. 15 Effects of different Int expression levels on attB×attP and attR×attL recombination on low copy number substrate plasmids. (A) Gel electrophoresis analysis of the *attB×attP* reaction. DS941 strain was transformed with substrate plasmid pZJ19off (BP) and one of Int expression plasmids (pZJ7m, pZJ7, pZJ20, pZJ21, and pZJ22), and cultured in medium with glucose (glu) or arabinose (ara) overnight. DNA extracted from cells was digested with XmnI and NheI (pZJ19off, BP: 92 bp, and 4562 bp, pZJ19on, RL: 806 bp and 3848 bp, pZJ7m, pZJ7, pZJ20, pZJ21, pZJ22: 2980 bp, 2229 bp, 1499 bp, 411 bp, and 56 bp) to check the *attB×attP* reactions (bands smaller than 600 bp are not shown on the gel picture). (B) Gel electrophoresis analysis of the *attR×attL* reaction. DS941 strain was transformed with substrate plasmid pZJ19on (RL), one of Int expression plasmids (pZJ7m, pZJ7, pZJ20, pZJ21, and pZJ22) and Gp3 expression plasmid pZJ24 and cultured in medium with glucose or arabinose overnight. DNA extracted from cells was digested with XmnI and NheI (pZJ24: 1133 bp and 2296 bp) to check the *attR×attL* reactions (bands smaller than 600 bp are not shown on the gel picture).

3.3.3.2 Effects of reaction time on inversion reaction efficiency

Before building up genetic devices based on the ϕC31 integrase inversion recombination system, the kinetics of *in vivo* recombination were tested in a simple *attB*×*attP* reaction time course. Plasmid pZJ21 was used because best inversion reaction performance was achieved with this Int expression plasmid. The time course was carried out in strain DS941/pZJ19off+pZJ21. Plasmid DNA was isolated from cells at different time points and checked by restriction digestion and agarose gel electrophoresis (Fig. 3.16).

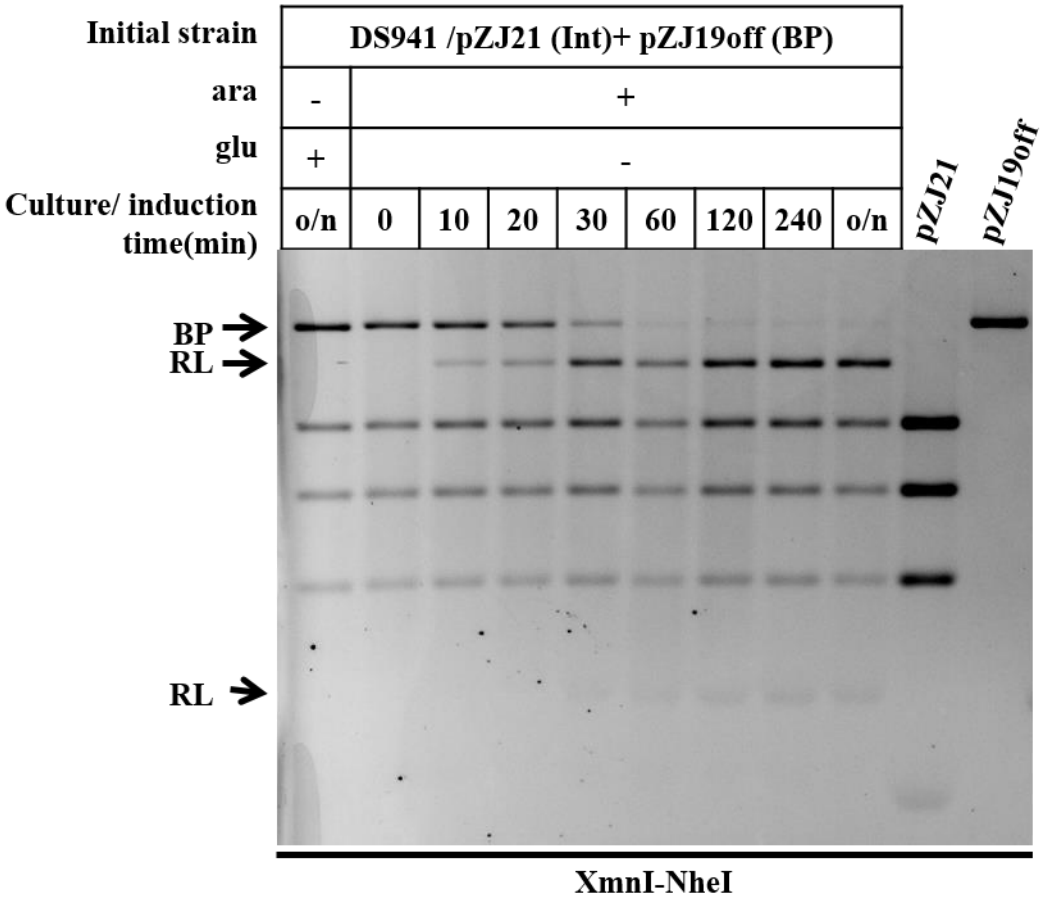


Figure 3. 16 Electrophoretic analysis of the attB×attP time course reaction. Strain DS941/ pZJ19off (BP) + pZJ21 (Int) was cultured in L-Broth with glucose (glu) overnight. The overnight culture was diluted 40-fold in L-Broth, precultured for 1.5 h to mid-log phase before adding arabinose (ara) for different induction time. Cells were centrifuged and frozen by liquid nitrogen. DNA extracted from cells was digested with XmnI and NheI (pZJ19off, BP: 92 bp and 4562 bp, pZJ19on, RL: 806 bp and 3848 bp, pZJ21: 2980 bp, 2229 bp, 1499 bp, 411 bp, and 56 bp) to check the recombination result (bands smaller than 400 bp are not shown on the gel picture).

The DNA state visualized by gel electrophoresis represented the instantaneous state at each time point, since all cellular activity was terminated instantaneously by a freezing in liquid nitrogen as soon as cells were harvested for DNA purification. The recombination product was first visible at 10 min, most of the substrate plasmids were recombined by 30 min, and

reaction reached it final state by one hour (Fig.3.16). This suggests that the expression of Int starts very quickly after adding inducer, and recombination starts very quickly and reaches completion within one hour.

3.4 Conclusion and discussion

In this chapter, ϕC31 integrase and Gp3 were investigated for efficiencies of inversion recombination *in vitro* and *in vivo*. For the *in vitro* assay, how recombination efficiencies varied with different concentrations of proteins and reaction times were analysed quantitatively. For the *in vivo* assay, how recombination efficiencies varied with different Int or Gp3 expression levels and the relative concentrations between proteins and binding sites were analysed qualitatively. In this section, results of these assays are discussed and compared in order to interpret the behaviours of inversion recombination mediated by ϕC31 integrase.

Overall, ϕC31 integrase alone at its optimal concentration can efficiently mediate *attB*×*attP* recombination reaction. Sufficient ϕC31 integrase and Gp3 together in an optimal concentration can efficiently mediate the *attR*×*attL* recombination reaction. For both *in vitro* and *in vivo* reactions, the maximum *attB*×*attP* reaction efficiency (*in vitro*: about 70%, *in vivo*: about 100%) achieved was higher than the maximum *attR*×*attL* reaction efficiency (*in vitro*: about 50%, *in vivo*: about 70%). This may be caused by the competition between the *attB*×*attP* and *attR*×*attL* reactions in the presence of both Int and Gp3, since the *attB*×*attP* reaction cannot be completely inhibited even with large amount of Gp3. To further improve this efficiency, more information about recombination mechanism is required.

Lower recombination efficiencies had been achieved *in vitro* than *in vivo* for both *attB*×*attP* and *attR*×*attL* recombination. The *in vitro* recombination efficiency cannot be further improved just through increasing the concentrations of proteins, since the saturated recombination efficiencies have been achieved even with medium protein concentrations (Fig. 3.3 and 3.4). The explanation for the lower efficient *in vitro* reaction might be that some inactive proteins bind to the sites and compete with the active proteins for recombination sites, or some of the proteins lose activity before all of the substrates are recombined. In living cells, the situation is different since the replication of substrate DNA may remove the inactive proteins from the binding sites and these sites can be recombined by newly expressed proteins to achieve highly efficient recombination. Different reaction conditions

for *in vivo* and *in vitro* reactions might also be a reason for the different recombination efficiencies.

In the *in vivo* test, intermolecular recombination was observed when recombining the high copy-number substrate pZJ16off, resulting in plasmid multimerisation. This problem was mitigated by adding a *cer* recombination site into the substrate plasmid to construct substrate pZJ18off. However, there was still a significant number of multimers produced from substrate plasmid with a *cer* site. One possible explanation was that intermolecular recombination produced large amount multimers in a short time, and Xer proteins in cells were not sufficient to recombine the *cer* sites completely. To test this hypothesis, some future work could be undertaken by retransforming multimer product from both pZJ16off and pZJ18off into strain DS941 and then culturing the transformant cells to check the DNA state from different multimers. If this hypothesis is right, some or most of the multimers with *cer* sites should be recombined to monomers, but the multimers without *cer* sites will be retained in the same form.

Later, the *in vivo* tests suggest that the recombination efficiency is influenced by the expression levels of Int and Gp3, as well as the relative concentrations of protein to DNA binding sites. Increasing the expression level of Int might contribute to a higher *attB*×*attP* reaction efficiency and increasing the ratio of Gp3 to Int might contribute to a higher *attR*×*attL* reaction efficiency. By lowering the substrate copy number and therefore reducing the intracellular *att* sites concentration for Int, both *attB*×*attP* and *attR*×*attL* reaction efficiencies were improved.

The time needed to complete the *in vivo attB*×*attP* reaction after induction (about one hour) was longer than that used for *in vitro* recombination (about half hour). This is probably because it takes time to express and accumulate sufficient Int for recombination, but Int was supplied directly in the *in vitro* reaction. Furthermore, the condition for *in vivo* and *in vitro* reactions can also influence the recombination rate. Cells completing recombination after one hour induction suggests that a one hour pulse of inducer should be able to trigger complete recombination since the expression of Int and the recombination cannot be abruptly terminated after removing the inducer from medium. The *attR*×*attL* reaction *in vivo* was also started with inducing the expression of Int by arabinose. Its kinetics was predicted to be same as the *attB*×*attP* reaction since the recombination rates of both directions were about the same in the *in vitro* time course.

In summary, the initial characterisation and testing demonstrates that efficient inversion reactions can be achieved with ϕC31 integrase and there are three factors influencing the inversion recombination result: 1) the relative concentration between Int and recombination sites; 2) the relative concentration between Gp3 and Int; 3) the recombination time. These factors will be considered during the construction of integrase-based synthetic gene circuits in the following chapters.

4 Design and engineering of plasmid-borne memory devices

4.1 Introduction

The creation of genetic devices for cells to store and process information is a major goal in Synthetic Biology. Biologists have successfully assembled these functional devices based on diverse technologies, such as DNA recombination and regulation of transcription and post-transcriptional process (Inniss and Silver, 2013). Encoding information into the DNA sequence by site-specific recombination, which makes the information heritable and easily detectable, and has low energy cost for the cells, is becoming an attractive technology (Ham *et al*., 2008). During the past decade a number of recombinase-based memory devices have been constructed and tested, including switches (Bonnet *et al*., 2012; Ham *et al*., 2008; Ham and Lee, 2006), logic gates (Bonnet *et al*., 2013; Siuti *et al*., 2013), and counters (Friedland *et al*., 2009). The design principles and behaviours of these devices have been introduced in detail in chapter 1.

The early recombinase-based devices used single-write architecture to record information. For instance, the promoter-inverting system based on λ phage Int (Podhajska *et al*., 1985), the inversion recombination switch based on FimE/FimB and Hin (Ham *et al*., 2008; Ham and Lee, 2006), and the invertase cascade counter based on Flpe and Cre (Friedland *et al*., 2009). In these systems, the amount of information stored was linearly proportional to the number of recombinases used, limiting the information storage capacity.

In the subsequent studies, researchers started to apply large serine-type phage integrases into genetic devices in order to increase the information storage capacity. The large serine-type phage integrase can flip DNA in one direction (write information to DNA), and the same integrase coexpressed with its Xis/RDF can flip the DNA in the reverse direction (rewrite information to DNA). The only successfully constructed device that can write and rewrite information to DNA is a recombinase addressable data module based on BxbI integrase and its RDF gp47 (Bonnet *et al*., 2012). However, this module cannot be used to record multiple occurrences of a single event since its switching is triggered by two different input signals. This property limited the design of a combinatorial genetic counter with this module. To break this limitation and make the most of the DNA rewritable function of serine integrase, new design principles are proposed in this chapter. Low copy-number plasmids are first used as vector for storing information since more complete recombination reactions were

achieved with low copy-number plasmid than with high copy-number plasmid as described in chapter 3, and it is easier to manipulate the plasmid than chromosome.

In chapter 3, it was proved that ϕC31 integrase can invert the orientation of a specific DNA sequence, and ϕC31 integrase coexpressed with its RDF gp3 can mediate the reaction in a reverse direction. This chapter presents the design and engineering of plasmid-borne memory devices by using ϕC31 integrase and RDF gp3 to invert and restore specific DNA sequence.

First, a simple set-reset latch was constructed to test the function of ϕC31 integrase in inverting and restoring the orientation of a specific DNA sequence. In this device, the expression of proteins for different recombination directions (integrase for inverting DNA sequence, and integrase and RDF for restoring DNA sequence) are controlled independently. Depending on the orientation of the target sequence, the expression of a fluorescent reporter can be in either the on or off state. Second, a state-based latch was constructed to explore the possibility of controlling the change of state with a single repeating signal. In this device, *gp3* is placed in the invertible DNA sequence and its expression depends on the sequence orientation and activation of an upstream promoter. There should be no Gp3 expressed when the device is in BP state, and Gp3 expression should be activated when the device is in RL state. If the Gp3 expression is only determined by the sequence state, this device should be expected to change between BP and RL states with a single repeating input pulse. Last, the state-based latch was modified to improve its reaction efficiencies in response to a single input signal. This was done by introducing a time delay for Gp3 expression, resulting in the third device, a binary counting module, which performed like a binary digit. The details about how these devices were constructed, tested, and their performances under different circumstances are introduced in the following sections.

4.2 Engineering of the genetic set-reset latch

4.2.1 Design and construction of the genetic set-reset latch

In electrical engineering, a latch is a circuit that uses two stable states to store information and is a fundamental component in computer science. A biological device which can change between two stable states will be a useful tool for living organisms to store information and control biological processes.

The inversion recombination based on ϕC31 integrase was tested and discussed in chapter 3. The previous experiments tested the inversion recombination in a single directional reaction starting from a substrate in either BP state or RL state. To investigate if Int can invert the substrate DNA via the *attB*×*attP* reaction (set reaction) and restore the substrate DNA via the *attR*×*attL* reaction (reset reaction) in a successive process, a set-reset latch was constructed (Fig. 4.1). In this device, the control element consists of a set generator and a reset generator, which is responsible for converting the extracellular input signals (arabinose or aTc) to intracellular components (Int or Int with Gp3) to write and rewrite information into the data storage. The data storage element is responsible for recording the information in the DNA sequence and transforming the information into an output signal (GFP expression).

This device has a similar structure to that of the rewritable recombinase addressable data module constructed by Bonnet *et al*. (2012), but using ϕC31 integrase instead of Bxb1 integrase. ϕC31 integrase belongs to the same large serine integrase family as Bxb1 integrase and functions in the same way as Bxb1 (Smith *et al*., 2010). Both of these integrases and their RDFs have been described in detail biochemically (Ghosh *et al*., 2006; Khaleel *et al*., 2011).

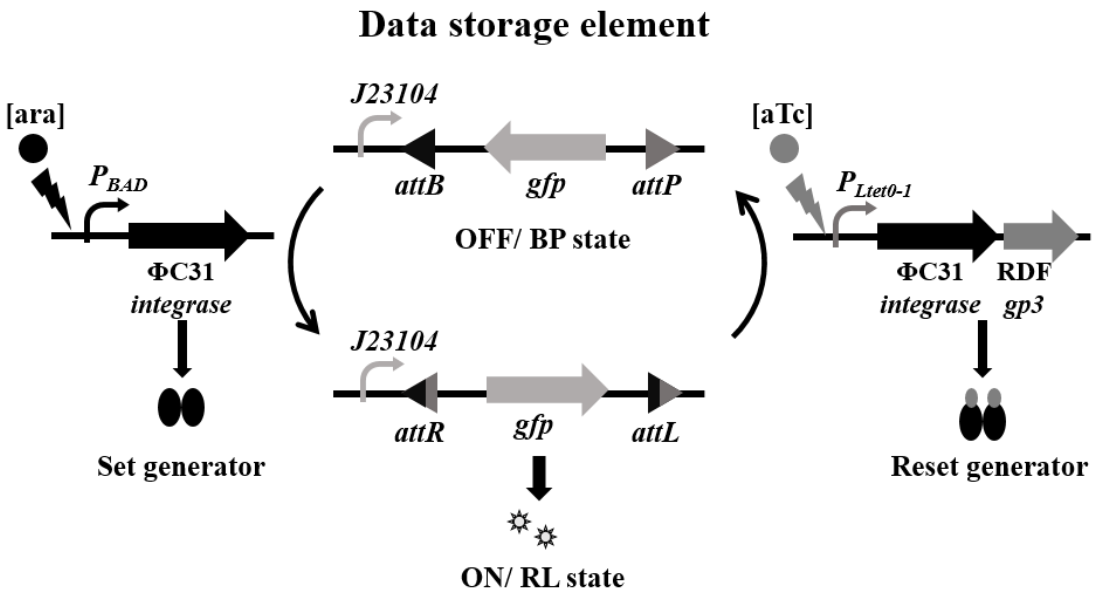


Figure 4. 1 Diagram showing the design principle of set-reset latch. There are two elements in the set-reset latch: in the middle is the data storage (the reporter gene *gfp* flanked by inverted repeat recombination sites *attB* and *attP*. The constitutive promoter J23104 is outside of the invertible segment); both sides of the data storage are the set generator (the ΦC31 *integrase* gene under the control of the inducible P_{BAD} promoter) and reset generator (the ΦC31 *integrase* and *gp3* genes under the control of the inducible $P_{Ltet0-1}$ promoter). For the set reaction, the inducer arabinose drives the expression of Int, inverting the orientation of reporter gene *gfp*, allowing its transcription from the promoter J23104, and switching on the expression of GFP. For the reset reaction: the inducer aTc drives the expression of both Int and Gp3, restoring the orientation the invertible sequence, placing the *gfp* gene in the opposite orientation to the promoter J23104, and switching off the expression of GFP.

To construct this system, the Int expression plasmid pZJ7, which led to efficient *attB*×*attP* recombination in previous tests (section 3.3.2 and section 3.3.3), was selected to work as the set generator. Plasmid pZJ123 was constructed by cloning the *int* gene and *gp3* gene under the control of an inducible $P_{Ltet0-1}$ promoter (activated by aTc) to work as the reset generator. The pZJ123 plasmid has the following characteristics: medium copy-number plasmid with pBR322 origin of replication, ampicillin resistance gene, and *gp3* (with a strong RBS sequence, predicted translation initiation rate: 19612, Salis *et al*., 2009) downstream of the *int* (with a weak RBS sequence, predicted translation initiation rate: 227, Salis *et al*., 2009), which was expected to give a high ratio of Gp3 to Int.

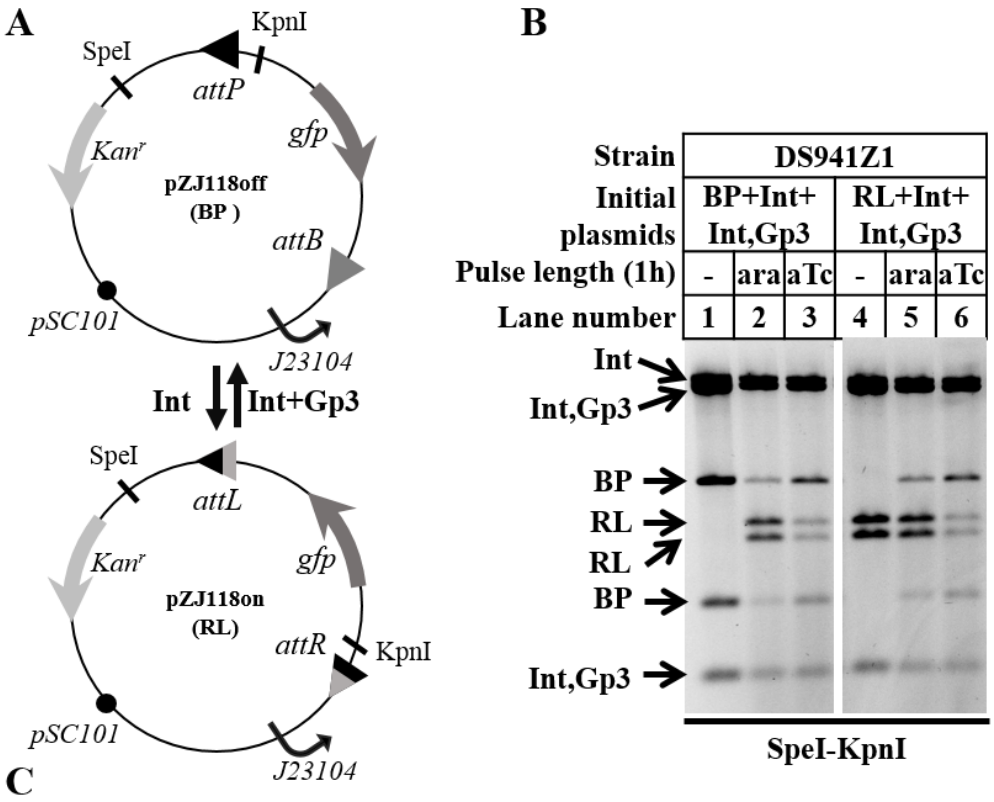
The OFF-state data storage element pZJ118off (Fig. 4.1) was constructed by cloning a synthetic sequence (inverted repeat *attB* and *attP* sites flanking a reporter gene *gfp* in the opposite orientation to a constitutive promoter J23104 in pSWITCH2, details of sequence in section 2.6) into the vector pMS183. The substrate plasmid pZJ118off has a pSC101 origin of replication (~5 copies/cell) and kanamycin resistance gene. The ON-state data storage element pZJ118on was selected from the *in vivo* recombination product of pZJ118off (as described in section 2.31). All these plasmids were used to test the performance of the genetic set-reset latch.

4.2.2 Investigations of the recombination efficiency with the set-reset latch

Previously, the *E.coli* strain DS941 was used as the host strain for the *in vivo* tests, in which the $P_{Ltet0-1}$ is a constitutive promoter since the Tet repressor is absent from this strain. To make the $P_{Ltet0-1}$ promoter inducible, strain DS941 was modified to express the Tet repressor. This was done by moving a Z1 module (Lutz and Bujard, 1997), containing two copies of the *lacI* gene and one copy of the *tetR* gene (encoding the lactose repressor and the tetracycline repressor) driven by the constitutive promoters P_{lacIq} and P_{N25} respectively, as well as the spectinomycin resistance gene from *E.coli* strain MG1655Z1 to DS941 via P1 transduction (section 2.10). The new strain was named DS941Z1.

The substrate plasmids pZJ118off or pZJ118on (data storage element), Int expression plasmid pZJ7 (set generator), and Int and Gp3 expression plasmid pZJ123 (reset generator) were transformed into strain DS941Z1. Single directional recombination was first carried out with DS941Z1/pZJ118off+pZJ7+pZJ123 and DS941Z1/pZJ118on+pZJ7+pZJ123. The previous time course showed that the set reaction reached completion in one hour after the expression of Int was activated by arabinose (Fig. 3.16). Therefore, one hour was used as

the length of input pulse for these experiments. Transformant cells were first cultured in medium with repressor (glucose) overnight, diluted 1:40 in medium without repressor and inducer, and then precultured for 30 min before adding the input pulse. The pulse was terminated by diluting cells 1:1000 in medium with repressor, and growing overnight prior to preparing DNA for analysis (as described in section 2.27.2). The DNA states were visualised by gel electrophoresis and the percentages of DNA in either BP or RL state under different induction conditions were calculated though quantifying the bands on the agarose gel (Fig. 4.2).



C

Initial state		BP			RL		
Set input arabinose		0	1	0	0	1	0
Reset input aTc		0	0	1	0	0	1
Output state	BP state	100%	25%	67%	0%	26%	72%
	RL state	0%	75%	33%	100%	74%	28%
Lane No.		1	2	3	4	5	6

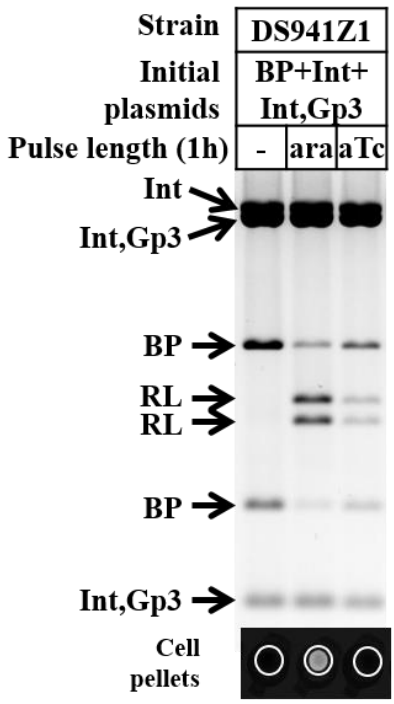
Figure 4. 2 Electrophoretic analysis of the set-reset latch in the single directional reaction. (A) Expected change of substrate plasmids upone the inversion recombination reaction. Int alone is expected to recombine the substrate plasmid pZJ118off (BP) to pZJ118on (RL). In the presence of Gp3, Int is expected to recombine the substrate plasmid pZJ118on to pZJ118off. Plasmid pZJ118off gives two fragments (size: 3145 bp and 1391 bp) and pZJ118on gives two fragments (size: 2419 bp and 2117 bp) after restriction digestion with SpeI and KpnI (the drawing is not to scale). (B) Electrophoretic analysis of the set or reset reaction. DS941Z1 containing substrate plasmid pZJ118off (BP) or pZJ118on (RL), Int expression plasmid pZJ7 and Int plus Gp3 expression plasmid pZJ123 was induced with one hour pulse of arabinose or aTc (section 2.27.2). DNA extracted from cells was digested with SpeI and KpnI (pZJ7: 7175 bp, pZJ123: 7154 bp and 776 bp) and run on the agarose gel. (C) Percentages of DNA in either BP or RL states were calculated by quantifying the bands corresponding to recombined and unrecombined DNA (section 2.33). Percentages shown in the grey boxes are discussed.

The substrate plasmids maintained their initial state when cells were cultured in medium without input signal (lanes 1 and 4, Fig. 4.2B). Highly efficient set reaction (75%) and less efficient reset reaction (26%) were observed when cells were induced by the set input arabinose (lanes 2 and 5, Fig. 4.2B). Whereas, less efficient set reaction (33%) and highly efficient reset reaction (72%) were observed when cells were induced by the reset input of aTc (lanes 3 and 6, Fig. 4.2B).

These results indicate that 1) Substrate DNA maintains its state well in the absence of any input signal. 2) Arabinose can induce Int expression, changing substrate DNA to RL state efficiently (~75%). 3) aTc can induce Int and Gp3 expression, changing substrate DNA to BP state efficiently (~70%). 4) There might be some leakage of Gp3 expression during the arabinose induction, resulting in the reset reaction (26%). 5) Similar final states can be reached with the same induction regardless of the initial state.

Based on the indications above, it was predicted that the set and reset reaction efficiencies would maintain the same when a full cycle of set and reset operations was conducted with strain DS941Z1/pZJ118off+pZJ7+pZJ123. One hour induction time was used for the pulsed induction (as described in section 2.27.2). The DNA states were visualised by gel electrophoresis and the fluorescent state of cell cultures were checked using Bio-Rad blue light Transilluminator (Fig. 4.3A). The percentages of DNA in either BP or RL states under different induction conditions were calculated though quantifying the bands on the agarose gel (Fig. 4.3B).

A



B

State	Initial state	After arabinose induction	After aTc induction
BP state	100%	23%	69%
RL state	0%	77%	32%

Figure 4. 3 Electrophoretic analysis of the genetic set-reset latch in a full cycle of set-reset reaction. (A) DS941Z1/pZJ118off (BP) + pZJ7 (Int) + pZJ123 (Int, Gp3) was induced in a cycle of set (with one hour pulse of arabinose) and reset (with one hour pulse of aTc) reactions (as described in section 2.27.2). DNA extracted from cells after each induction was digested with SpeI and KpnI and run on the agarose gel. Cells after each induction were centrifuged and the fluorescence of cell pellets in the bottom of microcentrifuge tubes was checked using Bio-Rad blue light Transilluminator and shown below the gel picture. (B) Percentages of DNA in either BP or RL states were calculated by quantifying the bands corresponding to recombined and unrecombined DNA (as described in section 2.33).

Similar to the efficiencies achieved in the single induction process (Fig. 4.2), about 77% of the DNA substrate was changed from BP state to RL state after cells were induced with arabinose, and about 69% of the DNA substrate was restored to BP state following the aTc induction according to the gel quantification results (Fig. 4.3B). The fluorescence of cell pellets in microcentrifuge tubes were changed from weak, to strong, then to weak, corresponding to the DNA states (higher percentage of RL state DNA would be expected to express higher level of GFP, resulting in higher fluorescent signal, *vice versa*).

Both the set and reset reactions with this genetic latch were efficient, and it should be possible to further improve the efficiency through adjusting the expression levels of Int and Gp3 according to the results in chapter 3. The set-reset latch showed that ϕC31 integrase can

invert the orientation of a specific DNA sequence and restore the sequence in a successive operation, which confirmed the feasibility of applying this integrase in extending the rewritable module and in the following design systems. As the set-reset latch is not the primary subject of this study, no further characterisation and optimisation of this device was undertaken.

4.3 Engineering of the state-based latch (two-signal controlled system)

For the latch device described previously (section 4.2), the set and reset operations were triggered by two different input signals. The ultimate aim of this study was to construct a device that would be switched between two distinct states under the control of a single signal, and its next state would depend on the current one. To take the project one step forward, a new genetic device, named state-based latch was constructed. The engineering process is described in this section.

4.3.1 Design and construction of the state-based latch

The device was named state-based latch because expression of RDF gp3 was connected to the state of the device. The control element of this latch consists of only an Int generator, in which Int expression is under the control of an arabinose inducible P_{BAD} promoter. In the data storage element, *gp3* gene and the constitutive J23119 promoter are flanked by inverted repeat recombination sites *attB* and *attP*, and the $P_{Ltet0-1}$ promoter and the reporter gene *gfp* are placed outside of the recombination sites (Fig. 4.4). The generator is responsible for converting the extracellular input signal (arabinose) to an intracellular component (Int), writing information into the data storage. The data storage element is responsible for recording the information in the DNA sequence, transforming the information into an output signal (GFP expression), as well as controlling the expression of RDF gp3.

In strain DS941Z1, the $P_{Ltet0-1}$ promoter is an aTc regulated promoter. After induction with arabinose, Int would mediate recombination between *attB* and *attP* (set reaction), changing the device to RL state, allowing transcription of *gp3* from the $P_{Ltet0-1}$ promoter and transcription of *gfp* from the J23119 promoter. Subsequently, the expression of Int and Gp3 would be induced with arabinose and aTc to mediate recombination between *attR* and *attL* (reset reaction), restoring the device to BP state, placing *gp3* gene to an opposite orientation to promoter $P_{Ltet0-1}$, and switching off the GFP expression. In this strain, the expression of Gp3 would be controlled by both the sequence state and the inducible $P_{Ltet0-1}$ promoter.

In contrast, the $P_{Ltet0-1}$ promoter is a constitutive promoter in strain DS941 and the device would behave differently in this *E. coli* strain. After induction with arabinose, Int would mediate recombination between *attB* and *attP*, changing the device to RL state, starting transcription of *gp3* from $P_{Ltet0-1}$ promoter and transcription of *gfp* from J23119 promoter. The expression of Gp3 would be switched on automatically, allowing Int to restore the device to BP state. The state of the system would be expected to be induced to switch both from OFF to ON and from ON to OFF by a single signal (arabinose).

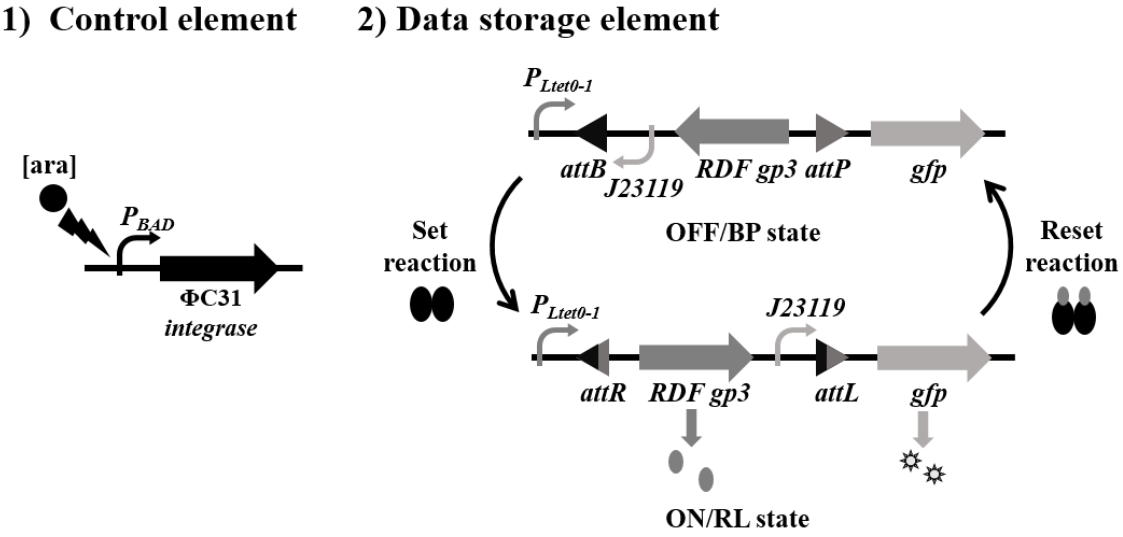
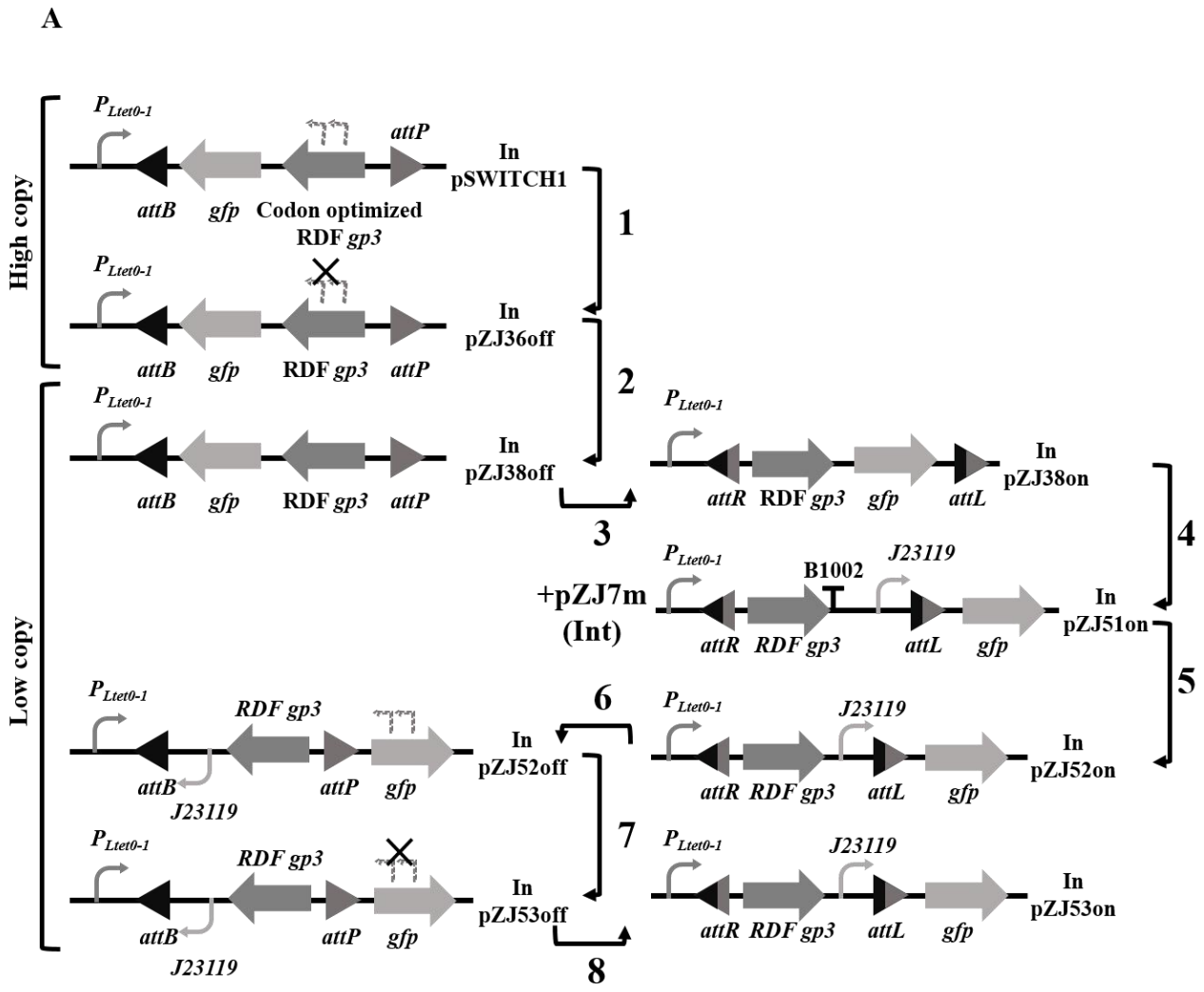


Figure 4. 4 Diagram showing the design principle of the state-based latch. There are two elements in the state-based latch: 1) the control element containing the set generator (ϕC31 integrase gene under the control of the inducible P_{BAD} promoter), and 2) the data storage element (*gp3* gene and the constitutive J23119 promoter are flanked by inverted repeat recombination sites *attB* and *attP*, and the $P_{Ltet0-1}$ promoter and reporter gene *gfp* are outside of the invertible segment). For set reaction from OFF to ON state, inducer arabinose drives the expression of Int, inverting the orientation of the *gp3* gene and the constitutive J23119 promoter, placing *gp3* gene in the same orientation as the $P_{Ltet0-1}$ promoter and switching on GFP expression. For the reset reaction from ON to OFF, Int expression can be induced with arabinose, and Gp3 expression can be induced with aTc (in strain DS941Z1) or be on automatically (in strain DS941), restoring the sequence state and switching off GFP expression.

The initially designed state-based latch described below was different from the one described in Figure 4.4. For the data storage element in the initial design, both *gfp* and *gp3* genes are flanked by the inverted repeat recombination sites *attB* and *attP* (Fig. 4.5, in pSWITCH1). This device was expected to have the similar performance under set and reset operation, and the potential to behave under the control of a single signal. However, experimental results of this initial state-based latch showed undesirable performance, and a series of recombination and optimisation experiments were carried out to improve the performance this genetic device (Fig. 4.5). Finally, an optimal design for the state-based latch was determined as described at the beginning of this section (Fig. 4.4)



B

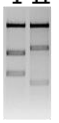
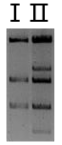
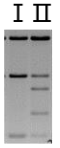
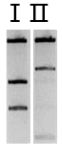
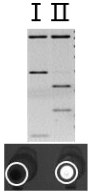
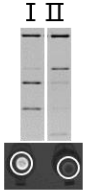
Strains		Performances			
			attB×*attP*		*attR*×*attL*
a	DS941Z1/ pSWITCH1				
b	DS941Z1/ pZJ36off				
c	DS941Z1/ pZJ38off/on+ pZJ7m		>90%		>80%
d	DS941Z1/ pZJ51off/on+ pZJ7m	Not determined			~20%
e	DS941Z1/ pZJ52off/on+ pZJ7m		~50%		>90%
f	DS941Z1/ pZJ53off/on+ pZJ7m		>90%		>90%

Figure 4. 5 Optimisation of the data storage element of the state-based latch.

(A) Pictures showing the architecture change of the data storage element (1-8).

1) Change pSWITCH1 to pZJ36off. The codon optimised *gp3* was changed to the natural ϕC31 *gp3*.

2) Change pZJ36off to pZJ38off. The invertible substrate sequence was moved from high copy number vector (ColE1 origin) to low copy number vector (pSC101 origin).

3) Selection of the substrate plasmid pZJ38on from the recombination product of pZJ38off.

4) Change pZJ38on to pZJ51on.The architecture was changed by moving the *gfp* out of the recombination sites and having a constitutive promoter to control its expression.

5) Change pZJ51on to pZJ52on. The terminator B1002 was removed.

6) Select the substrate plasmid pZJ52off from the recombination product of pZJ52on.

7) Change pZJ52off to pZJ53off. The *gfp* was mutated to remove the predicted promoter.

8) Selection of the substrate plasmid pZJ53on from the recombination product of pZJ53off.

(B) Performance of strains containing corresponding data storage element (a-f).

a) DS941Z1/pSWITCH1 was cultured in L-Broth overnight. To check the fluorescent signal, two droplets (5 μl) of this culture on a clean plate were visualized by Bio-Rad blue light Transilluminator (the droplets were highlighted by white circles)

b) DS941Z1/pZJ36off was cultured in L-Broth overnight. To check the fluorescent signal, two droplets (5 μl) of this culture on a clean plate were visualized by Bio-Rad blue light Transilluminator (the droplets were highlighted by white circles)

c-f) The set reactions were triggered by culturing strain containing substrate plasmid (BP state) and Int expression plasmid in L-Broth with glucose or arabinose overnight, DNA extracted from cells was digested and run on the agarose gel (Ⅰ, glucose ; Ⅱ, arabinose). The reset reactions were carried out by culturing strain containing substrate plasmid (RL state) and Int expression plasmid in L-Broth with glucose or arabinose plus aTc overnight, DNA extracted from cells was digested and run on the agarose gel (Ⅰ, glucose; Ⅱ, arabinose+aTc). The overnight cultures from strain DS941Z1/pZJ53off+pZJ7m or DS941Z1/pZJ53on+pZJ7m were centrifuged and the fluorescence of cell pellets in the microcentrifuge tubes were checked using Bio-Rad blue light Transilluminator and shown below the gel pictures (f, the pellets are highlighted with white circles).

To construct the initial designed state-based latch, the sequence for the data storage element was synthesised by gene synthesis company GeneArt (details of sequence in section 2.6). The synthesised DNA is in a high copy-number plasmid named pSWITCH1, which has a ColE1 origin of replication and kanamycin resistance gene. Since in strain DS941Z1 (*TetR*$^+$) the expression of Gp3 was more controllable, this strain was first used to test the recombination efficiencies with the state-based latch.

After strain DS941Z1 was transformed with the plasmid pSWITCH1 and cultured in L-Broth overnight, a strong fluorescent signal of DS941Z1/pSWITCH1 was detected (Fig. 4.5B-a). The *gfp* gene was designed to be in an opposite orientation to the promoter $P_{Ltet0-1}$ which should not be expressed when the substrate was in BP state. According to the online promoter prediction program (Reese, 2001), there are two predicted promoters (Table 4.1) in the codon optimised *gp3* gene upstream of *gfp* gene, which might lead to the leakage of GFP. To remove the predicted promoters, the codon optimised *gp3* was replaced by the natural ϕC31 *gp3* which does not have any predicted internal promoter sequence, resulting in plasmid pZJ36off (process as shown in Fig. 4.5A-1). Strain DS941Z1 containing the resultant plasmid pZJ36off showed much weaker fluorescent intensity than strain DS941Z1 containing substrate plasmid pSWITCH1 (Fig. 4.5B-a and 4.5B-b).

Then, the sequence for the data storage (from pZJ36off) was cloned into a low copy-number vector pMS183 (pSC101 origin of replication, kanamycin resistance gene) to construct pZJ38off plasmid, containing convenient restriction sites (SacI and BamHI) for easy recombination results analysis (Fig. 4.5A-2). Since the *gp3* gene was carried by a low copy-number plasmid, Gp3 expression was expected to be low. Therefore, a weaker Int expression plasmid (pZJ7m with a weak GTG start codon) was used in this section to provide a relative high ratio of Gp3 to Int.

To test the performance of the newly constructed substrate plasmid, the set reaction was first carried out with strain DS941Z1/pZJ38off+pZJ7m in an overnight induction process, and efficient set reaction (>90%, Fig. 4.5B-c) was achieved. To investigate the efficiency of the reset reaction, the substrate plasmid pZJ38on was selected from the recombination product of pZJ38off *in vivo* (section 2.31), which has the $P_{Ltet0-1}$ promoter in the same orientation as *gfp* and *gp3* genes. The reset reaction was carried out with strain DS941Z1/pZJ38on+pZJ7m in an overnight induction process. More than 80% of the data storage sequence was recombined under the induction of arabinose and aTc (Fig. 4.5B-c).

Though both set and reset reactions with this plasmid were efficient, it was difficult to observe the fluorescence change of cells after recombination. The expression of GFP could not be switched ON even when the substrate was in RL state, since the $P_{Ltet0-1}$ promoter is repressed in strain DS941Z1. To solve this problem, the sequence of the data storage element was changed by placing a constitutive promoter between the recombination sites to control the transcription of *gfp* which was moved outside the recombination sites, resulting in pZJ51on (Fig. 4.5A-4). Thus, GFP would be expressed all the time when the substrate was in RL state.

The newly constructed substrate plasmid pZJ51on gave decreased *attR*×*attL* recombination efficiency (~20%, Fig. 4.5B-d), suggesting Gp3 was not being expressed at a high enough level. It was hypothesized that *gp3* transcript might be unstable because terminator B1002 was too close to the end of the *gp3* gene, leading to rapid degradation of the mRNA by 3' to 5' exoribonucleases (Zuo and Deutscher, 2001). To solve this potential problem, the terminator was deleted (Fig. 4.5A-5), as a result, the reset reaction efficiency was increased from 20% to more than 90% with the new substrate plasmid pZJ52on (Fig. 4.5B-e).

However, the set reaction from the corresponding substrate plasmid pZJ52off only had 50% efficiency. Analysis of the *gfp* sequence revealed potential promoter sequences (Table 4.1) which could lead to expression of Gp3 in the BP state, inhibiting the set reaction. The *gfp* gene was therefore mutated to remove the potential promoters, resulting in the substrate plasmid pZJ53off (Fig. 4.5A-7). Optimal set and reset reactions (greater than 90% efficiency) were achieved with the final constructed substrate plasmid (pZJ53off and pZJ53on, respectively) in the overnight induction process (Fig. 4.5B-f). What is more, the fluorescence of cells containing the final constructed state-based latch (based on substrate pZJ53off or pZJ53on) showed the expected change after corresponding induction process (cell fluorescence changed from weak to strong after set reaction; cell fluorescence changed from strong to weak after reset reaction, Fig. 4.5B-f).

Table 4. 1 Predicted promoters in Gp3 and GFP

Name	Sequence information
Gp3	1.TTG GAC GCT GTT CAA AGT CCT GCA TGA AGC GGA GGA TGA TGT GGA GCG (0.92) 2.TGG ACG CTC TTC AAG GTT CTG CAC GAA GCC GAA GAC GAC GTT GAG CGC 3. W T L F K V L H E A E D D V E R
Gp3	1.GCG GAT GGT ATT CAC TGG GAC ATG AAA CTG TGG TTA AAT GGC AAG CTT (0.92) 2.GCT GAC GGT ATC CAC TGG GAT ATG AAG CTT TGG CTG AAC GGC AAG CTG 3. A D G I H W D M K L W L N G K L
Gfp	1.CCG ATT CTG GTT GAA TTA GAT GGC GAT GTC AAC GGA CAT AAA TTC AGT (0.92) 2.CCG ATC CTG GTT GAA CTG GAT GGC GAT GTC AAC GGA CAT AAA TTC AGT 3. P I L V E L D G D V N G H K F S
Gfp	1.ACG GGT GTA GTA CCG ATT CTG GTT GAA TTA GAT GGC GAT GTC AAC GGA (0.97) 2.ACG GGT GTA GTA CCG ATC CTG GTT GAA CTG GAT GGC GAT GTC AAC GGA 3. T G V V P I L V E L D G D V N G

The predicted promoter sequences are shown in the 1st line with promoter scores in parentheses (site-directed mutagenesis nucleic acids are highlighted using grey box). The predicted promoter sequences were replaced with the sequences shown in the 2nd line. Changing of the nucleic acides did not change the corresponding amino acid sequences as shown in the 3rd line.

4.3.2 Characterisation of the state-based latch

In this section, experiments were carried out in detail to test the performance of the state-based latch under different conditions, such as varying induction times, Int expression levels, and strains.

4.3.2.1 Process of continuous time induction reaction

To study instantaneous DNA states during different culture and induction stages, a continuous time course including a full cycle of set and reset operations was carried out with strain DS941Z1/pZJ53off+pZJ7m (section 2.23.2). Cells were centrifuged and frozen to stop the recombination at different induction time points from the same culture. Plasmids in the cells were extracted, digested, and checked by gel electrophoresis (Fig. 4.6).

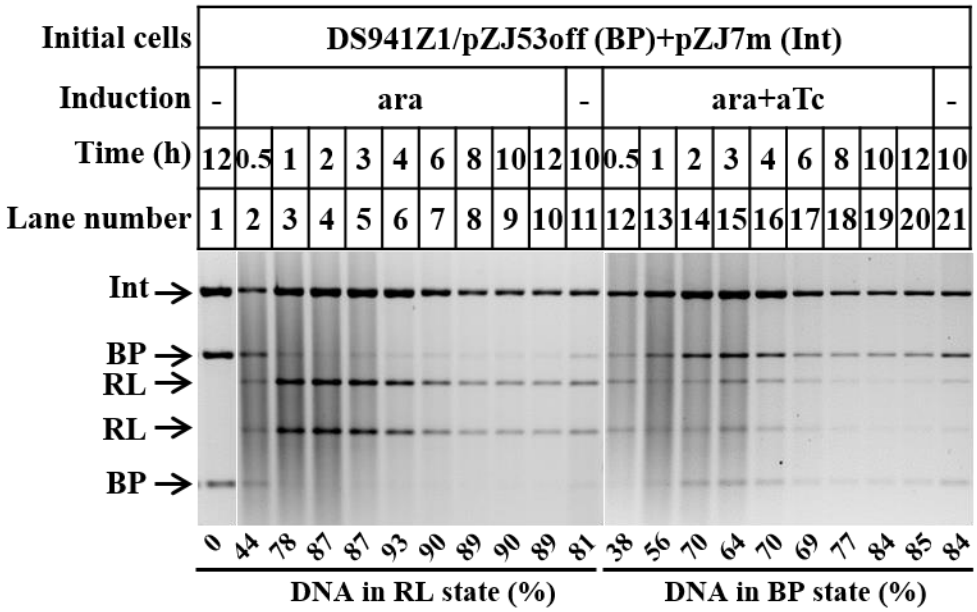


Figure 4. 6 Electrophoretic analysis of the state-based latch under continuous induction time course. DS941Z1/pZJ53off (BP) + pZJ7m (Int) was inoculated in L-Broth with glucose overnight (lane 1). 4 ml of overnight culture was centrifuged and washed into 160 ml L-Broth, precultured for 90 min to mid-log phase before adding arabinose. Cells were centrifuged and frozen at different time points after adding arabinose (lane 2-10). After 12 hour induction, 5 ul of liquid culture was diluted into 5 ml L-Broth with glucose and cultured overnight (lane 11). 4 ml of overnight culture was centrifuged and washed into 160 ml L-Broth, precultured for 90 min to mid-log phase before adding arabinose and aTc. Cells were centrifuged and frozen at different time points after adding arabinose (lane 12-20). After 12 hour induction, 5 ul of liquid culture was diluted into 5 ml L-Broth with glucose and cultured overnight (lane 21). To get roughly similar amount of cells, the volumes of cell culture centrifuged at different time points were: 3 ml (lanes 1,8-11,18-21), 20 ml (2-4 and 12-14), 10 ml (lanes 5,6,15, and 16), 4.5 ml (lanes 7 and 17). DNA extracted from cells was digested with SpeI (pZJ7:7175 bp, pZJ53off, BP:1459 bp and 3917 bp, pZJ53on, RL:2198 bp and3178 bp) and run on the agarose gel. Percentages of DNA in either RL or BP state were calculated after gel quantification.

For the set operation, the substrate plasmid maintained BP state stably when cells were cultured in L-Broth with glucose (Fig.4.6, lane 1). About 45% of the substrate plasmid was recombined to RL state at 30 min with arabinose induction (lane 2, Fig.4.6) and the reaction reached to a final state in two hour with about 90% efficiency (lanes 4, Fig.4.6). For the subsequent reset operation, recombination started with a mixture of substrate plasmids in BP or RL state (81% in RL state, lane 11, Fig.4.6), because the previous set operation did not recombine all of the substrate plasmid to RL state. Nevertheless, more than 70% substrate plasmid was restored to BP state in two hour (lane 14, Fig.4.6).

4.3.2.2 Reactions with different lengths of induction time

To test the influence of the length of induction time on recombination efficiency, different lengths of induction were used in a pulsed time course reaction including a full cycle of set and reset operations with strain DS941Z1/pZJ53off+pZJ7m. This strain was cultured and induced with the pulsed induction (section 2.23.2). After different pulses of induction, cell

cultures were diluted 1:1000 in L-Broth without inducer and cultured overnight. Plasmids from the overnight culture were extracted, digested, and checked by gel electrophoresis (Fig. 4.7).

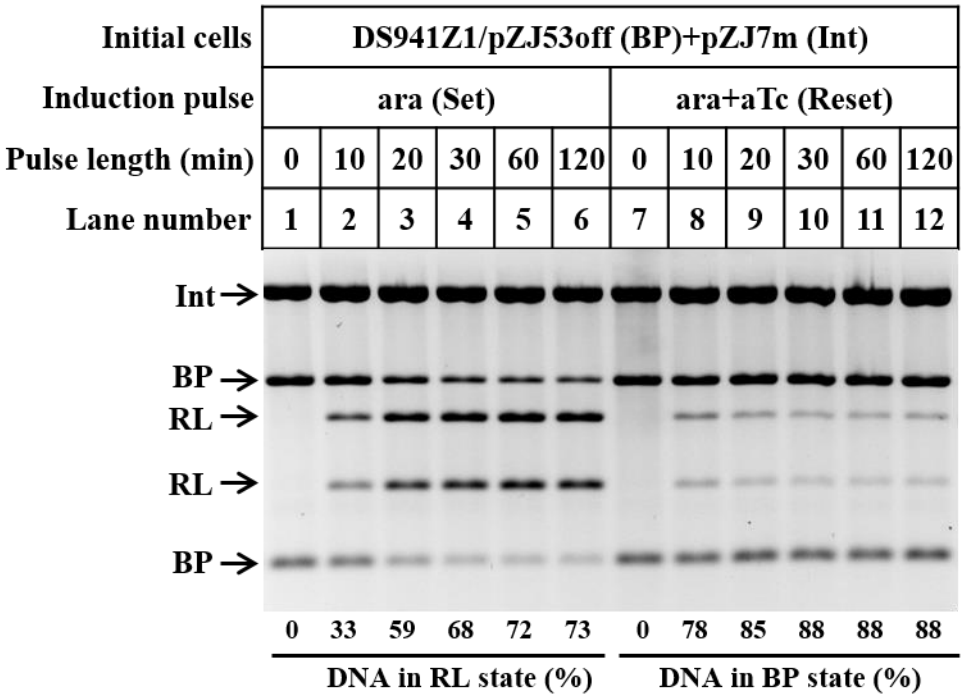


Figure 4. 7 Electrophoretic analysis of the state-based latch under pulsed induction time course. DS941Z1/pZJ53off (BP) + pZJ7m (Int) inoculated in L-Broth with glucose overnight. The overnight culture was diluted 40-fold in 5 ml L-Broth, precultured for 90 min to mid-log phase before adding arabinose pulse with different lengths, and then cells were diluted 1: 1000 into L-Broth with glucose and cultured overnight (lanes 1-6). Each overnight culture was diluted 40-fold in 5 ml L-Broth, precultured for 90 min to mid-log phase before adding arabinose and aTc pulse with different lengths, and then cells were diluted 1: 1000 into L-Broth with glucose and cultured overnight (lanes 7-12). DNA extracted from cells after each induction was digested with SpeI and run on the agarose gel. Percentages of DNA in either RL or BP state were calculated after gel quantification.

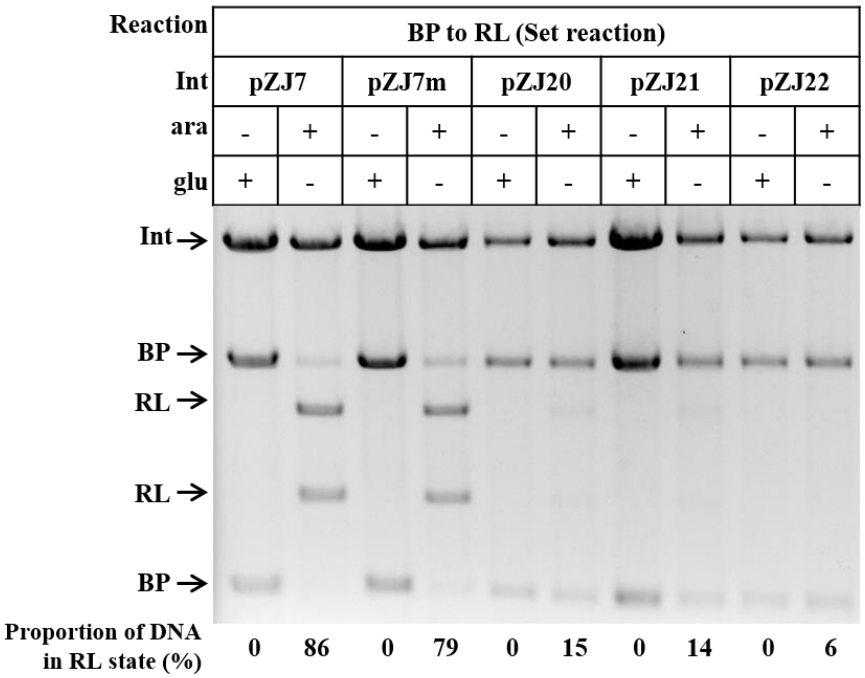
The pulsed induction time course indicated that both set and reset reactions reached the most efficient recombination points (~75% and ~90%, respectively) within one hour induction time (Fig. 4.7), and longer induction time did not improve the recombination efficiency. However, the highest set reaction efficiency (~75%, lane 6, Fig. 4.7) achieved with pulsed induction time was lower than the efficiency (~90%) achieved in the continue induction (Fig. 4.6). The reason might be that the concentration of Int decreased after removing its inducer and the leakage of Gp3 may promote Int to reset some of the data storage to PB state.

4.3.2.3 Reactions with different Int expression plasmids

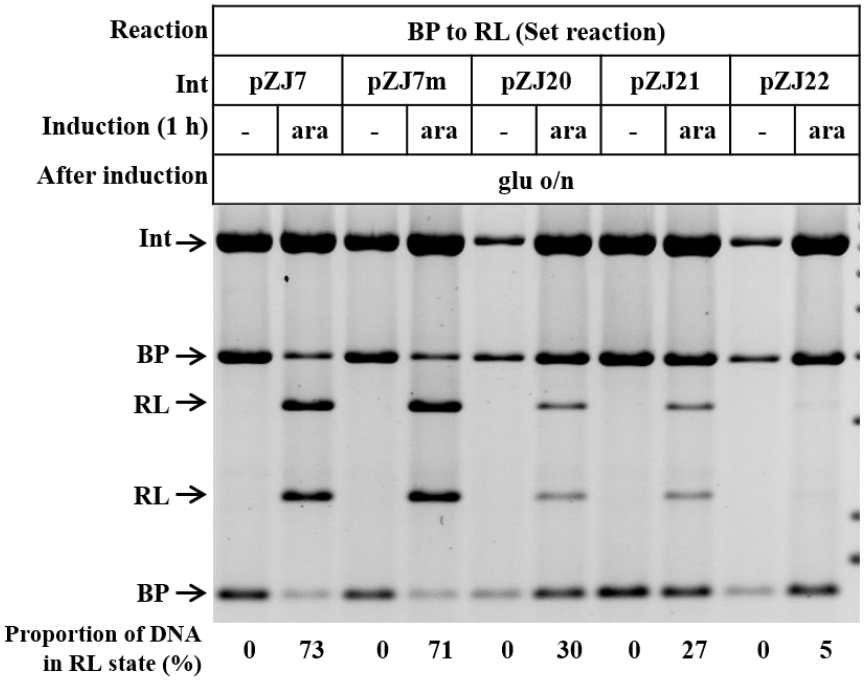
To study if the Int expression level influences the recombination efficiency, experiments were carried out with the previously constructed Int expression plasmids (section 3.3.2.4).

Both continuous and pulsed inductions were carried out with strain DS941Z1 containing substrate plasmid pZJ53off (for set reaction) or pZJ53on (for reset reaction) and one of Int expression plasmids (pZJ7, pZJ7m, pZJ20, pZJ21, or pZJ22, Fig.4.8). The details of these integrase expression plasmids were described in chapter 3 (Table 3.1).

A. Overnight induction with arabinose



B. One hour induction with arabinose



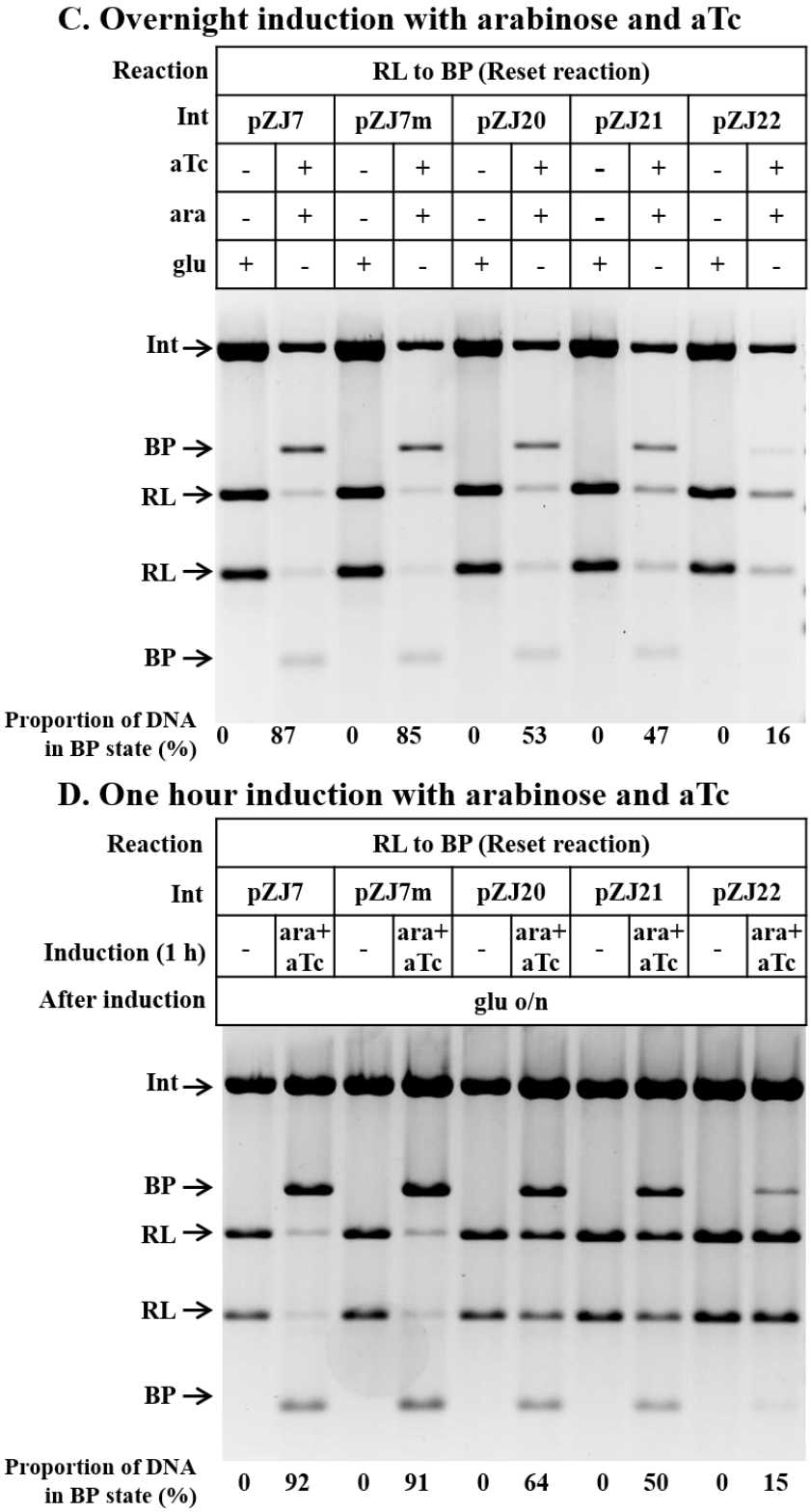


Figure 4. 8 Set and reset reaction of the state-based latch with different expression levels of Int. For the set operation, DS941Z1 was transformed with substrate pZJ53off (BP) and Int expression plasmid pZJ7, pZJ7m, pZJ20, pZJ21, or pZJ22, and induced with either overnight (12h, A) or pulsed (1 h, B) arabinose induction. DNA extracted from cells was digested with SpeI and run on an agarose gel. For the reset operation, DS941Z1 was transformed with substrate pZJ53on (RL) and Int expression plasmid pZJ7, pZJ7m, pZJ20, pZJ21, or pZJ22, and induced with either overnight (12 h, C) or pulsed (1 h, D) arabinose and aTc induction. DNA extracted from cells was digested with SpeI and run on an agarose gel. The proportation of substrate in either BP or RL state were calculated after gel quantification.

The gel results indicated that different RBS sequences for Int affected recombination efficiencies (Fig. 4.8). Recombination with plasmid pZJ7 and pZJ7m (same RBS sequence and different start codons for Int) showed similar recombination efficiency in all these reactions. Int expressed from plasmid pZJ20, pZJ21, and pZJ22 containing a weaker predicted RBS sequence than that of pZJ7 and pZJ7m, showed insufficient recombination ability, especially when it was used to mediate the set reaction. The inefficient set reaction by Int from pZJ20, pZJ21, and pZJ22 suggested that there might be not enough Int expressed from these plasmids or some leaky expression of Gp3 inhibiting the set reaction. Since highly efficient set (~75%) and reset (~90%) reactions were achieved with Int expression plasmid pZJ7 under one hour pulsed induction, this plasmid was used in the following tests.

4.3.2.4 Reactions in different strains

To determine whether the *E. coli* strain used had an effect on the recombination efficiency, reactions were carried out in strain MG1655Z1 (*rec*$^+$), DS941Z1 (*rec*F), or DH5αZ1 (*rec*A1). Each strain containing substrate plasmid pZJ53off or pZJ53on and Int expression plasmid pZJ7 was induced by either overnight or one hour induction (Fig. 4.9).

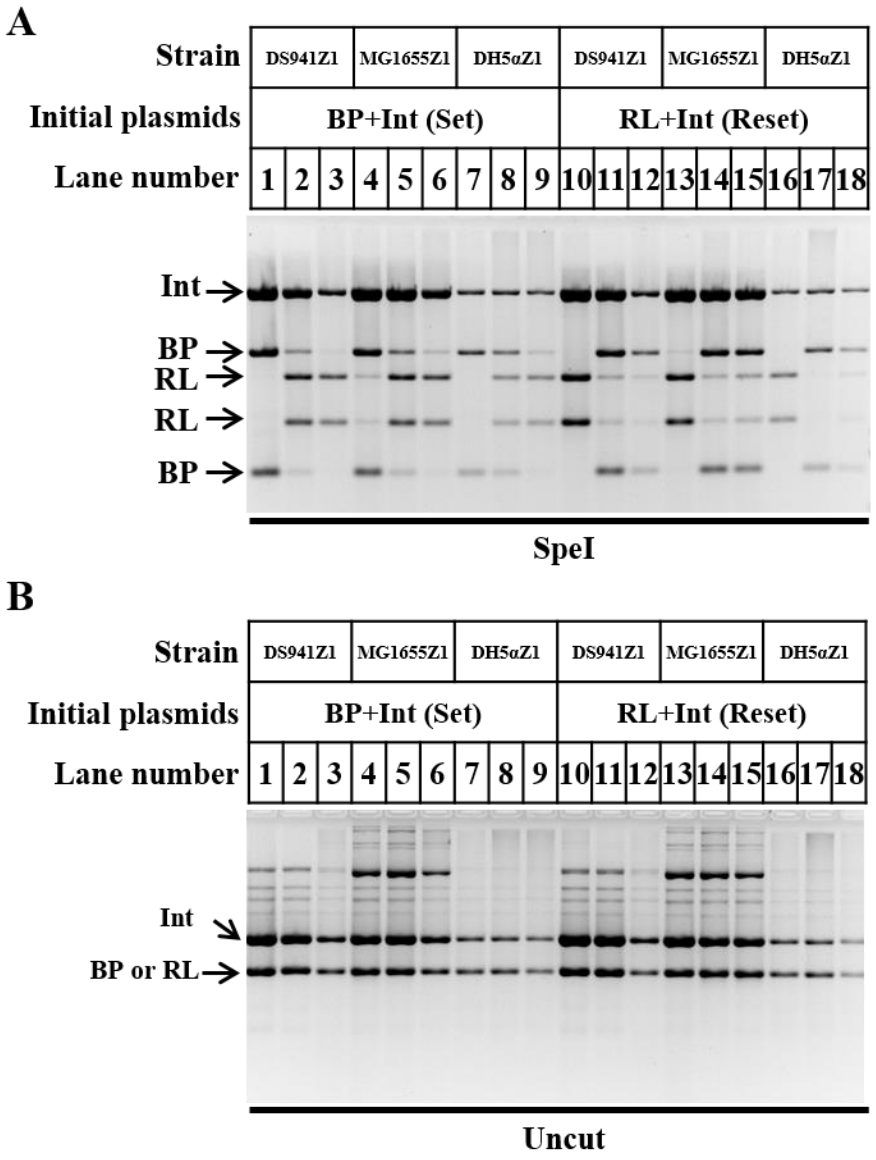


Figure 4. 9 Performance of state-based latch in different strains. Strain DS941Z1, MG1655Z1 or DH5αZ1 was transformed with substrate plasmid pZJ53off (BP) or pZJ53on (RL), and Int expression plasmid pZJ7, cultured and induced with either pulsed (1 h) or overnight (12 h) induction (using arabinose for the set reaction and arabinose and aTc for the reset reaction). DNA was extracted from cells, and run cut with SpeI (A) and uncut (B) DNA on the agarose gel. The induction conditions used were: culturing in L-Broth with glucose overnight (lanes 1, 4, 7, 10, 13, and 16); inducing by one hour pulse of arabinose (lanes 2, 5, and 8); inducing by 12 h of arabinose (lanes 3, 6, and 9); inducing by one hour pulse of arabinose and aTc (lanes 11, 14, and 17); inducing by 12 h of arabinose and aTc (lanes 12, 15, and 18).

The digestion results suggested that recombination reached similar efficiencies in different strains (Fig. 4.9A), except the pulsed set induction in strain DH5αZ1, which was less efficient than the reaction in other strains (Fig. 4.9A, lane 8). Furthermore, the amount of DNA collected from strain DH5αZ1 was lower compared to other strains under the same culturing conditions. Strain DH5αZ1 has a *rec*A mutation which may cause the cells to grow

slower and be unhealthy, so less plasmid DNA was harvested and the short induction time might not be long enough to induce complete recombination.

The uncut DNA results revealed that more multimers were formed in strain MG1655Z1 compared to others, which was probably caused by homologous recombination (Fig. 4.9B). The overnight induction appeared to reduce plasmid yield in all these strains (Fig. 4.9B, lanes 3, 6, 9, 12, 15, and 18), more obviously in DS941 than in MG1655Z1. This might be because the continuous protein expression increased the load for cells, affecting the cell growth and plasmid yield. Or it was possible that some of the DNA was damaged under continuous exposure to Int, and the damage can be repaired in strain MG1655Z1 (*rec*$^+$) due to the functional DNA repair system, while the plasmid DNA had a chance to recover in overnight growth after removing the input signal in the one hour pulsed induction process (Fig. 4.9B, lanes 2, 5, 8, 11, 14, and 17). In subsequent experiments, the pulsed induction was used as a common way to induce the reaction, since it is less harmful to cell growth and DNA yield.

The strain types did not obviously influence the site-specific recombination efficiency, but had some effects on the plasmid profiles (monomeric or multimeric) and the maintenance of plasmid DNA in cells. Strain DS941Z1 was used as the host strain for the following test because of the limited homologous recombination between plasmids due to the *recF* mutation and good recombination performance under pulsed induction.

4.3.3 Further optimisation of the state-based latch by changing the expression levels of RDF gp3

The relatively low efficiency (~75%) of the set reaction previously seen with Int expression plasmid pZJ7 under one hour pulsed induction was suggested to be caused by the leakage of Gp3 (section 4.3.2.3). It was possible that some unexpected promoters upstream of *gp3* gene lead to the leakage both in BP and RL states, or the leakage was from the $P_{Ltet0\text{-}1}$ promoter in RL state. A weaker RBS sequence for Gp3 was expected to be helpful for lowering the leakage and improving the set reaction efficiency. However, too weak RBS would lower the efficiency of the *attR*×*attL* recombination. Therefore, a library of RBSs for Gp3 was built and a selection process was used to find out the optimal Gp3 expression level for the most efficient recombination in both directions. The whole process is outlined in Figure 4.10, which include design, experimental construction, and screening of the RBSs library. First of all, to make sure that functional Gp3 can be expressed from sequence after changing its RBS, the library was constructed in the RL-state substrate and only the ones that can be

recombined to BP state were selected. This was done by transforming the substrate plasmids containing the RBSs library into *E.coli* cell already carrying Int expression plasmid, culturing the transformants together, and inducing the culture with one hour pulse of arabinose and aTc to recombine the substrate from RL (GFP ON) state to BP state (GFP OFF). Only cells that changed from fluorescent to non-fluorescent were selected. Then further characterisation was carried out to find out the optimal Gp3 expression for both efficient set and reset reaction.

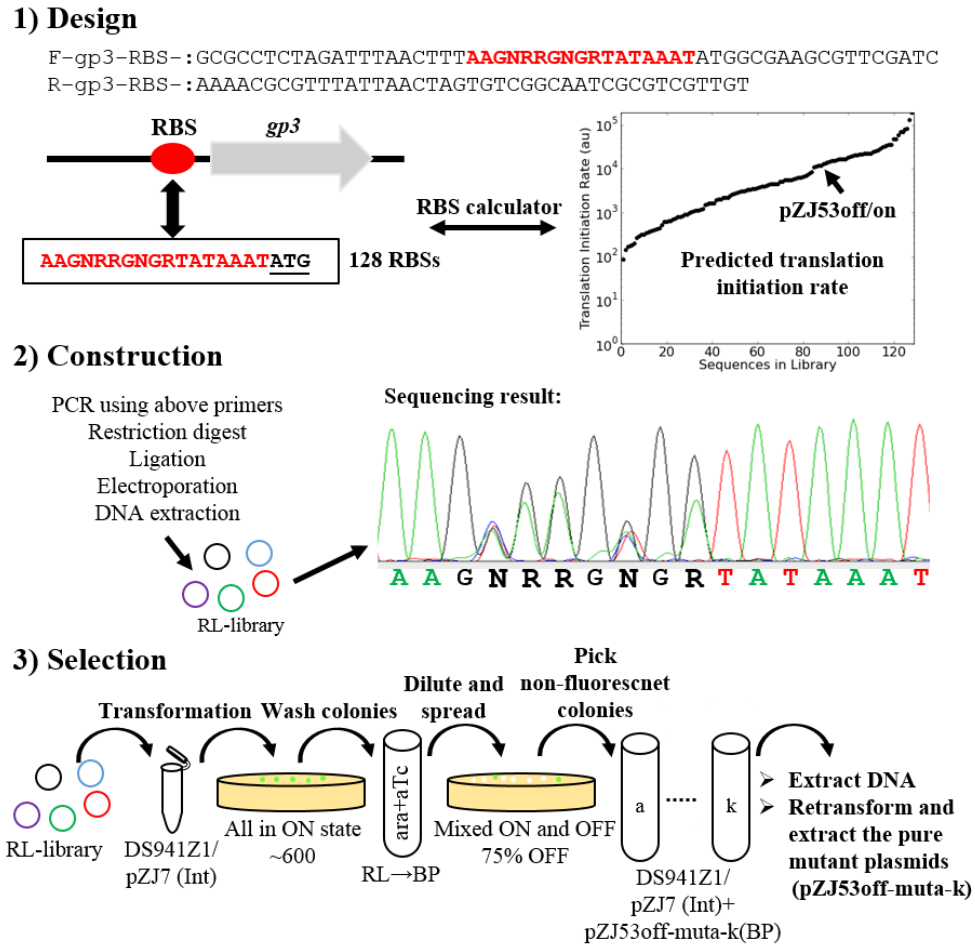


Figure 4. 10 Process of seeking optimal Gp3 expression level for efficient recombination.

1) Design of primers containing the RBS library to amplify *gp3*, providing 128 different RBS sequences with a range of predicted Translation Initiation Rate from 84 to 186955 (Salis *et al.*, 2009).

2) Experimental construction of the substrate plasmids containing the RBSs library. The *gp3* gene was amplified using the designed primer, and the PCR products were digested, inserted to the substrate plasmid pZJ53on (RL) to replace the original *gp3* sequence, and transformed into strain DS941Z1 by electroporation. Resultant colonies were washed from the plate, and the DNA extracted from the cell mixture was sequenced.

3) Experimental selection of substrate plasmids recombined by Int efficiently *in vivo*. The substrate plasmids (RL state) containing RBSs library of *gp3* were transformed into strain DS941Z1/pZJ7. The transformant colonies were washed from plate, induced by one hour pule of arabinose and aTc to recombine the substrate from RL (ON) state to BP (OFF) state. Next, the culture was diluted (10^6 folds) and spread (100 ul) on plate to get single colonies (75% non-fluorescent), and 11 non-fluorescent colonies were picked randomly and inoculated in L-Broth with glucose overnight. DNA extracted from cells was diluted (10 folds) and retransformed into empty strain DS941Z1 to achieve cells only contain substrate plasmids. Then, the pure substrate plasmids (BP) were purified from cells.

Following the process of seeking optimal Gp3 expression level for efficient recombination (Fig. 4.10), 11 random substrate candidates that switched from RL to BP state were proved to have functional *gp3* gene. They were picked and transformed into strain DS941Z1/pZJ7 separately. These substrate candidates were named pZJ53off-a to pZJ53-k. One and a half cycles of set and reset operations were carried out with each tranformant cell, and the fluorescence of cells were visualized and quantified using Typhoon Fluorimager (Fig. 4.11, described as section 2.29.1).

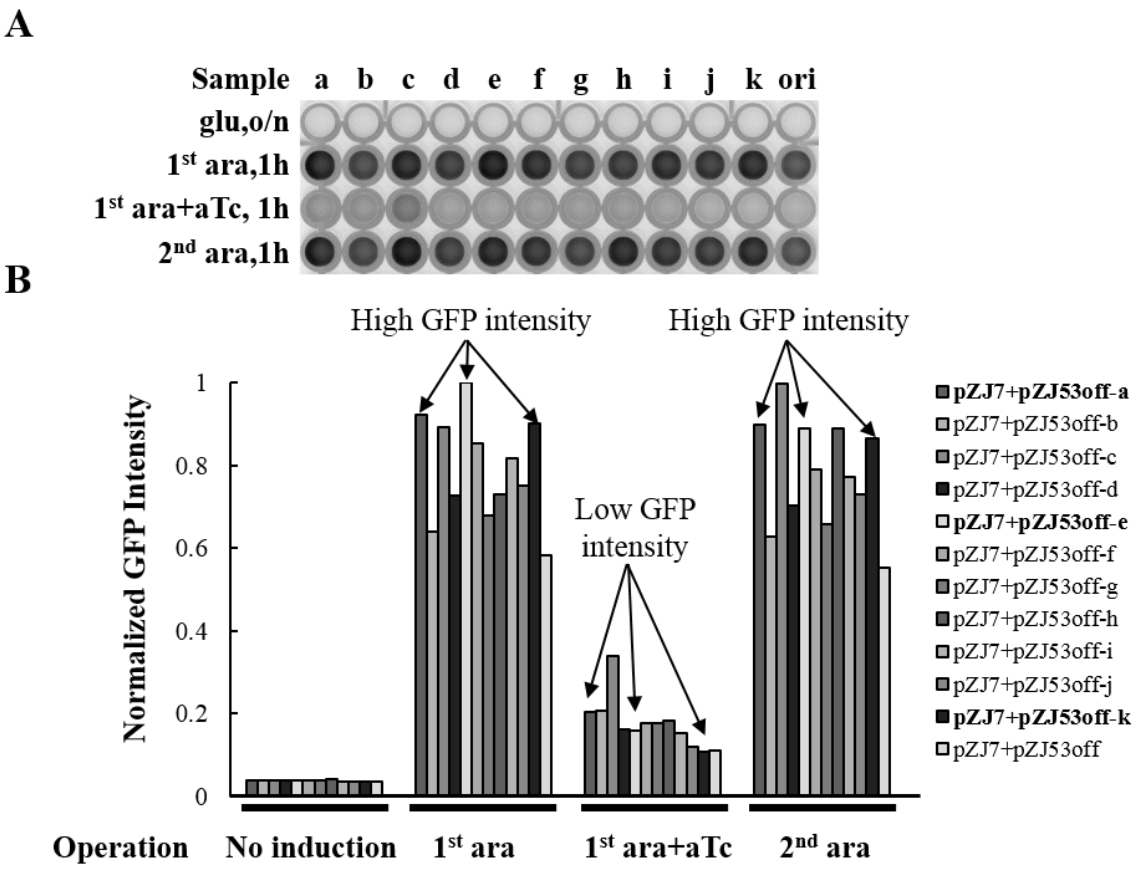


Figure 4. 11 Fluorescent analysis of cells containing substrate with different gp3 RBSs under set and reset reactions. (A) Fluorescent scanning showing fluorescent strength of cell cultures. DS941Z1/pZJ7 (Int) was transformed with substrate plasmid (a-k) selected from the *gp3* RBSs library or the original one (ori). One and half cycles of set (with one hour pulse of arabinose) and reset (with one hour pulse of arabinose and aTc) reactions were carried. Cells after each induction were fixed, and the fluorescence of the cell cultures were imaged using a Typhoon FluorImager. (B) Quantification of the cell culture fluorescence. The GFP intensity of each culture was quantified using Typhoon Fluorimager and the quantification numbers were normalized by setting the highest one to 1.

According to the fluorescence quantification result (Fig. 4.11B), cells containing any of these substrates changed in a "low-high-low-high" pattern of GFP intensity (Fig. 4.11B). The fluorescent signal change indicated that cells containing the substrate mutation pZJ53off-a, pZJ53off-e, and pZJ53off-k gave better set (high GFP intensity) and reset (low GFP intensity)

reaction performance than that with the original substrate pZJ53off. To further confirm this indication, plasmid DNA was extracted from these cells and analysed by restriction digestion and gel electrophoresis (Fig. 4.12). The reactions with substrate pZJ53off-b (low GFP intensity after set reaction), and pZJ53off-c (high GFP intensity after reset reaction) were checked as negative control. All of these substrate plasmids were sequenced to analyse the relationship between the Gp3 RBS sequence and the recombination efficiencies as determined by quantitating the gel (Table 4.2).

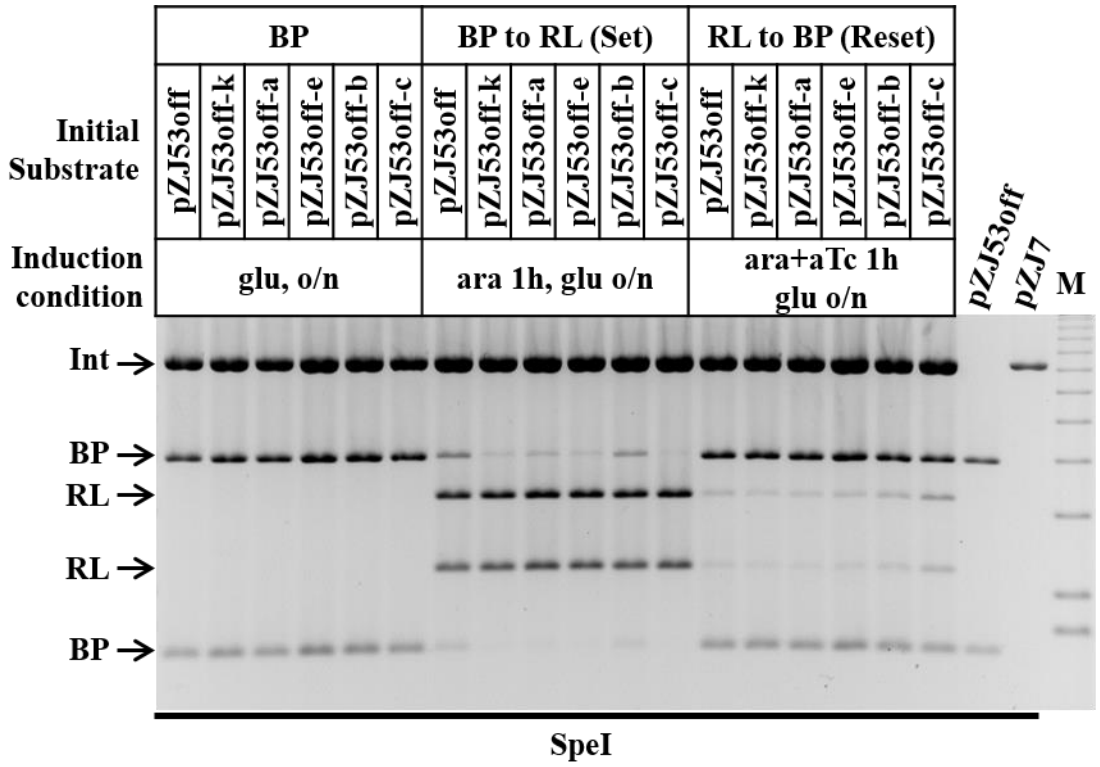


Figure 4. 12 Electrophoretic analysis of cells containing mutational state-based latch in a full cycle of set-reset reaction. Strain DS941/pZJ7 was transformed with substrate plasmid (pZJ53off, pZJ53off-k, pZJ53off-a, pZJ53off-e, pZJ53off-b, or pZJ53off-c). Each resultant strain was induced in a cycle of set (with one hour pulse of arabinose) and reset (with one hour pulse of arabinose and aTc) reactions. DNA extracted from cells after each induction was digested with SpeI and run on the agarose gel (pZJ7:7175 bp, BP:1459 bp and 3917 bp, RL:2198 bp and3178 bp)

Table 4. 2 Effect of Gp3 RBS on reaction efficiencies

Plasmid DNA	RBS sequence	Predicted Translation Initiation Rate (au)	Set	Reset
pZJ53off	AAGAAGGAGATATAAAT	19612	76%	89%
pZJ53off-k	AAGTAAGTGGTATAAAT	1106	95%	93%
pZJ53off-a	AAGAAAGTGGTATAAAT	644	94%	92%
pZJ53off-e	AAGGGAGTGGTATAAAT	3989	95%	92%
pZJ53off-b	AAGCGGGAGGTATAAAT	23586	84%	88%
pZJ53off-c	AAGGGGGTGGTATAAAT	34733	99%	76%

RBS with different sequence underlined. The predicted Translation Initiation Rate was calculated by an online RBS Calculator (Salis *et al.*, 2009). Recombination efficiencies were calculated through quantifying the bands on the agarose gel (Fig. 4.12, described as section 2.33), and substrates shown better performance were highlighted using grey shading.

Substrate plasmid pZJ53off-k, pZJ53off-a, and pZJ53off-e in cells were recombined efficiently in both set and reset reactions (Fig. 4.12). The set reaction on these mutants were better than the recombination on the substrate plasmid pZJ53off under pulsed induction, and reset reaction on these mutants maintained the same efficiency as the original plasmid. The *gp3* in these mutants had weaker predicted RBSs than that in pZJ53off, which supported the previous hypothesis that lowering the Gp3 expression level by using a weaker RBS can decrease the Gp3 leakage, improving the set reaction efficiency.

The control substrate mutant pZJ53off-b was recombined less efficiently in the set reaction probably because strong RBS of Gp3 led to more leakage of Gp3 to inhibit the set reaction. The control substrate mutant pZJ53off-c was recombined efficiently in the set reaction and less efficiently in the reset reaction probably because the Gp3 expression was not enough even with a strong predicted RBS (Table 4.2). This was the only sample that Gp3 expression was inconsistent with the RBS strength prediction.

As shown in Figure 4.11 and Figure 4.12, all of the cell samples gave consistent fluorescent and DNA state (higher fluorescent intensity corresponded to a larger proportion of substrate plasmid in ON state and *vice versa*). One of the most efficient substrate plasmid pZJ53off-k was used for the reliability test in the following section and its name was shortened to pZJ80off.

4.3.4 Reliability test of the optimised state-based latch under multicycle set-reset operation using two different input signals

To investigate the reliability of the final optimised state-based latch over multiple cycles of operation, strain DS941Z1/pZJ80off+pZJ7 was cultured and induced for five cycles using one hour pulsed induction (arabinose for set reaction and arabinose plus aTc for reset reaction). The DNA states were checked by agarose gel electrophoresis and the fluorescence of cells was checked using a Typhoon Scanner (Fig. 4.13).

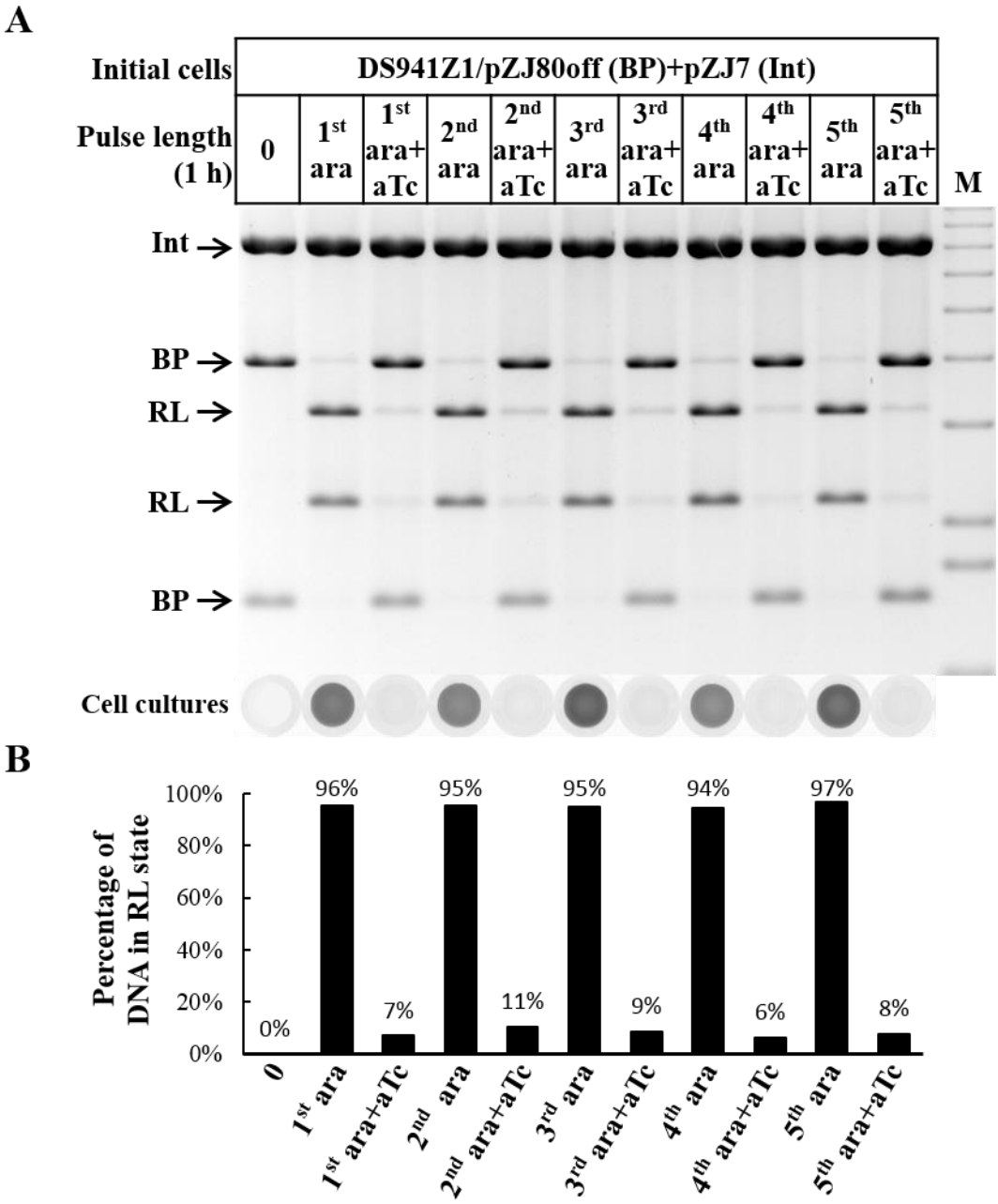


Figure 4. 13 Five cycles of set-reset operation with the optimised state-based latch. (A) Electrophoretic analysis of the recombination efficiencies starting from strain DS941Z1/pZJ80off (BP) + pZJ7 (Int). Five cycles of set (with one hour pulse of arabinose) and reset (one hour pulse of arabinose and aTc) reactions were carried out. Plasmid DNA extracted from cells after each induction was digested with SpeI and run on the agarose gel. Cell cultures were fixed, the fluorescence of the cell cultures was imaged using a Typhoon FluorImager, and shown below the gel. (B) The percentages of substrate plasmids in RL state were calculated though quantifying the bands on the agarose gel, and shown in the graph.

The substrate plasmid in the cells was set and reset efficiently multiple times with different pulsed signals (arabinose for set reaction and arabinose plus aTc for reset reaction), and the fluorescent intensities of cells changed between low and high accordingly (Fig. 4.13A). Alternate set and reset reactions remained highly efficient (~95% set and ~90% reset) for

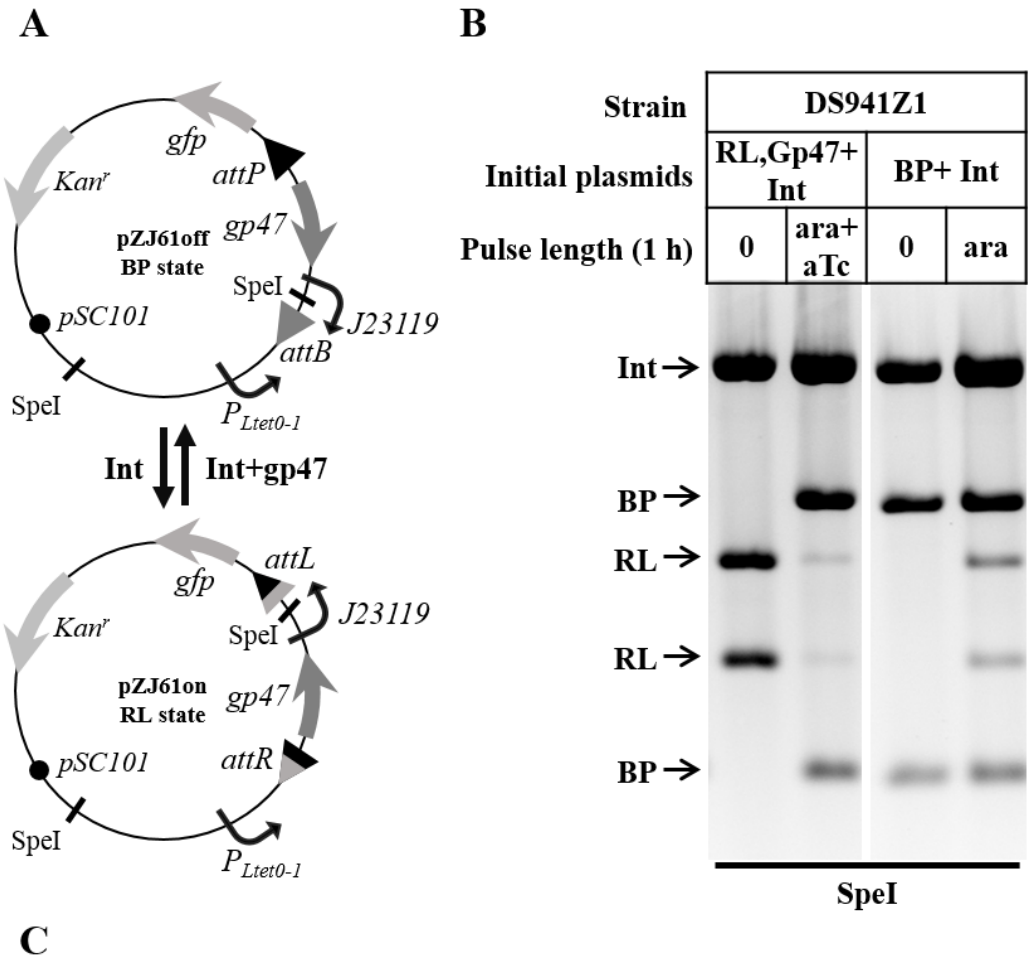
five full cycles of operation, demonstrating the practical reliability of the state-based latch in the multicycle operation.

4.3.5 Test of the state-based latch using Bxb1 integrase instead of ϕC31 integrase

To build a genetic system which can record a large number, multiple binary memory modules are needed. For example, a 4-bit genetic binary counter which can count up to 16, would consist of four modules each using a different integrase. Each module would have two states ("0" and "1"), and store a single binary digit. To use the state-based latch as the submodule of a binary counter or some other system, other integrases would have to behave as efficiently as ϕC31 integrase in the same architecture. Therefore, another well characterised integrase, Bxb1 integrase, with its corresponding RDF gp47, were tested in the state-based latch.

Bxb1 integrase gene was synthesised by total DNA synthesis by GeneArt (Life Technology) and inserted into the Int expression plasmid pZJ7 to replace the ϕC31 *integrase* gene, resulting in plasmid pZJ57. This plasmid has the following characteristics: medium copy-number plasmid with p15a origin of replication and chloramphenicol resistance gene, Bxb1 integrase gene with the same RBS sequence (predicted translation initiation rate: 81, Salis *et al*., 2009) as that for ϕC31 integrase gene (in pZJ7) under the control of the P_{BAD} promoter. For the substrate plasmid, a sequence with inverted repeat Bxb1 *attR* and *attL* flanking the RDF *gp47* gene (with same RBS as that for Gp3, predicted translation initiation rate:10779, Salis *et al*., 2009) and a constitutive promoter J23119 (from pGP47-INV-RL, section 2.27) was used to replace the equivalent invertible sequence in pZJ53on, resulting in plasmid pZJ61on (Fig. 4.14A). To test the activity of Bxb1 and RDF gp47, the reset (*attR*×*attL*) reaction was first carried out in strain DS941Z1 with substrate plasmid pZJ61on and Int expression plasmid pZJ57 using a one hour arabinose and aTc pulse. The DNA states were checked by agarose gel electrophoresis (left, in Fig. 4.14B).

To test the set (*attB*×*attP*) reaction, the substrate pZJ61off (Fig. 4.14A) was selected from the *in vivo* recombination product of pZJ61on as described in section 2.31. The set (*attB*×*attP*) reaction was carried out in strain DS941Z1 with substrate plasmid pZJ61off and Int expression plasmid pZJ57 using a one hour arabinose pulse. The DNA states were checked by agarose gel electrophoresis (right, in Fig. 4.14B).



C

Initial state		BP state		RL state	
Set input arabinose		0	0	0	1
Reset input arabinose+aTc		0	1	0	0
Output state	BP state	100%	10%	0%	20%
	RL state	0%	90%	100%	80%

Figure 4. 14 Reset or Set operation with the state-based latch based on Bxb1 integrase. (A) Expected change of substrate plasmids upone the inversion recombination reaction. Int alone is expected to recombine the substrate plasmid pZJ61off (BP) to pZJ61on (RL), and place the *gp47* gene in the same orientation as $P_{Ltet0-1}$ promoter and place J23119 promoter upstream of gfp. Int with Gp47 are expected to recombine the substrate plasmid pZJ61on (RL) to pZJ61off (BP) and restore the DNA state. Plasmid pZJ61off gives two fragments (size: 3949 bp and 1450 bp) and pZJ61on gives two fragments (size: 3172 bp and 2227 bp) after restriction digestion with SpeI (the drawing is not to scale). (B) Electrophoretic analysis of the reset and set reactions. Strain DS941Z1/pZJ61on (RL) + pZJ57 (Int) was induced with one hour pulse of arabinose and aTc for the reset reaction. DNA extracted from cells was digested with SpeI (pZJ57: 6994 bp) and run on the agarose gel (Left). Strain DS941Z1/pZJ61off (BP) + pZJ57 (Int) was induced with a one hour pulse of arabinose for the set reaction. DNA extracted from cells was digested with SpeI and run on the agarose gel (Right). (C) Percentages of DNA in BP and RL state before and after induction were calculated though quantifying the bands on the agarose gel and shown in the table. Percentages shown in the grey boxes are discussed.

The reset reaction mediated by BxbI integrase was very efficient (more than 90%, Fig. 4.14C), indicating that there were sufficient amount of BxbI integrase and RDF gp47 in cells for efficient *attR*×*attL* recombination. However, the set reaction was much less efficient (~

20%, Fig. 4.14C). These results indicated that the set reaction might be inhibited by the leakage of Gp47 in the absence of input signal aTc.

One possibility that caused the leaky Gp47 expression is that the RBS for Gp47 was too strong (predicted translation initiation rate: 10779). If this is correct, the same process used for optimising Gp3 expression level (section 4.3.3) could also be applied to optimise the expression level of Gp47 to decrease its leakage. Another possibility is that *attR* or *attP* upstream of *gp47* gene have some promoter activities in *E. coli*, leading to the leakage. If so, the location of the *gp47* gene, or the orientations recombination sites might need to be adjusted to remove the leakage caused by its nearby sequence.

4.4 Engineering of the binary counting module (single-signal controlled system)

4.4.1 Developing the state-based latch into a binary counting module to be controlled efficiently in strain DS941

Previous operations with the state-based latch (substrate showing in Fig. 4.15A) were carried out in strain DS941Z1, in which the inducible $P_{Ltet0-1}$ promoter is activated in the presence of aTc. Whereas, in strain DS941 which does not express the Tet repressor, the expression of Gp3 was expected to be governed only by the substrate sequence state. Therefore, when the substrate is in the RL state, Gp3 should be expressed independent of any input signal.

With Gp3 expression controlled only by the state of the device, starting from the BP state a single pulse of integrase would push the device to RL state through *attB*×*attP* recombination (switch-on) reaction, switching on Gp3 and GFP expression. Then, the next pulse of integrase together with Gp3 would push the device to BP state through *attR*×*attL* recombination (switch-off) reaction, switching off Gp3 and GFP expression.

However, since the expression of Gp3 was likely to be switched on as soon as the device was pushed to RL state, initial switching maybe inhibit further switch-on reaction and promoting the switch-off reaction. It was predicted that by using a short enough pulse of inducer to activate the expression of Int for a short time, Int might be degraded before the Gp3 built up to high enough level to reverse the recombination reaction. In this way, the RL state might be reached and maintained until the next pulse of induction. This was tested by inducing strain DS941 containing substrate plasmid pZJ53off and Int expression plasmid

pZJ7 with short pulses of arabinose (half hour or one hour). After induction, the DNA states were checked by agarose gel electrophoresis (Fig. 4.15B).

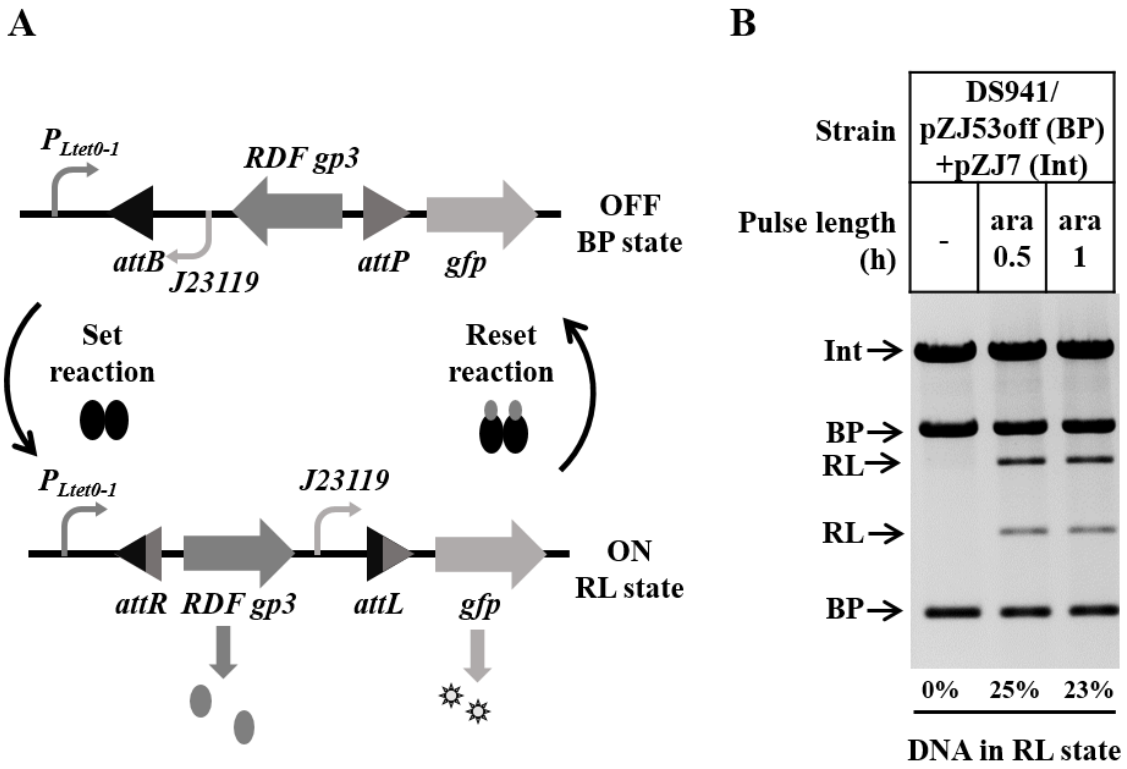


Figure 4. 15 The switch-on reaction of state-based latch in strain DS941. (A) The architecture showing the switching of the state-based latch in switch-on and switch-off reactions. (B) Electrophoretic analysis of the reaction. DS941/pZJ53off (BP) + pZJ7 (Int) was induced with half hour or one hour pulse of arabinose for the switch-on reaction. DNA extracted from cells after each induction was digested with SpeI (pZJ7:7175 bp, pZJ53off:1459 bp and 3917 bp, pZJ53on:2198 bp and 3178 bp) and run on the agarose gel. Percentages of DNA in RL were calculated after gel quantification.

Even with a short induction time (half hour), the switch-on reaction was very inefficient (about 20%) with the state-based latch in strain DS941 (Fig. 4.15B). It was hypothesized that the expression of Gp3 was very fast after the substrate DNA was recombined to RL (ON) state, which might inhibit the switch-on reaction and promote the switch-off reaction before the Int was sufficiently degraded or diluted with cell division.

To delay the expression of Gp3, a Tet repressor gene was cloned into the state-based latch substrate sequence, resulting in plasmid pZJ68off (substrate sequence showing in Fig. 4.16A). In this device, the expression of Tet repressor is on when the substrate is in BP state. The first pulse of integrase will push the device to RL state, switching on GFP expression and switching off Tet repressor expression. The residual Tet repressor is expected to repress the $P_{Ltet0-1}$ promoter, delaying the Gp3 expression during switching. As the Tet repressor degrades or is diluted by cell growth, Gp3 expression will be switched on, but by this point

Int will have also degraded so that the device does not reset. However, at the next pulse of integrase, there will be enough Gp3 to allow the switch-off reaction, resulting GFP expression off.

To check the set reaction on this new substrate sequence, strain DS941 containing substrate plasmid pZJ68off and Int expression plasmid pZJ7 was induced using short pulses of arabinose (Fig. 4.16B). Following the set reaction, the reset reaction was carried out using the same input pulse as used in the set reaction, to find out whether the substrate state can be restored or not in a successive operation (Fig. 4.16B).

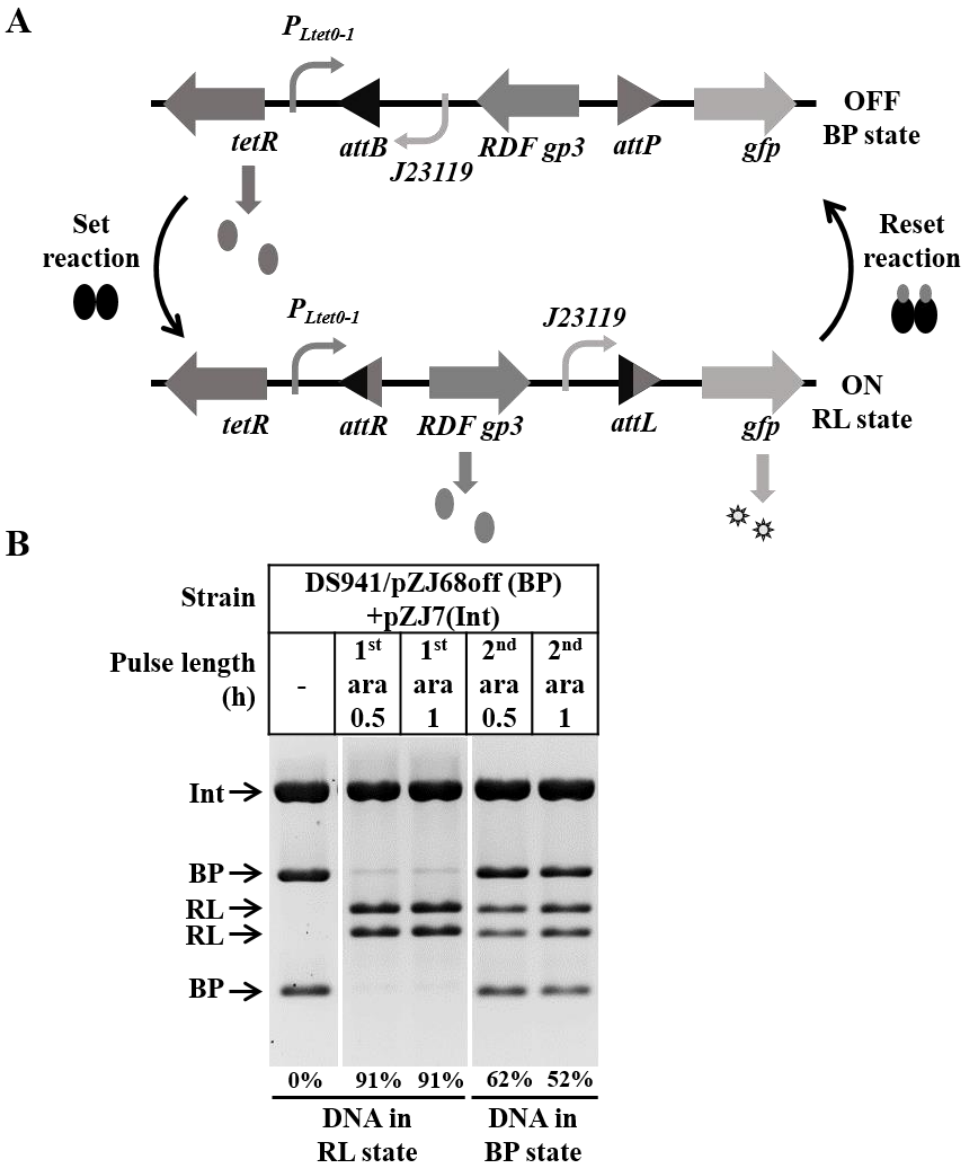


Figure 4. 16 The switch-on and switch-off reactions of the new device in strain DS941. (A) The architecture showing the switching of the new device in switch-on and switch-off reactions. (B) Electrophoretic analysis of the reactions. Strain DS941/pZJ68off (BP) +pZJ7 (Int) was induced in a cycle of switch-on (with half or one hour pulse of arabinose) and switch-off (with half or one hour pulse of arabinose) reactions. DNA extracted from cells after each induction was digested with SpeI (pZJ7:7175 bp, pZJ68off, BP: 2056 bp and 3917 bp, pZJ68on, RL:2795 bp and 3178 bp) and run on the agarose gel. Percentages of DNA in either RL or BP state were calculated after gel quantification.

Compared to the switch-on reaction efficiency with the state-based latch in DS941 (~20%, Fig. 4.15B), the efficiency was greatly improved with the new device (~90%, Fig. 4.16B) after either half or one hour arabinose induction. This result indicated that adding *tetR* into the substrate sequence increased the switch-on reaction efficiency. Meanwhile, the efficiencies greater than 60% and 50% were achieved for the switch-off reaction with the new device after the second half hour and one hour arabinose induction, respectively (Fig 4.16B).

The new device was demonstrated to change between two distinct states (high proportion of BP or RL substrate) by using a single repeating signal pulse, and it was named binary counting module. However, the best switch-off reaction (~60% efficiency with half hour induction) was not as efficient as the switch-on reaction (90% efficiency). The following sections present the experiments carried out to improve the efficiencies of the binary counting module through tuning the Gp3 expression level. It was expected that a higher Gp3 expression would be helpful for improving the switch-off reaction efficiency.

4.4.2 Optimisation of the binary counting module by changing the expression levels of RDF gp3

The efficiency of the state-based latch was improved by selecting Gp3 with optimized RBS sequence from a library (section 4.3.3). To find out if this method could also improve the efficiency of the binary counting module, a substrate library named pZJ68on-library (RL state) was constructed by cloning the *TetR* gene into the pZJ53on-library constructed in section 4.3.3. Cells containing these substrate plasmids would be changed from high fluorescent to low or non fluorescent if the substrate plasmids could be recombined efficiently from RL (GFP ON) state to BP state (GFP OFF). Based on this theory, strain DS941/pZJ7 containing substrate plasmids with RBSs library were induced with a half hour pulse of arabinose, followed by selection of cells that had a low fluorescence using the same process described previously (Fig. 4.10). As time was limited, only five random plasmids (named pZJ68off-a to pZJ68off-e) were selected from the library and their performances were compared.

One and a half cycles of switch-on switch-off reaction were carried out with strain DS941/pZJ7 containing these five selected plasmids using half hour arabinose pulses, the fluorescence of cell cultures was visualized and quantified by Typhoon Fluorimager (Fig. 4.17).

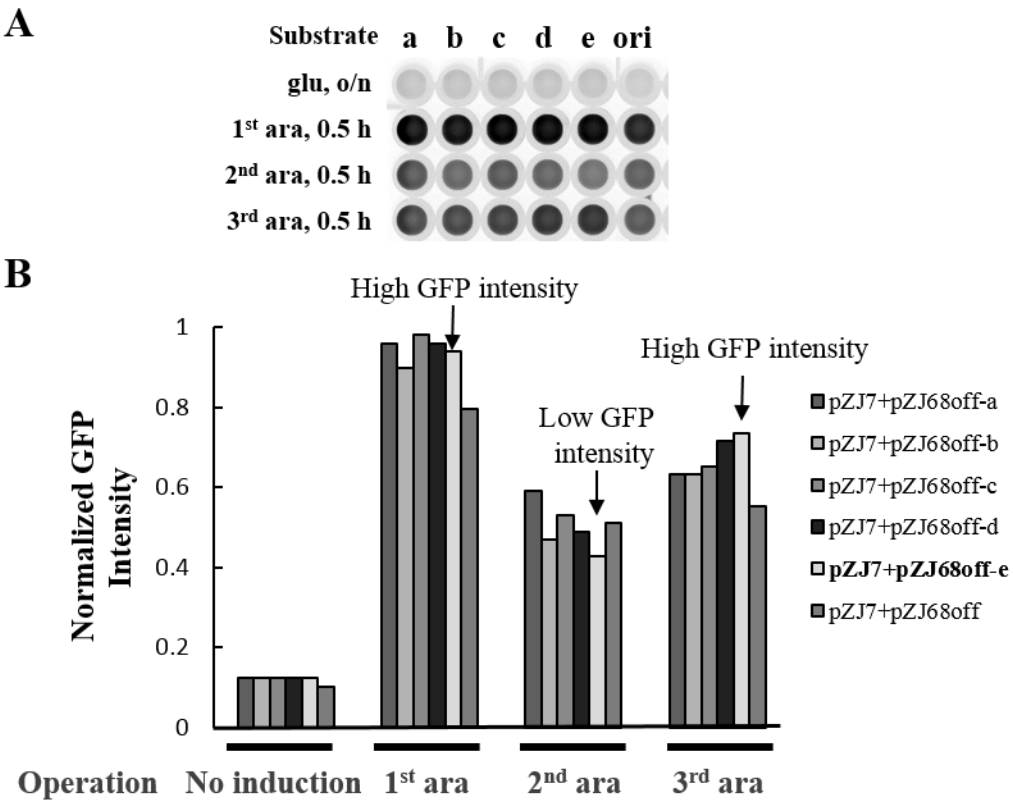


Figure 4. 17 Fluorescent analysis of cells containing substrate with different RBSs for Gp3 under switch-on and switch-off reactions. (A) Fluorescent scanning showing fluorescent strength of cell cultures. DS941/pZJ7 (Int) was transformed with substrate plasmid (a-e) selected from the *gp3* RBSs library or the original one (ori). One and half cycles of switch-on switch-off reactions with half hour pulse of arabinose for both reactions were carried out. Cells after each induction were fixed, and the fluorescence of the cell cultures were imaged using Typhoon FluorImager. (B) Quantification of the cell culture fluorescence. The GFP intensity of each culture was quantified by Quantity One software and the quantification numbers were normalized by setting the highest one to 1.

The fluorescent intensity of all cells was changed from low to high, then to low during the first cycle of switch-on switch-off reaction (Fig. 4.17). After the 3rd time of arabinose induction the fluorescent intensity of cells containing the original binary counting module was not changed, but cells containing substrate mutant pZJ68off-d or pZJ68off-e showed obvious change (Fig. 4.17).

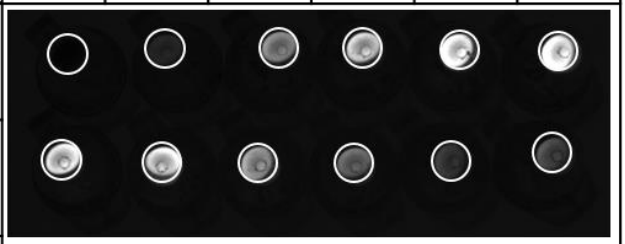
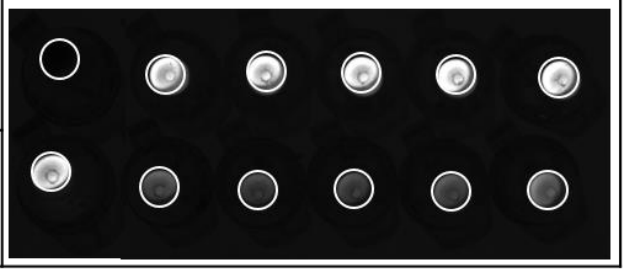
Substrate pZJ68off-e was chosen for the further test due to its lowest fluorescent intensity after the switch-off reaction (2nd arabinose pulse) and its highest fluorescent intensity after the second switch-on reaction (3rd arabinose pulse, Fig. 4.17). This substrate was named pZJ79off and sequenced to analyse how RBS sequence affected the recombination efficiency. Its RBS (sequence: AAGTGAGCGGTATAAAT) had a lower translation initiation rate (predicted translation initiation rate: 393, Salis *et al.*, 2009) than the one (sequence:

AAGAAGGAGATATAAAT) in the original substrate sequence (predicted translation initiation rate: 19612, Salis *et al*., 2009). This result might go against the previous expectation that the switch-off reaction efficiency could be improved by increasing the Gp3 expression level. But the one selected from the library had a predicted low Gp3 expression level because of its weaker RBS. Probably, lowering the Gp3 expression level decreased its leakage, leading to more efficient switch-on reaction. This would switch off TetR expression more completely, lowering its repression on Gp3 expression during the switching from RL to BP state. However, more data are needed to support this surmise since only a slight advantage of the cell fluorescence change was observed with the selected substrate (Fig. 4.17) and no more substrate was sequenced for analysing the relation between RBS strength and recombination efficiencies.

4.4.3 Characterisation of the optimised binary counting module to determine the optimal induction time

Previously, to obtain efficient protein expression, the pulsed inductions were started by adding inducer during exponential growth (after 1.5 hour preculture). It was predicted that extending the preculture length would lead to adequate Tet repressor accumulation when the substrate was in the BP state, resulting in more efficient switch-on reaction. Meanwhile, a longer preculture length would lead to adequate Gp3 accumulation when the substrate was in RL state, resulting in more efficient switch-off reaction. To test this hypothesis, a switch-on operation was carried out with strain DS941/ pZJ79off+pZJ7 by adding input signal after 1.5 or 2 hour preculture. Substrate plasmid pZJ79on was selected from the recombination product of pZJ79off *in vivo* (as described in section 2.31). A switch-off operation was carried out with strain DS941/pZJ79on+pZJ7. Cell cultures were spun down after each induction and checked by Bio-Rad blue ligth Transilluminator (Fig. 4.18A), the fluorescent strengths quantified using Quantity One software were shown in the diagram (Fig. 4.18B).

A

Preculture length (hour)	Pulse length (min) / Initial Strain	0	10	15	20	30	60
1.5 OD_{600}≈0.35	DS941 (pZJ7+pZJ79off)						
	DS941 (pZJ7+pZJ79on)						
2 OD_{600}≈0.7	DS941 (pZJ7+pZJ 79off)						
	DS941 (pZJ7+pZJ79on)						

B

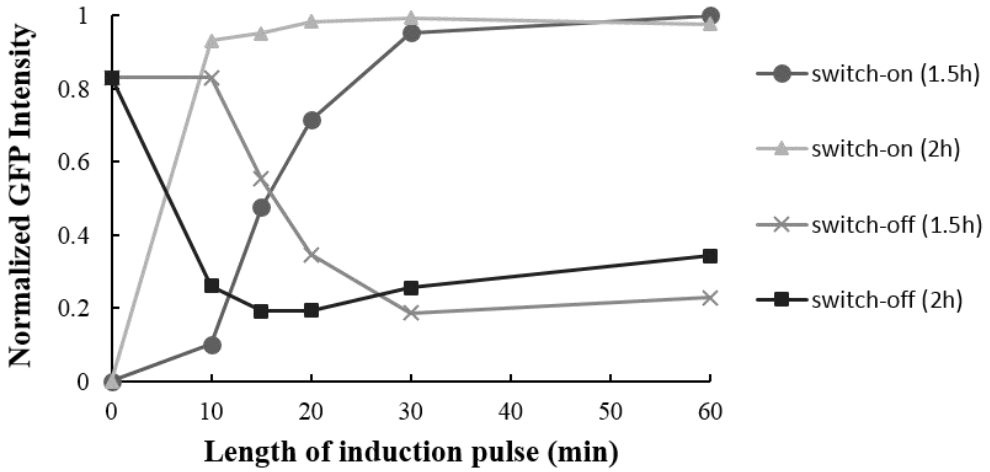


Figure 4. 18 Pulsed induction time course with the binary counting module by adding input signal after different preculture length. (A) Fluorescent analysis of cell cultures after induction. DS941/pZJ79off (BP) + pZJ7 (Int) and DS941/pZJ79on (RL) + pZJ7 (Int) were induced with pulsed induction. Each resultant colony was inoculated in L-Broth with glucose overnight, and the overnight culture was diluted 40-fold in 5 ml L-Broth. After 1.5 or 2 hour preculture, arabinose was added for different lengths of time followed by dilution and growth in L-Broth with glucose overnight. Each overnight culture was centrifuged in a microcentrifuge tube and the cell pellets labelled by white circles were checked using Bio-Rad blue light Transilluminator. (B) Quantification of the fluorescence strengths of cell pellets. The GFP intensity of each cell pellet with different induction time was quantified using Quantity One software, the quantification numbers were normalized by setting the highest one to 1.

Fluorescent signal change of cell pellets suggested that two hour preculture length leads to fast response to input signal for both directional reactions (Fig. 4.18). A shorter length of pulse was enough to change cells to a maximum opposite state (from low GFP intensity to high GFP intensity or *vice versa*) when added after a longer preculture length (two hour, Fig. 4.18B). However, the preculture length did not influence the maximum efficiency for both switch-on and switch-off reaction since cells reached to similar steady fluorescent state (Fig. 4.18B).

As cells gave a fast response to signal after two hour preculture, the time course with two hour preculture length was continued and a full cycle of switch-on switch-off operation was carried out to confirm the optimal pulse length for induction. DNA was extracted from cells that were induced with different lengths of pulse, checked by restriction digestion, and agarose gel electrophoresis (Fig. 4.19).

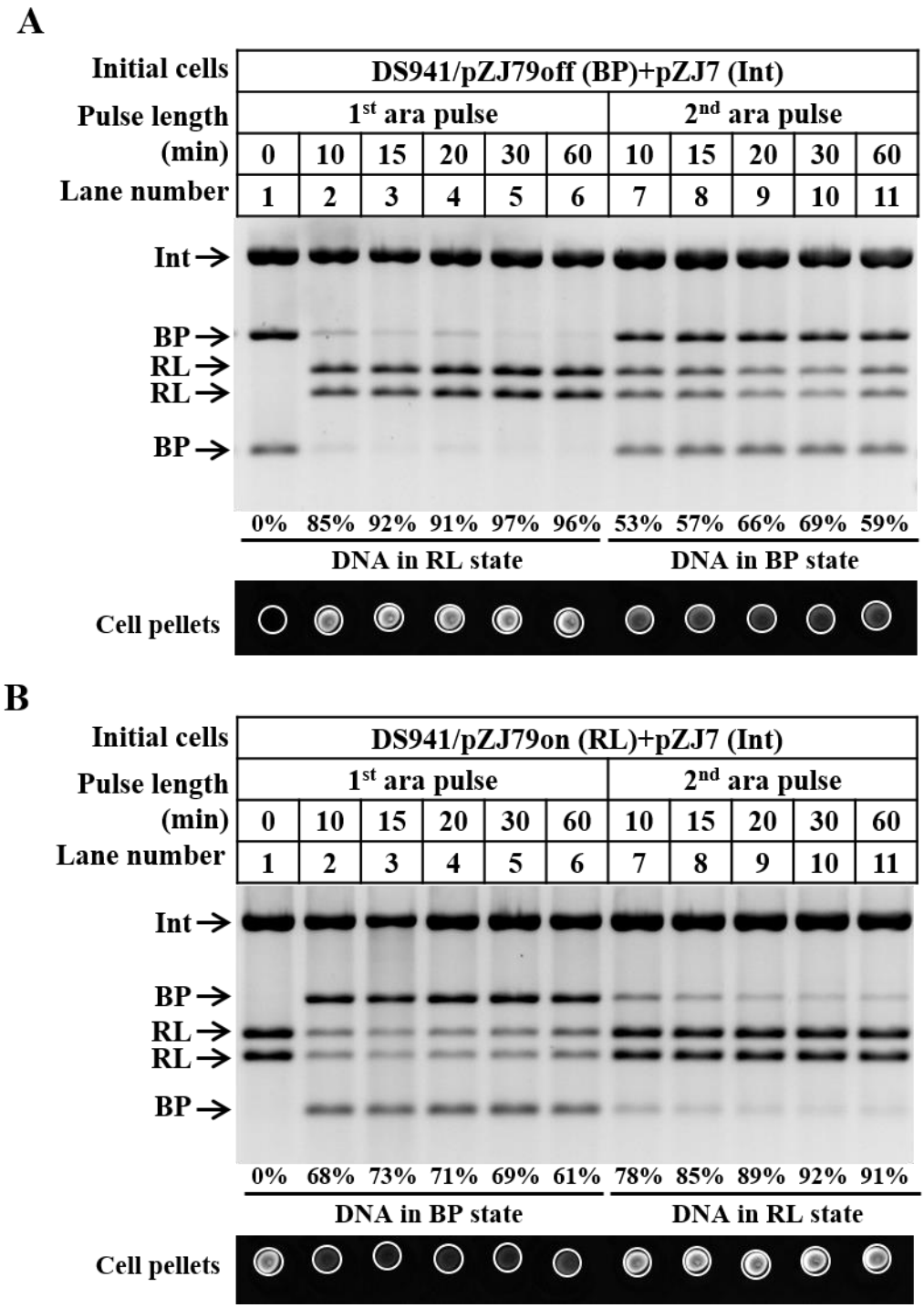


Figure 4. 19 Pulsed induction time course of the binary counting module. (A) Electrophoretic analysis of time course reaction starting with BP-state substrate. DS941/pZJ79off (BP) + pZJ7 (Int) was induced in a cycle of switch-off switch-on reaction with different lengths of arabinose pulse. DNA extracted from cells after each induction was digested with SpeI (pZJ79off, BP: 2056 bp and 3917 bp, pZJ79on, RL:2795 bp and 3178 bp, pZJ7: 7175 bp) and run on the agarose gel. (B) Electrophoretic analysis of time course reaction starting with RL-state substrate. DS941/pZJ79on (RL) + pZJ7 (Int) was induced in a cycle of switch-on switch-off reaction with different lengths of arabinose pulse. DNA extracted from cells after each induction was digested with SpeI and run on an agarose gel. Percentages of DNA in either RL or BP state were calculated through gel quantification, and shown below each gel. Cells after each induction were centrifuged and the fluorescence of cell pellets in the microcentrifuge tubes was checked using Bio-Rad blue light Transilluminator and shown at the bottom of figure A and B.

In vivo time courses starting from either DS941/pZJ79off+pZJ7 (containing substrate in BP state) or DS941/pZJ79on+pZJ7 (containing substrate in RL state) showed that a pulse longer than 15 min was good for highly efficient switch-on reaction (> 90% efficiency, lanes 3-6 in Fig. 4.19A, and lanes 8-11 in Fig. 4.19B), but a specific range of pulse lengths (15 min to 30 min) was required for efficient switch-off reaction (~70% efficiency, lanes 8, 9, and 10 in Fig. 4.19A, and lanes 3, 4, and 5 in Fig 4.19B). The fluorescence of cell culture changed consistently with the DNA state.

For fast and efficient induction, the two hour preculture length and the shortest workable induction time 15 min (92% switch-on reaction efficiency, lane 3 in Fig. 4.19A, and 73% switch-off reaction efficiency, lane 3 in Fig. 4.19B) were used in the following operation on the binary counting module.

4.4.4 Reliability test of the binary counting module under multicycle switch-on switch-off operation using a single input signal

To study the reliability of the binary counting module in a multicycle switch-on switch-off operation, strains DS941/pZJ79off+pZJ7 and DS941/pZJ79on+pZJ7 were cultured and induced for four cycles using a 15 min induction pulse of arabinose. The DNA extracted from cells was checked by restriction digestion and agarose gel electrophoresis (Fig. 4.20)

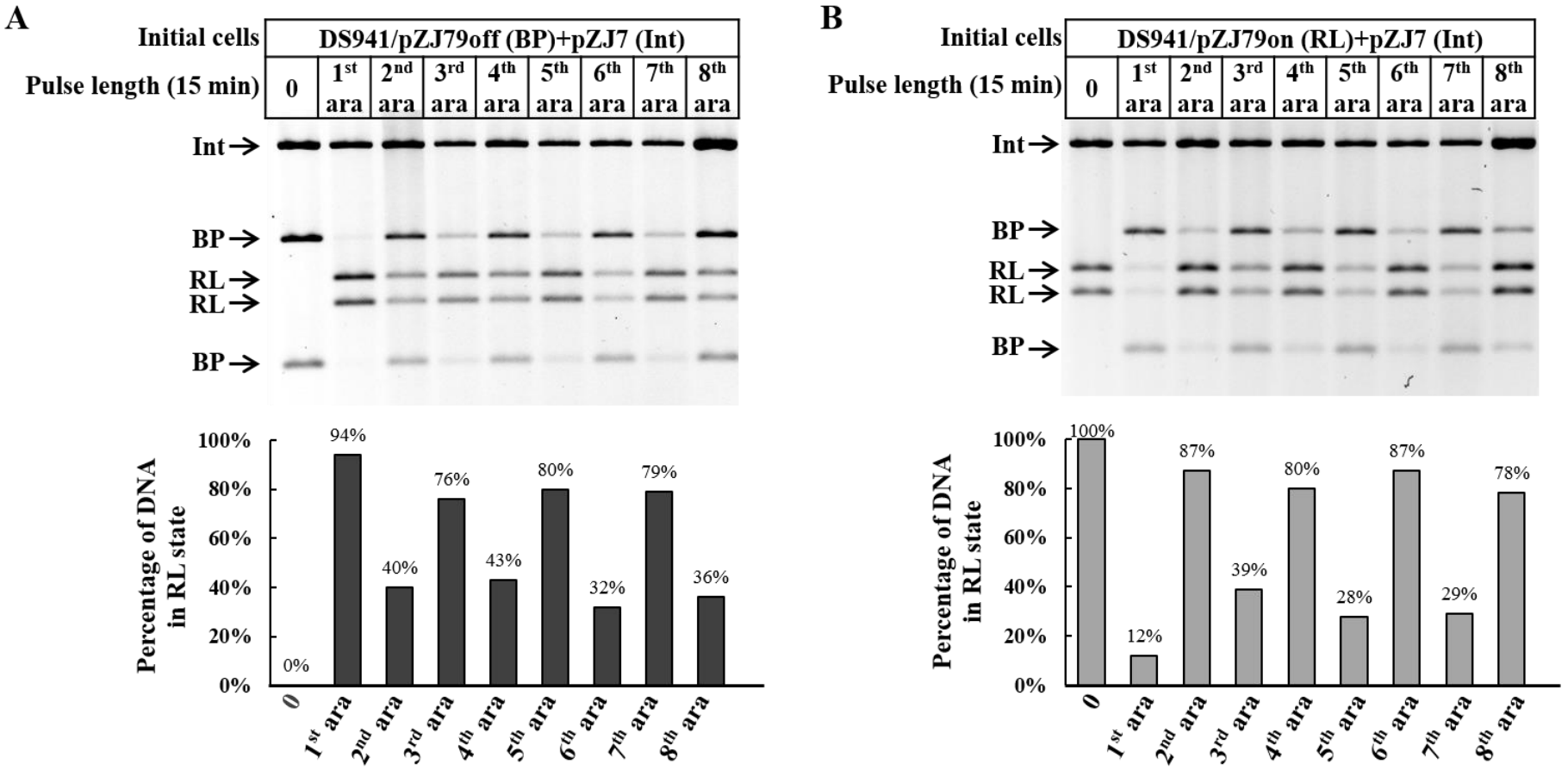


Figure 4. 20 Multicycle operations with the binary counting module. (A) Electrophoretic analysis of the recombination efficiencies starting from BP-state binary counting module. DS941/pZJ79off (BP) +pZJ7 (Int) was induced with 15 min pulse of arabinose for eight times. Plasmid DNA extracted from cells after each induction was digested with SpeI and run on the agarose gel. The percentages of substrate plasmids in RL state were calculated though quantifying the bands on the agarose gel, and shown in the graph below the gel. (B) Electrophoretic analysis of the recombination efficiencies starting from RL-state binary counting module. DS941/pZJ79on (RL) + pZJ7 (Int) was induced with 15 min pulse of arabinose for eight times. Plasmid DNA extracted from cells after each induction was digested with SpeI and run on the agarose gel. The percentages of substrate plasmids in RL state were calculated though quantifying the bands on the agarose gel, and shown in the graph below the gel.

The state of the substrate plasmid was changed multiple times by the induction with an arabinose pulse. In the absence of any input signal, the counting module was maintained in its initial state, either fully OFF (BP) or ON (RL). With the first pulse of arabinose, the counting module was efficiently changed from BP to RL (~95% efficiency, Fig. 4.20A) through the switch-on reaction, or from RL to BP (~90% efficiency, Fig. 4.20B) through the switch-off reaction. The efficiency for subsequent switch-on and switch-off reactions went down to about 80% and 60%-70% respectively, and the alternation between 80% switch-on reaction and 60%-70% switch-off reaction could be maintained for three cycles. These results demonstrated the reliability of the binary counting module in a multicycle reaction.

To observe the fluorescence of cell cultures after each reaction, cell samples collected from the experiment described above were fixed and checked by Typhoon Scanner (liquid in the wells, Fig. 4.21A and B). To further investigate the fluorescence of individual cells, the fixed samples were checked by CyAn ADP Analyzer (as described in section 2.29.3). The fluorescent data for 30, 000 cells of each sample was collected and analysed (flow cytometry images, Fig. 4.21A and B).

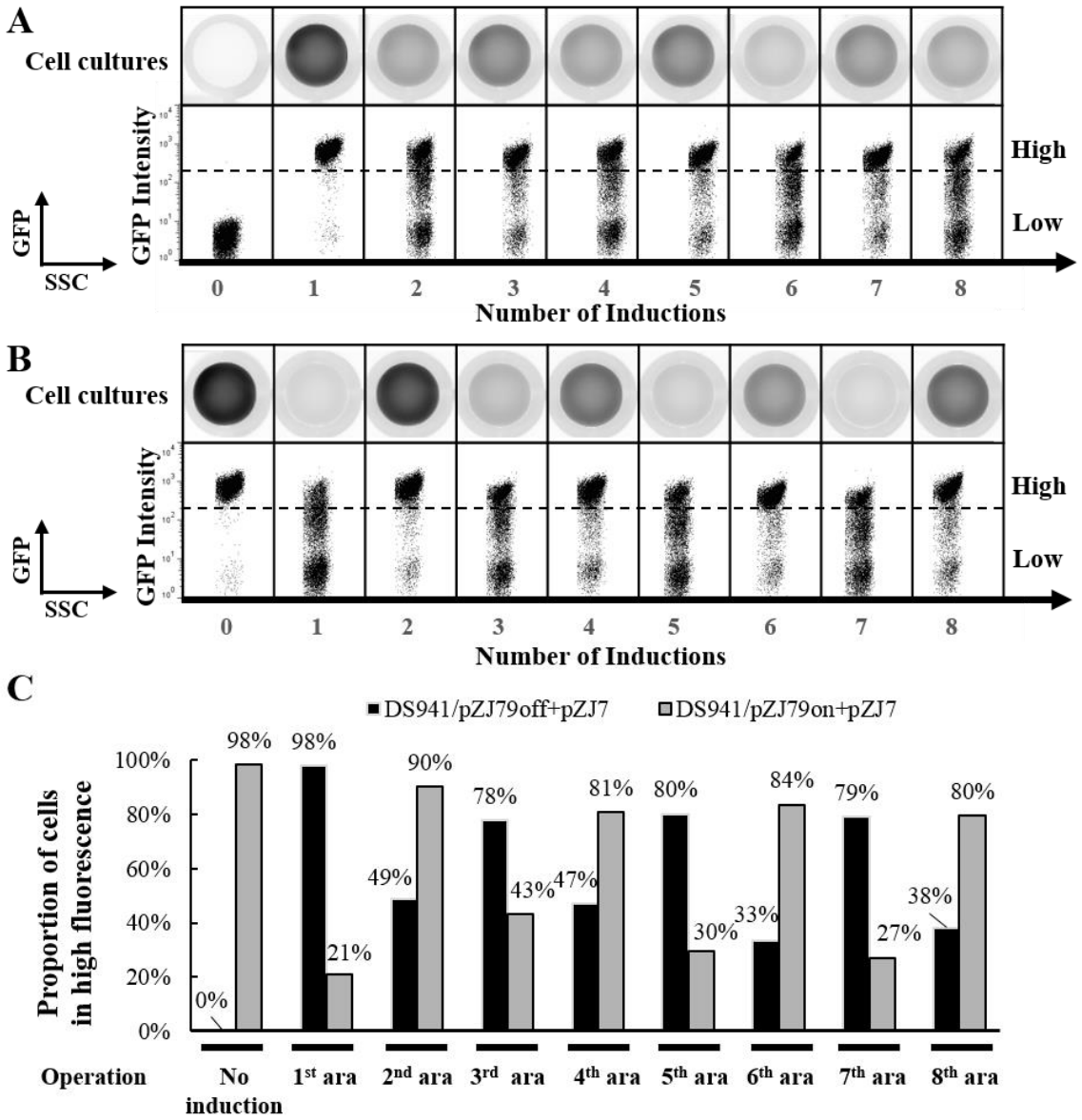


Figure 4. 21 Fluorescent analysis of cell cultures and individual cells containing the binary counting module in multicycle reactions. (A) DS941/pZJ79off (BP) + pZJ7 (Int) was induced eight times with 15 min pulses of arabinose (Same process as Fig. 4.20). Fluorescence of cell cultures was recorded using Typhoon FlourImager (top), and fluorescence of individual cells was analysed using Flowjo program after gating cells by forward and side scatter (bottom, cells in high and low fluorescent state are divided by dashed line). (B) DS941/pZJ79on (RL) + pZJ7 (Int) was induced eight times with 15 min pulses of arabinose (Same process as Fig. 4.20). Fluorescence of cell cultures was recorded using Typhoon FlourImager (top), and fluorescence of individual cells was analysed using Flowjo program after gating cells by forward and side scatter (bottom, cells in high and low fluorescent state are divided by dashed line). (C) The percentages of cells in high fluorescent state detected by Flowjo program are shown in the diagram.

The fluorescent states of cells changed with each arabinose pulse, alternating between predominantly fluorescent and predominantly non-fluorescent each time. In the absence of input signal, strain DS941/pZJ79off+pZJ7 retained non-fluorescent and strain DS941/pZJ79on+pZJ7 retained high GFP fluorescence. With the first pulse of arabinose, nearly all of the non-fluorescent cells were changed to express high-level GFP (98% cells in

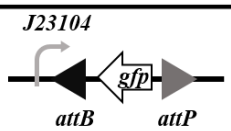
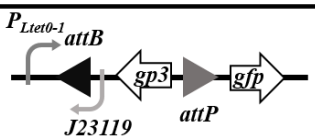
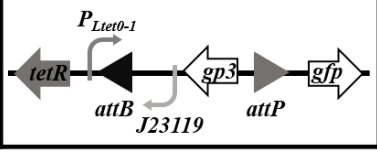
high fluorescent state, Fig. 4.21A) through switch-on reaction, and most of the highly fluorescent cells were changed to express low-level GFP (79% cells in low fluorescent state, Fig. 4.21B) through switch-off reaction. The efficiency for the subsequent switch-on and switch-off reactions went down to about 80% and 50%-70% respectively (based on the fluorescence change of cell population, Fig. 4.21A and B), and the alternation between switch-on reaction and switch-off reaction carried on for three cycles at the same efficiencies. These results agreed with the recombination efficiencies based on the overall DNA state (80% switch-on efficiency and 60%-70% switch-off efficiency, Fig. 4.20).

It was noticed that cells containing the binary counting module produced a mixed fluorescent state population (Fig. 4.21A and B). This was more obvious after the switch-off reaction: some cells switched to complete off of GFP expression; some cells entered an intermediated level of GFP expression; while others retained high GFP expression. The inconsistency of the cell fluorescence might be because the substrate DNA sequence was in a multi-copy plasmid, and that the proportion of plasmid in BP or RL state in each individual cells influenced their fluorescent intensity of each individual cell.

4.5 Conclusion and discussion

In this chapter, three main plasmid-borne memory devices were constructed, which are 1) set-reset latch, 2) state-based latch, and 3) binary counting module. The set (or switch-on) and reset (or switch-off) reactions with these devices were based on an inversion recombination mediated by serine integrase. These reactions, which can change the device between two states (GFP ON and GFP OFF), were triggered using either the same or different input pulses. The induction condition and the range of recombination efficiencies (through checking the overall DNA state) of each device are described (Table 4.3), and the advantages and disadvantages of each device are discussed in this section.

Table 4. 3 Induction times and recombination efficiencies for different memory devices

Device	Architecture of the substrate in BP state	A cycle of set-reset reaction					
		Set reaction (BP to RL)			Reset reaction (RL to BP)		
		Input signal	Length of induction pulse	Efficiency	Input signal	Length of induction pulse	Efficiency
Set-reset latch		**ara**	**60 min**	**~75%**	**aTc**	**60 min**	**~70%**
State-based latch		**ara**	**60 min**	**~95%**	**ara +aTc**	**60 min**	**~90%**
Binary counting module		**ara**	**15 min**	**80-95%**	**ara**	**15 min**	**60-90%**

ara, arabinose; aTc, anhydrotetracycline.

4.5.1 Set-reset latch

An efficient set-reset latch was described at the beginning of this chapter. This device has a similar architecture to the rewritable recombinase addressable data module described by Bonnet *et al*. (2012), but using ϕC31 integrase to record information into plasmid DNA rather than using Bxb1 integrase to record information into chromosome. Both the set (*attB*×*attP*) and reset (*attR*×*attL*) operations led to distinct DNA state and fluorescent signal change, which proved that ϕC31 integrase is an efficient tool to write information into DNA. The advantage of this latch system is that the components (set control, reset control, and data storage) are decoupled from each other, enabling flexible change or selection of any of these components to fulfill different application requirements.

There should be space to further improve the recombination efficiencies (now is 75% for set reaction and 70% for reset reaction) of the set-reset latch. Experiments described in chapter 3 proved that the recombination efficiencies can be improved through adjusting the integrase and RDF expression levels, which was not tested with the set-reset latch. It was possible that placing the RDF gene downstream of the *int* gene might lead to some leaky expression of RDF from the reset generator, inhibiting the set reaction. The result in Figure 4.2 showed that even just activating Int expression, the *attR*×*attL* recombination was observed, which supported the assumption of Gp3 leakage. Additionally, the *gp3* gene was far away from the

promoter, which might result in insufficient Gp3 expression for highly efficient *attR*×*attL* recombination. To overcome these deficiencies, future work could be done by changing the position of ϕC31 *int* and *gp3* genes on the reset operator. It might also be necessary to tune the expression levels and the degradation rates of Int and RDF for more efficient reactions.

4.5.2 State-based latch

Next, the state-based latch was constructed. It was recombined efficiently to switch on and off the expression of GFP using different signals. The performance of this device was improved in three ways: 1) exchanging the positions of genetic parts, 2) modifying specific DNA sequence, and 3) tuning the strength of expression regulatory part, RBS.

Adjusting the position of the genetic parts (genes, recombination recognition sites, promoters, and terminators) can minimize the unwanted interacting between each part, and help give the desired protein expression. For example, the behaviour of reporter protein expression was improved by changing *gfp*'s position and placing a constitutive promoter in the invertible sequence to control GFP expression (section 4.3.1). The position of genetic parts in circuits is crucial for a device's behaviour, especially when a large number of genetic parts are involved. This should be considered during the design process and sometimes need to be adjusted according to the experimental results.

Unexpected promoters were one of the main factors that interfered with the performance of the genetic circuits made in this study. The *attB*×*attP* recombination was thought to be inhibited by the leaky Gp3 expression caused by the unexpected promoters upstream of *gp3* (section 4.3.1). These promoters were removed from the sequence by introducing silent mutation. Sequence modification is useful when part order should be fixed and adding a terminator between two parts is not applicable if they are sharing the same promoter.

Tuning the strength of the RBS sequence varied the expression levels of the RDF gp3 and allowed selection of improved behaviour (section 4.3.3). A RBSs library was constructed with expected range of strength according to the calculation from an online program (Salis *et al*., 2009). Using a functional screening process, one of the device with new RBS for Gp3 was proved to be more efficient than the original one (section 4.3.3), which indicated the practicability of the RBS screening process.

Using these methods summarized above, the efficiency of the state-based latch was increased to greater than 95% for set reaction and greater than 90% for reset reaction in a successive

set-reset reaction. This latch was also demonstrated to be highly reliable during a multicycle set-reset reaction according to the quantification result of the substrate DNA state (section 4.3.4). These results suggest that ϕC31 integrase behaves as well as Bxb1 used by Bonnet *et al*. (2012) in recording information into DNA sequence. The state-based latch is an important step towards building a system controlled by a single signal.

Integrase Bxb1 was also used to test the usability in the state-based latch (section 4.3.5). After induction, the reset reaction was efficient, whereas, the set reaction was inefficient. It was inferred that the reason for the inefficient set reaction was the leakage of RDF gp47 due to its strong RBS or the promoter activity in nearby sequence, such as the Bxb1 *attP* and *attR* sites. In order to improve the efficiencies of the Bxb1 integrase based device, future work need to be done to reduce the Gp47 leakage through tuning the RBS strength or changing the position of *gp47*.

4.5.3 Binary counting module

Though the state-based latch was recombined efficiently under the operation of two different signals (set reaction: arabinose; reset reaction: arabinose and aTc) in strain DS941Z1, the device did not perform as anticipated, switching between BP and RL state under the control of a single signal in DS941 (section 4.4.1). The switch-on reaction was very inefficient in DS941 even with a very short pulse of arabinose. It was hypothesized that the inefficient switch-on reaction was caused by the expression of Gp3 as soon as the device started switching, inhibiting further switch-on reaction. To extend the interval between degradation of Int and expression of Gp3, a delay for Gp3 expression was added by inserting the *tetR* gene under the control of sequence state, resulting in the binary counting module. The expected performance of the binary counting module is described as below (Fig. 4.22).

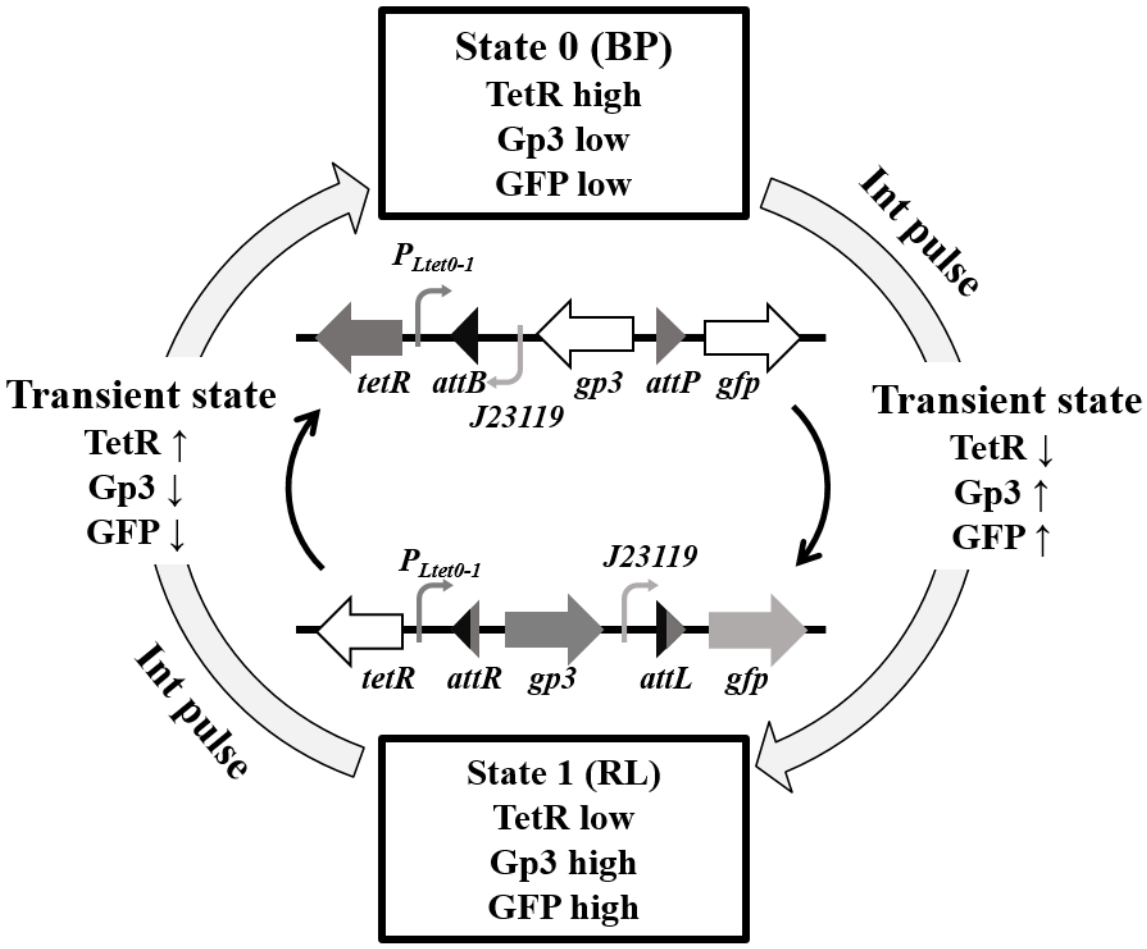


Figure 4. 22 Expected performance of the binary counting module with single induction pulse. This device can be switched between state 0 (BP) and state 1 (RL) by the Int pulse induced using a single input signal. The change of protein expression level is represented using symbols (↑, increase; ↓,decrease).

In this module (with *tetR*), the expression of Gp3 was delayed and the switch-on reaction efficiency (with 30 min arabinose pulse) was improved to about 90% compared to the device without *tetR* (section 4.4.1).

It was proved that using a slightly longer preculture length (two hour) could promote quicker recombination reactions. The explanation might be that the induction of Int was more efficient later in the exponential growth phase. It was also thought that there would be more Tet repressor accumulated in cells at this time point, leading to tighter repression of Gp3 and improved switch-on reaction. Furthermore, a longer preculture length might also beneficial to accumulate Gp3 when the substrate was in RL state, improving switch-off reaction efficiency.

Such binary counting module proved the ability of changing state in response to a single repeating input signal. After encountering an odd number of input signal, the device switched

to an opposite state to the initial one. Whereas, after encountering an even number of input signal, device switched to same state as initial. By tuning the RBS strength of *gp3* and using optimal induction conditions (two hour preculture length and 15 min length of input pulse), the efficiencies of the binary counting module maintained a range of 80%-95% for switch-on reaction and 60%-90% for switch-off reaction during a muticycle operation (Fig. 4.20). There was no obvious degradation in efficiency after multiple induction operation, but the fluctuations of efficiencies were obvious. These fluctuations might be caused by the stochastic factors during the culture and induction (such as temperature, inducer, and time). More repeated experiments are needed to confirm these performances and to find out an optimal operation condition.

It was noticed that starting the recombination from the pure substrate plasmid (BP state or RL state) showed better recombination efficiency than starting from a mixture with the same induction pulse (e.g. 94% of lane 2 in Fig. 20A compared to 87% of lane 3 in Fig. 4.20B; 88% of lane 2 in Fig.4.20B compared to 60% of lane 3 in Fig. 4.20A). The explanation might be that the module state had great influence on the recombination efficiency since the expressions of both Gp3 and TetR were governed by the state of the substrate sequence. Starting from pure BP substrate (TetR expression on, Gp3 expression off), all of the cells expressed sufficient amount of Tet repressor to delay Gp3 expression, resulting in more efficient switch-on reaction; starting from pure RL substrate (TetR expression off, Gp3 expression on), none of the cells expressed Tet repressor to repress Gp3 expression, resulting in more efficient switch-off reaction.

A limitation with the binary counting module is that the reaction efficiencies are influenced by the length of the induction pulse, and only a small range of pulse lengths can fulfill the requirement of high recombination efficiency. The previously constructed recombinase-based counter (Friedland *et al*., 2009) and push-on push-off switch (Lou *et al*., 2010) lost function with a long time input signal. To avoid this problem, Subsoontorn and Endy (2012) proposed the design principle of an activation-inhibition toggle flip-flop by connecting two independent recombinase based set-reset latches to store state during and between input pulses (Subsoontorn and Endy, 2012). However, this system is complex and two integrases are involved in recording one bit information which limits the information storage capacity. Thus, the future direction of developing the binary counting module constructed in this study is to extend its functional pulse width.

Though the plasmid-borne memory devices gave promising recombination efficiency when viewing the overall DNA states, the individual cell states were not consistent because mixed populations of multi-copy substrate plasmids were generated in each cell. To solve this problem, the invertible substrate sequence could be integrated into the genome, which would decrease the copy number to one. This change was expected to divide cells into either a completely OFF or completely ON state, without any other intermediate states. The next chapter will discuss how to transfer the substrate sequence into the chromosome of *E. coli* strain and the performances of different chromosomal memory devices.

5 The design and engineering of chromosomal memory devices

5.1 Introduction

In chapter 4, three different plasmid-borne memory devices, the set-reset latch (section 4.2), the state-based latch (section 4.3), and the binary counting module (section 4.4), were constructed. All three devices showed efficient recombination performances as measured by the overall substrate DNA state (Fig. 4.3, Fig. 4.13, and Fig. 4.20). The fluorescent performance of individual cells which contained the binary counting module was checked by flow cytometry. These cells showed variant fluorescent intensities after recombination, especially after the *attR*×*attL* recombination (Fig. 4.21). The mixed cell populations were thought to be generated by recombination on a multi-copy substrate plasmid, so that each individual cell can contain a mixture of plasmids in different states (Brophy and Voigt, 2014; Moon *et al*., 2011). In order to mitigate the inconsistency of cell states and increase the stability of the genetic device, the target sequence was integrated into the chromosome of *E.coli* strain DS941 or DS941Z1 as a single copy. It was expected that this change would also improve the recombination efficiency, since the initial data collected in chapter 3 showed more efficient recombination on a low copy-number substrate plasmid than that on a high copy one (section 3.3.2 and section 3.3.3).

In the first part of this chapter, the method used to construct the chromosomal memory devices is introduced. To integrate a single copy substrate sequence into the chromosome, the invertible substrate sequence (from the state-based latch or the binary counting module) was placed on a suicide ISY100 transposon delivery plasmid and mobilized by RP4 conjugative machinery (Demarre *et al*., 2005; Ferrières *et al*., 2010; Urasaki *et al*., 2002).

Later in this chapter, methods used to check the recombination efficiencies of the chromosomal devices are discussed. These methods include fluorescence microscopy, flow cytometry, and colony PCR. Flow cytometry was found to be an efficient method and was therefore used to check the behaviours of the chromosomal devices in all the subsequent experiments. Based on this initial characterisation, tests of the chromosomal devices are presented and ways to improve their efficiencies are discussed.

5.2 Chromosomal delivery of memory devices using ISY100 transposition

ISY100 is a member of IS630/Tc1/mariner superfamily, and is demonstrated to transpose in *E.coli* at a very high frequency (Urasaki *et al*., 2002). What is more, *in vitro* studies have demonstrated that IS630/Tc1/mariner members do not need host factors to perform transposition (Feng and Colloms, 2007; Lampe *et al*., 1996). Any piece of DNA flanked by two ISY100 inverted repeat sequence (IRs) can be mobilised on expression of the transposase. Therefore, ISY100 transposon was used to integrate the substrate sequence into the chromosome.

The invertible substrate sequences of the state-based latch and the binary counting module were integrated into the chromosome of *E. coli* using suicide ISY100 transposon delivery plasmids mobilised by broad-host-range RP4 conjugation machinery (Fig. 5.1). The suicide plasmid pSW23T is based on the oriVR6Kγ origin of replication which is dependent on the *pir* encoded protein for plasmid replication, and contains the oriT transfer origin of the broad host-range conjugative plasmid RP4 (Demarre *et al*., 2005). The invertible substrate sequence between ISY100 IRs was cloned into this plasmid along with a kanamycin resistance gene, and the manipulation was carried out in a pir^+ strain π3 (Demarre *et al*., 2005). The pSW23T based plasmid will be unable to replicate in pir^- stain, therefore selection for a marker (antibiotic resistance gene) adjacent to the invertible segment between ISY100 IRs will select for chromosomal integration of the sequence.

It was reported that delivery of this conditionally replicating plasmid (e.g. pSW23T) through transformation was highly inefficient (Ferrières *et al*., 2010), and this was confirmed in this study too. So, plasmid pSW23T containing the oriT transfer origin was mobilized through RP4 conjugation (Thomas and Smith, 1987). Strain MFDpir (Ferrières *et al*., 2010) was used as the donor strain for the RP4 conjugation. This strain contains the *pir* gene for the replication of plasmid pSW23T (oriVR6Kγ) and functional region for RP4 conjugation on the chromosome. This strain is also auxotrophic for diaminopimelic acid (DAP), allowing counterselection on rich complex media lacking DAP after conjugation (Ferrières *et al*., 2010). When the suicide delivery plasmid was moved into the target strain DS941 (pir^-) or DS941Z1 (pir^-) through conjugation, the successful ex-conjugants containing the invertible substrate sequence on the chromosome would be selected efficiently using medium containing the appropriate antibiotic and lacking DAP.

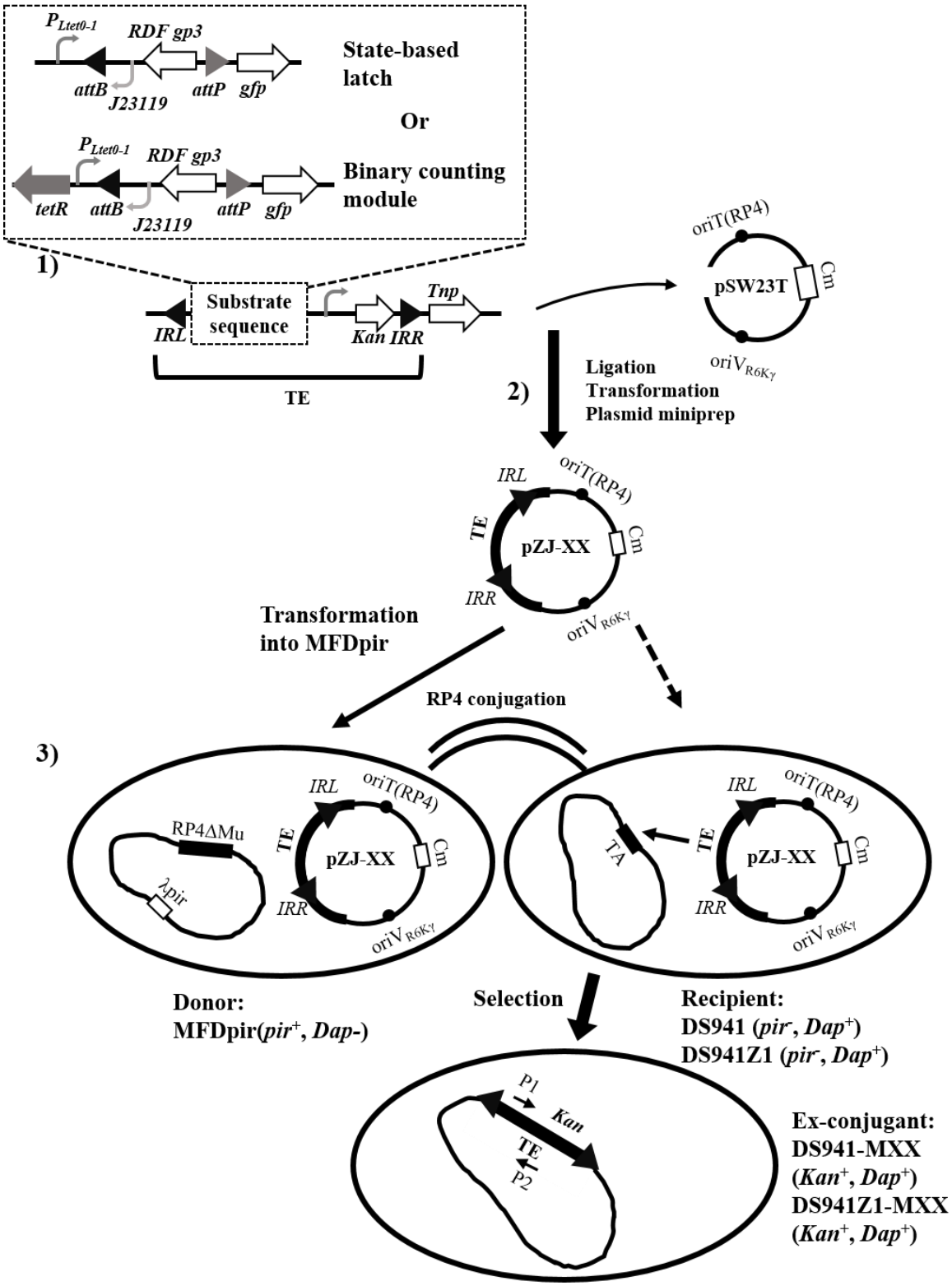

P Ltet0-1
RDF gp3
attB
J23119
attP
gfp
State-based latch
Or
tetR
Binary counting module
1)
Substrate sequence
Tnp
IRL
Kan
IRR
TE
oriT(RP4)
pSW23T
Cm
oriV R6Kγ
2)
Ligation
Transformation
Plasmid miniprep
pZJ-XX
Transformation into MFDpir
RP4 conjugation
3)
RP4ΔMu
λpir
TA
Donor:
MFDpir(pir+, Dap-)
Selection
Recipient:
DS941 (pir-, Dap+)
DS941Z1 (pir-, Dap+)
P1
P2
Kan
Ex-conjugant:
DS941-MXX
(Kan+, Dap+)
DS941Z1-MXX
(Kan+, Dap+)

Figure 5. 1 Process of integrating the invertible substrate sequence into the chromosome.

1) Construction of the transposable element (TE). The TE was first constructed on a high copy number (ColE1 origin) plasmid pZJ71 containing a kanamycin resistance gene (between the ISY100 IRs, IRL and IRR) and ISY100 transposase gene (outside of the ISY100 IRs) in tandem and controlled by the same constitutive promoter. One of the invertible substrate sequence of memory device (from the state-based latch or the binary counting module) was cloned between the ISY100 IRs on plasmid pZJ71.

2) Construction of the suicide transposon delivery plasmid. The sequence containing the TE and transposase gene was moved to the suicide plasmid pSW23T and transformed into *pir*$^{+}$ strain π3. The newly constructed plasmid pZJ-XX was extracted from π3.

3) Mobilization of the suicide transposon delivery plasmid. Both transformation and conjugation were used to move the suicide plasmid into target strain DS941 (*pir*$^{-}$) or D9S41Z1 (*pir*$^{-}$). No transformants were obtained after either chemical transformation or electroporation. For the conjugation reaction, plasmid pZJ-XX was introduced into strain MFDpir by transformation, and conjugation was carried out between strain MFDpir/ pZJ-XX and DS941 or DS941Z1. The donor strain (MFDpir) is a diaminopimelic acid (DAP) auxotroph, and only grows in medium supplement with DAP. The recipient cell (DS941 or DS941Z1) is kanamycin sensitive. The successful ex-conjugant strain (named DS941-MXX or DS941Z1-MXX) containing a copy of TE sequence (including the kanamycin resistance gene) on the chromosome was selected on a kanamycin selection L-Broth agar plate without DAP.

The "XX" represents different plasmid or strain constructions. Chromosomal DNA is represented as a wiggly line, and plasmid is shown as circle. The drawing is not to scale.

Since the *gp3* gene was part of the invertible substrate sequence for both the state-based latch and the binary counting module, its expression was expected to be lower in these single copy chromosomal devices than in the previously characterised plasmid-borne devices. To keep sufficient Gp3 expression for the *attR*×*attL* recombination, the substrate sequences containing stronger RBSs for Gp3 were chosen to create chromosomal switches. The invertible substrate sequence for the state-based latch was from pZJ53off (section 4.3.1), and the invertible substrate sequence for the binary counting module was from pZJ68off (section 4.4.1).

5.3 Methods used to analyse the performance of the chromosomal device

To construct the chromosomal state-based latch (module 1, M1), the invertible substrate sequence (shown in Fig.4.4) from pZJ53off was integrated to the chromosome of *E. coli* strain DS941Z1 using the process described above (section 5.2), resulting in strain DS941Z1-M1. The performance of this strain was tested using fluorescent microscopy, flow cytometry, and colony PCR.

Strain DS941Z1-M1 was transformed with Int expression plasmid pZJ7 (Int expression plasmid constructed in section 3.3.1), resulting in strain DS941Z1-M1/pZJ7. The Int gene was under the control of an arabinose inducible P_{BAD} promoter, and *gp3* was under the

control of the invertible switch and the aTc inducible $P_{Ltet0-1}$ promoter. It was hypothesized that a short pulse induction should be enough to complete the recombination on the single copy substrate (section 4.3.2.2), so a half hour induction pulses of arabinose (for the set reaction), and arabinose and aTc (for the reset reaction) were used as an initial test condition to trigger the reaction. The recombination efficiencies were assayed by checking the fluorescence and DNA sequence states of cell samples (Fig. 5.2).

A

Architecture of the substrate sequence	Cell fluorescent state	Colony PCR product (P1-P2)	Colony PCR product (P2-P3)
$P_{Ltet0-1}$ p1 BP state *attB* *gp3* *gfp* *attP* J23119 p3 p2	GFP OFF	1125 bp	No product
$P_{Ltet0-1}$ p3 J23119 RL state *attR* *gp3* *gfp* *attL* p1 p2	GFP ON	No product	955 bp

B

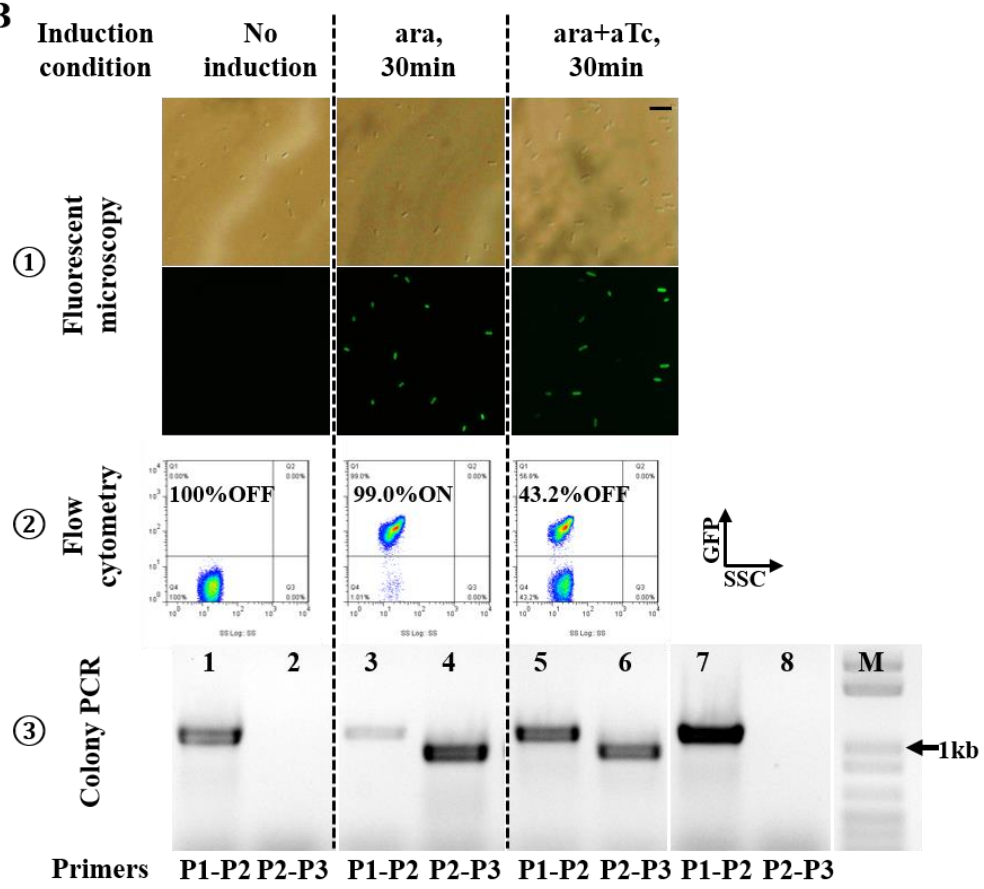


Figure 5. 2 Methods used to check the recombination performance of the chromosomal device.

(A) The architecture of the substrate sequence in BP (OFF) or RL (ON) state, and expected performances of cells containing either sequence on the chromosome. Cells containing the OFF state substrate sequence should not be fluorescent and the colony PCR product is 1125 bp with primers P1 and P2. Cells containing the ON state substrate sequence should be fluorescent and the colony PCR product is 955 bp with primers P3 and P2.

(B) Analysis of the recombination efficiencies. Strain DS941Z1-M1/pZJ7 was induced in a cycle of set (with half hour pulse of arabinose) and reset (half hour pulse of arabinose and aTc) reactions (as described in section 2.27.2). Cells collected after each induction were checked by three methods:

1) Fluorescent microscopy, the phase contrast images (top) and fluorescent images (bottom) were recorded using fluorescent microscope (scale bar: 5um).

2) Flow cytometry, the population distribution GFP (y axis) and side scatter (x axis) levels are shown using the data collected from the CyAn ADP Analyzer (cells were gated by forward and side scatter, and the proportion of cells on either the ON or OFF state is shown on top of the each image);

3) colony PCR, the overnight culture (1 μl) of cells was used for colony PCR (in the Thermo Scientific Taq master mix solution, with two pairs of primers P1 and P2, P2 and P3 respectively), products were visualized by agarose gel electrophoresis (lanes 1-6, overnight cultures amplified with corresponding primers; lane 7, the control plasmid DNA containing the substrate sequence in OFF state was amplified by P1 and P2 primers; lane 8, the same control plasmid was amplified by P2 and P3 primers).

In each area separated by dashed lines, the tested cells were from the same cell population.

The fluorescent state of individual cells was visualized using fluorescent microscopy and also analysed by flow cytometry. Both methods indicated similar fluorescent states of cell populations. The whole population remained non-fluorescent before induction. Nearly 100% of cells were changed to fluorescent state by set induction with 30 min pulse of arabinose, and then nearly 50% of cells were restored to non-fluorescent state by reset induction with 30 min pulse of arabinose and aTc (Fig. 5.2B). Microscopy images showed the fluorescent state of individual cells, but visualising a large number of cells using microscope was relatively difficult, and long-time exposure to blue light caused GFP bleaching, which might influence the counting result. Flow cytometry results revealed the fluorescent strength of individual cells in a large population (30,000 cells) and showed the fluorescent strength of individual cells. Therefore, in the following sections, flow cytometry was used as the main method for checking recombination reactions.

PCR was performed on liquid cell culture directly to check the state of the substrate DNA on the chromosome. The relative amount of substrate in OFF or ON state within one sample was indicated by the corresponding PCR products. Only OFF state product was amplified from cells before induction, much less OFF state product than ON state product was amplified from cells induced by arabinose for 30 min, and about the same amount of OFF and ON state products were amplified from cells induced by arabinose and aTc for 30 min (Fig. 5.2B). This method is able to detect if a specific DNA sequence is present in cells or not, but is not reliable for indicating the relative amounts of the substrate sequence in OFF or ON state. To achieve precise ratios of OFF/ON substrates, more accurate technology should be used, like real-time PCR.

It was predicted that cells containing a single chromosomal copy of the invertible substrate sequence should either be in the ON or the OFF state and it should not be possible for a single cell to be in an intermediate state. In agreement with this prediction, flow cytometry results showed that cells were in one of two different states (fluorescent or non-fluorescent), with very few cells (< 0.5%) in an intermediate fluorescent state (Fig. 5.2B). Therefore, the single copy chromosomal substrate avoided the multiple cell states caused by different ratios of OFF/ON substrate plasmids in different cells. However, it was noticed that after the reset operation (30 min arabinose and aTc induction), only part of the cell population was restored to the OFF state and there were two distinct cell populations. The efficient set reaction (30 min arabinose induction, converting 99% of cells to ON) demonstrated that sufficient Int for *attB*×*attP* recombination was expressed with 30 min induction pulse. Therefore, insufficient

expression of Int was probably not the reason for the inefficient *attR*×*attL* reaction. Three possible reasons were proposed: 1) induction time was not enough for activating the expression of sufficient Gp3; 2) the expression level of Gp3 from a single copy DNA sequence was low, resulting in insufficient ratio of Gp3 to Int; 3) Int with Gp3 caused both set and reset reaction. Based on these speculations, experiments were carried out to characterise and improve the behaviours of the chromosomal state-based latch.

5.4 Investigation of factors affecting recombination efficiencies of chromosomal state-based latch

5.4.1 Effects of induction pulse length on recombination efficiencies

Since the *gp3* encoded gene within the substrate sequence was integrated into the chromosome, the 30 min induction pulse might not be long enough to express sufficient Gp3 protein for efficient *attR*×*attL* recombination. A time course was carried out by inducing strain DS941Z1-M1/pZJ7 with different lengths (0min, 10 min, 20min, 30min, 1h, and 2h) of arabinose pulse for the set reaction to test if a short induction was the reason for inefficient recombination. Subsequently, the reset reaction was carried out using same length (30min, 1h, and 2h) of arabinose and aTc as the pulse used for the set reaction. The fluorescent states of cell populations were checked by flow cytometry (Fig. 5.3). Since the set recombination with 0 min, 10 min, or 20 min induction pulse was not efficient (Fig. 5.3), no subsequent reset operation was conducted using these lengths of pulse.

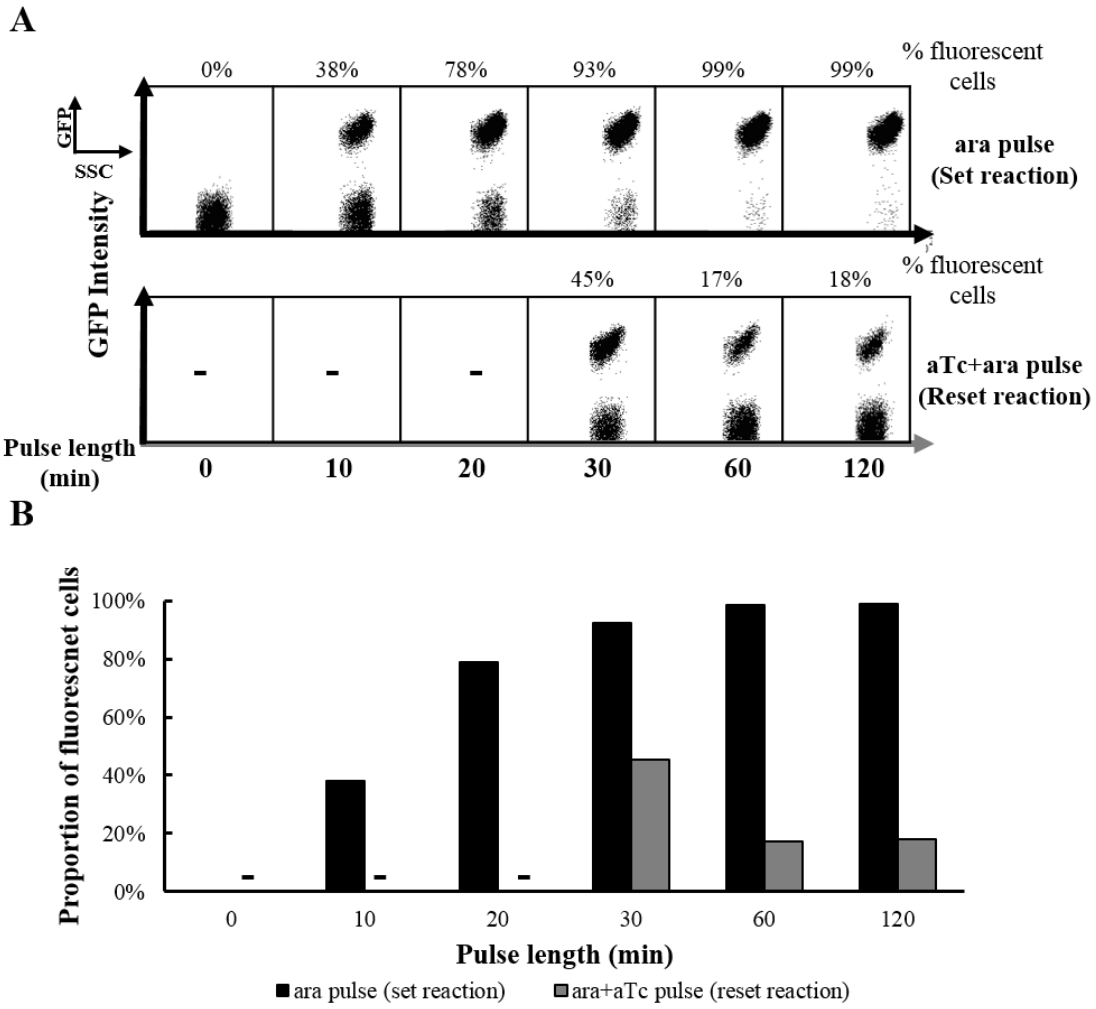


Figure 5. 3 Flow cytometry analysis of the chromosomal state-based latch under pulsed induction time course. Strain DS941Z1-M1/pZJ7 was induced in a cycle of set-reset reaction with different lengths (0-120 min) of arabinose pulse for the set reaction and different lengths of arabinose and aTc (30-120 min) for the reset reaction. After pulses of indicated lengths of time, cells were diluted 1:1000 into LB-medium with glucose and allowed to grow overnight. (A) Cells collected after each overnight culture were centrifuged, fixed, and checked by CyAn ADP Analyzer. The data were analysed by flowJo software, and the population distribution GFP (y axis) and side scatter (x axis) levels are shown in cytometry images (cells are gated by forward and side scatter). (B) The proportion of fluorescent cells in each sample is shown in the graph. The dark grey bars represent the samples after the set reaction, and the light grey bars represent the samples after the reset reaction.

Reactions reached their maximum efficiencies in one hour for both set reaction (greater than 99%) and reset reaction (greater than 83%, Fig. 5.3). The set reaction efficiency (93%, Fig. 5.3) with 30 min arabinose induction was slightly lower than that achieved previously (99%, Fig. 5.2). These results suggested that the 30 min length of induction pulse might not be enough for stable and efficient set and reset reactions. Therefore, the one hour length of induction pulse was used in the following tests.

The efficiency of *attR*×*attL* recombination was still not ideal even with longer induction pulse. The possible way to improve this efficiency was raising the ratio of Gp3 to Int or

increasing the relative degradation rate of Int to Gp3. To raise the ratio of Gp3 to Int, the Gp3 expression level could be increased or the Int expression level could be decreased through adjusting their RBS strengths. To increasing the relative degradation rate of Int to Gp3, an ssrA-mediated degradation tag could be added to the C-terminus of Int (Andersen *et al*., 1998; Keiler and Sauer, 1996). Since the *gp3* gene was on the chromosome and the plasmid can be altered more easily than the chromosome, experiment was carried out to change the expression levels and degradation rate of Int.

5.4.2 Effects of Int expression level and degradation rate on recombination efficiencies

It was hypothesized that lower Int expression level and faster Int degradation rate might be helpful for efficient *attR*×*attL* recombination, since the ratio of Gp3 to Int would be high enough. To test this hypothesis, Int expression plasmid pZJ7m (weaker start codon and lower predicted translation initial rate than pZJ7, Table 5.1) was used to reduce Int expression level.

A degradation tag was added to the C-terminus of Int to attempt to increase its degradation rate. Proteins carrying a C-terminal AANDENYALAA peptide tag, are recognized and rapidly degraded by tail-specific Tsp proteases present in both the cytoplasm and periplasm of *E. coli*. (Andersen *et al*., 1998; Keiler and Sauer, 1996). Mutants in the last three residues of this peptide consensus sequence are known to alter protein stability (Andersen *et al*., 1998; Keiler and Sauer, 1996). Protein with peptide sequence AANDENYALAA (LAA) was suggested to be less stable than protein with peptide sequence AANDENYAAAV (AAV) (Andersen *et al*., 1998). New Int expression plasmids were constructed by cloning a nucleotide sequence corresponding to these two peptides just before the stop codon of Int in plasmid pZJ7, resulting in pZJ46 (LAA) and pZJ47 (AAV) (Table 5.1).

It was suggested that the use of degradation tags might decrease the steady state concentration of protein (Arpino *et al*., 2013). To maintain sufficient steady state Int concentration, a stronger RBS (predicted translation initial rate: 7263, Salis *et al*., 2009) was also used, replacing the RBS (predicted translation initial rate: 227, Salis *et al*., 2009) for Int in plasmid pZJ7, pZJ46, and pZJ47, resulting in plasmid pZJ88, pZJ91(LAA), and pZJ94(AAV), respectively. This gave a total number of seven different Int expression plasmids with either different predicted translation initiation rates or degradation rates (Table 5.1).

To test the effects of the Int expression levels and degradation rates on recombination efficiencies, one cycle of set-reset operation was carried out with strain DS941Z1-M1 containing each of these seven Int expression plasmids (Table 5.1), using one hour induction pulses (arabinose for set reaction, and arabinose and aTc for reset reaction). The fluorescence of cell populations were then checked by flow cytometry (Fig. 5.4).

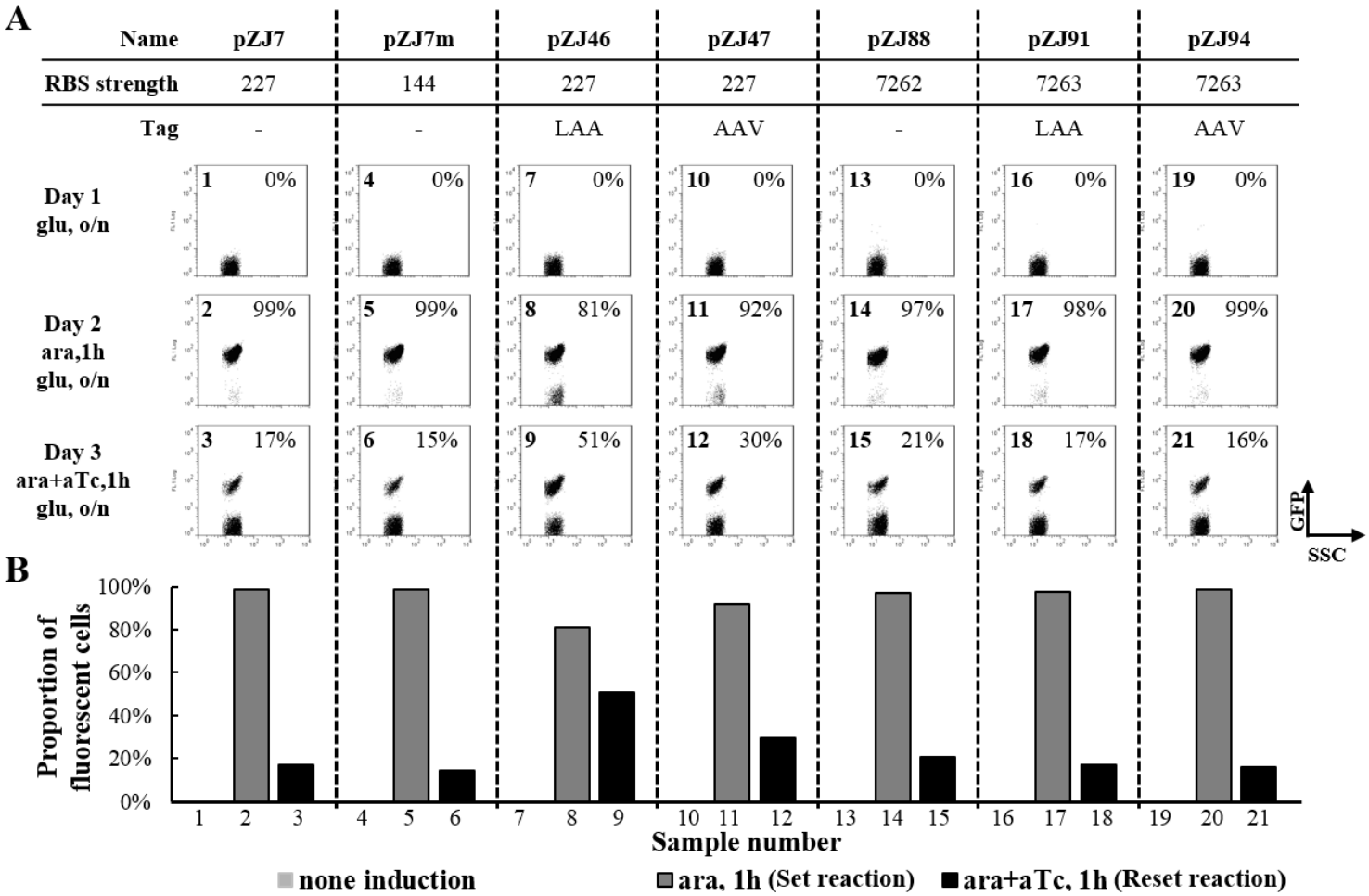


Figure 5. 4 Recombination efficiencies with different Int expression and degradation conditions. Strain DS941Z1-M1 containing different Int expression plasmids (pZJ7, pZJ7m, pZJ46, pZJ47, pZJ88, pZJ91, and pZJ94) was induced in a cycle of set (with one hour pulse of arabinose) and reset (one hour pulse of arabinose and aTc) reactions. Induced cells were diluted 1:1000 in LB-medium with glucose and allowed to grow overnight. The overnight cultures were centrifuged, fixed, and checked by CyAn ADP Analyzer. (A) The data were analysed by flowJo software, and the population distribution GFP (y axis) and side scatter (x axis) levels are shown in cytometry images (cells were gated by forward and side scatter) with information of Int expression plasmids above. (B) The proportions of fluorescent cells are shown both on the corresponding flow cytometry images and in the graph.

Table 5. 1 Int expression plasmids and recombination performance abtained

Name	RBS sequence and Start codon	Predicted translation initiation rate (au)	Degradation tag	Efficiency	
					Reset
pZJ7	GAGCTCAGGAGTAACAT**ATG**	227	None	99%	83%
pZJ7m	GAGCTCAGGAGTAACAT**GTG**	144	None	99%	85%
pZJ46	Same as pZJ7	227	LAA	81%	49%
pZJ47	Same as pZJ7	227	AAV	92%	70%
pZJ88	GAGCTCCAGCTCGAGAGGAG GTACAT**ATG**	7263	None	97%	79%
pZJ91	Same as pZJ88	7263	LAA	98%	83%
pZJ94	Same as pZJ88	7263	AAV	99%	84%

Comparing the recombination behaviours mediated by Int without degradation tag (pZJ7m, pZJ7, and pZJ88), similar efficiencies were obtained for *attB*×*attP* recombination, but Int with higher predicted translation initiation rate mediated slightly less efficient *attR*×*attL* recombination (Fig. 5.4). These results showed that the amounts of Int expressed from all three plasmids were enough for efficient *attB*×*attP* recombination (with greater than 95% efficiency, Fig. 5.4). However, it appears that a higher Int expression level leads to lower ratio of Gp3 to Int, resulting in poor efficiencies of *attR*×*attL* recombination.

Comparing the recombination performances mediated by Int with weaker RBS sequence from pZJ7, pZJ46 (LAA), and pZJ47 (AAV), Int with degradation tag mediated less efficient *attB*×*attP* and *attR*×*attL* recombination than Int without a tag, and the strong degradation tag LAA caused a great reduction in recombination efficiency (Fig. 5.4). These results can be explained by the statement that the degradation tag reduces the induced steady state concentration of Int, resulting in inefficient recombination. Using stronger predicted RBS sequence (predicted translation initial rate: 7263, Salis *et al*., 2009) to replace the weaker RBS of Int, both the *attB*×*attP* and *attR*×*attL* recombination efficiencies were increased (comparing pZJ91 with pZJ46, and pZJ94 with pZJ47, Table 5.1). These results suggested that tuning the RBS strength of Int could compensate for the decrease of steady state protein levels caused by the addition of degradation tag.

Furthermore, comparing the recombination performances mediated by Int with stronger RBS from pZJ88, pZJ91 (LAA), and pZJ94 (AAV), the tagged Int from pZJ91 and pZJ94 showed higher recombination efficiencies, especially for the *attR*×*attL* recombination (Fig. 5.4). This

indicated that by adding a C-terminal degradation tag to Int, an optimised ratio of Gp3 to Int was achieved or an effective ratio was maintained during the proteins degradation stage after removal of arabinose and aTc.

In summary, either by reducing the expression level of Int (pZJ7m), or by combining a high expression level with a strong degradation tag (pZJ94), *attB*×*attP* recombination of greater than 95% and *attR*×*attL* recombination of greater than 85% could be obtained.

5.5 Reliability test of the chromosomal state-based latch under multicycle set-reset operation

The Int expression plasmids pZJ7m and pZJ94, which showed the best recombination efficiencies were selected to test the reliability of the chromosomal state-based latch in a multicycle set-reset operation. Reactions were carried out with strain DS941Z1-M1/pZJ7m and strain DS941Z1-M1/pZJ94 for five cycles using one hour induction pulse (arabinose for set reaction and arabinose plus aTc for reset reaction). The distribution of fluorescence intensities from different cells was checked by flow cytometry (Fig. 5.5).

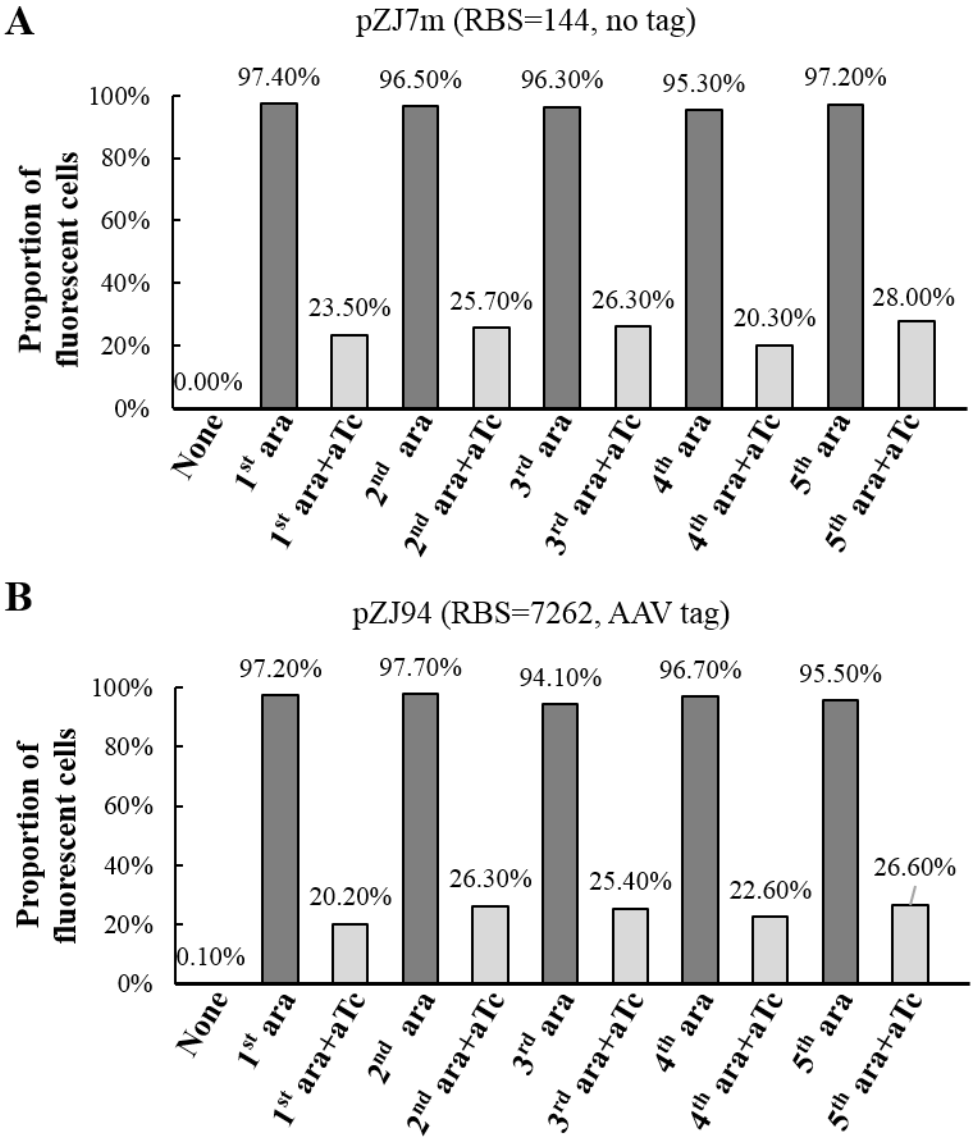


Figure 5. 5 Five cycles of set-reset operation with chromosomal state-based latch. (A) DS941ZI-M1 strain was transformed with Int expression plasmid pZJ7m. Resultant colony was inoculated and induced by five cycles of set (with one hour pulse of arabinose) and reset (one hour pulse of arabinose and aTc) reactions. Cells collected after each induction, were centrifuged, fixed and checked by CyAn ADP Analyzer. The data were analysed by flowJo software, and the proportions of fluorescent cells are shown in the graph. (B) DS941ZI-M1 strain was transformed with Int expression plasmid pZJ94, and same cycles of operation were carried out as described in (A).

The proportions of fluorescent cells changed repeatedly with alternating arabinose and arabinose plus aTc induction pulses (Fig. 5.5). Int from both Int expression plasmids (pZJ7m and pZJ94) mediated highly efficient set (*attB*×*attP*) reaction (greater than 95% of cells were switched to GFP ON state) and efficient reset (*attR*×*attL*) reaction (about 75% of cells were restored to GFP OFF state), and no apparent decrease of efficiency was detected during these

five cycles operation. This demonstrated that the state-based latch operated reliably for at least five set-reset cycles.

However, the recombination efficiencies were not totally consistent in different cycles, and the highest recombination efficiencies achieved in this experiment were slightly lower than that achieved in the previous testing (Fig. 5.4). It was possible that the culture and induction conditions were not identical during different induction time, resulting in the fluctuation of recombination efficiencies. For instance, the freshness of inducers might affect the strength of the inducible promoters. It was reported that inducer aTc binds to TetR with a higher affinity in the presence of Mg^{2+} (Leypold *et al*., 2004), thus different batches of medium containing various amounts of Mg^{2+} might lead to different induction of $P_{Ltet0-1}$ promoter for Gp3 expression.

5.6 Test of the chromosomal devices with a single signal

The chromosomal state-based latch performed reliable set (*attB*×*attP*) and reset (*attR*×*attL*) recombination under multicycle operation with two alternative input signals (arabinose for set, and arabinose and aTc for reset). As the purpose of this study was to construct a device that can change between two distinct states in response to a single signal, this section mainly talks about the performance of the chromosome devices when using a single repeating signal to trigger the reaction.

To control the change of this system with a single signal, the invertible substrate sequence of the state-based latch (from pZJ53off) was integrated into the chromosome of the DS941, a strain lacking the Tet repressor, resulting in strain DS941-M1. In this strain, the $P_{Ltet0-1}$ promoter is a constitutive promoter, and Gp3 expression will be determined only by the substrate sequence state. In constrast to the state-based latch where the set reaction was triggered by arabinose, and the reset reaction was triggered by arabinose and aTc (section 5.4 and 5.5), arabinose was used as the single input signal to trigger both switch-on (*attB*×*attP*) and switch-off (*attR*×*attL*) reactions in the following experiments. It was expected that the first arabinose pulse would turn on the expression of Int, mediating *attB*×*attP* recombination and switching the device from the BP (GFP and Gp3 OFF) state to the RL (GFP and Gp3 ON) state. Ideally, Gp3 would not be expressed quickly enough to reverse the reaction, but then Gp3 would be present for the second pulse of integrase

allowing *attR*×*attL* recombination and switching the device from the RL (GFP and Gp3 ON) state back to the BP (GFP and Gp3 OFF) state.

In chapter 4, the efficiency of the plasmid borne single-signal controlled system was improved by introducing a time delay for Gp3 expression, allowing the construction of a binary counting module. To test if the binary counting module (module 2, M2) could also perform higher recombination efficiency on the chromosome, the invertible sequence of the binary counting module (from plasmid pZJ68off, section 4.4.1) was integrated into the chromosome of DS941 (*tetR*$^-$), resulting in strain DS941-M2. Delay of Gp3 expression from Tet repressor was expected to be helpful for efficient *attB*×*attP* recombination.

The Int expression plasmids pZJ7m and pZJ94 which showed the best recombination performance among all of the tested Int expression plasmids (section 5.3.2) were used to execute recombination in DS941-M1 and DS941-M2. Different lengths of arabinose pulse were used in a full cycle of switch-on switch-off time course reaction to investigate the effect of induction pulse length on recombination efficiencies. The fluorescent states of cell populations were checked by flow cytometry (Fig. 5.6 and Fig. 5.7).

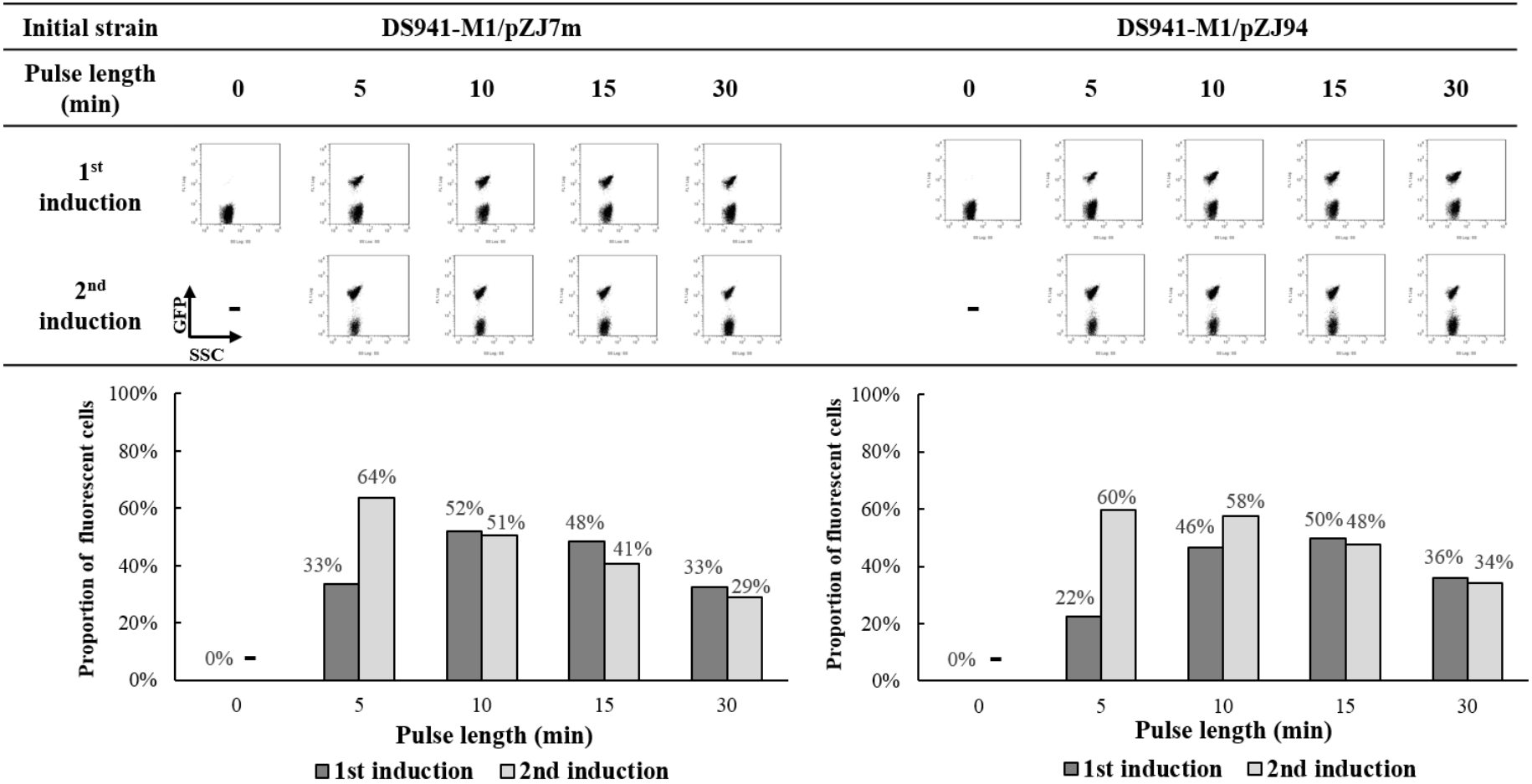


Figure 5. 6 Pulsed induction time course with chromosomal state-based latch (without time delay from TetR) using a single repeating signal. Strain DS941-M1/pZJ7m (left) and DS941-M1/pZJ94 (right) were induced for the 1st time with two hour preculture time and different lengths (5-30 min) of arabinose pulse followed by dilution and growth overnight in L-Broth with glucose. Subsequently, arabinose pulses with the same length (5-30 min) as used for the 1st induction were used for the 2nd induction followed by dilution and growth overnight in L-Broth with glucose. Cells collected after each induction were centrifuged, fixed and checked by CyAn ADP Analyzer. The data were analysed by flowJo software, and the population distribution GFP (y axis) and side scatter (x axis) levels are shown in cytometry images (cells were gated by forward and side scatter). The proportions of fluorescent cells in each sample are shown in the graph. The dark grey bars represent the samples after the 1st induction and the light grey bars represent the samples after the 2nd induction.

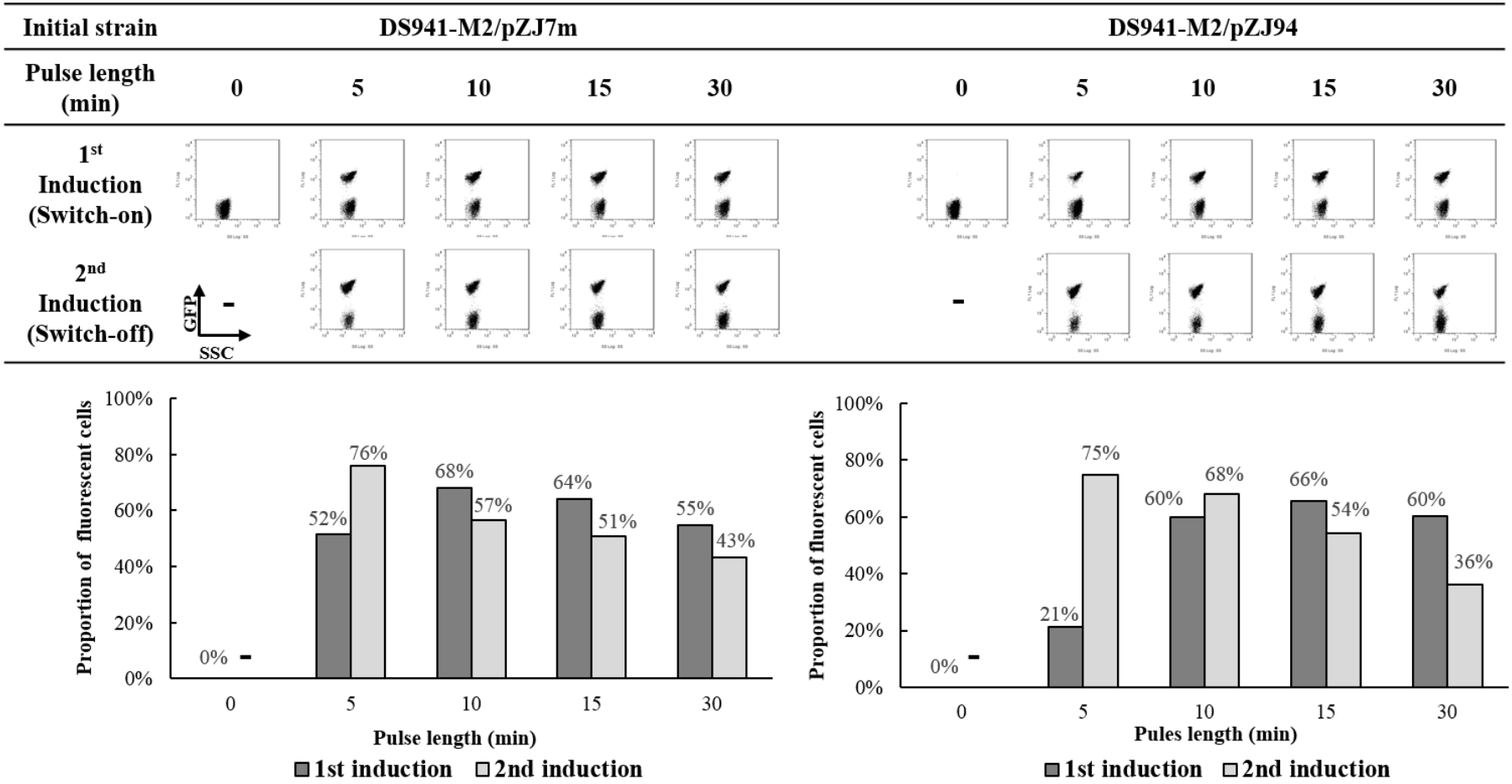


Figure 5. 7 ***Pulsed induction time course with chromosomal binary counting module (with time delay from TetR) using a single repeating signal.*** Strain DS941-M2/pZJ7m (left) and DS941-M2/pZJ94 (right) were induced for the 1st time with two hour preculture time and different lengths (5-30 min) of arabinose pulse followed by dilution and growth overnight in L-Broth with glucose. Subsequently, arabinose pulses with the same length (5-30 min) as used for the 1st induction were used for the 2nd induction followed by dilution and growth overnight in L-Broth with glucose. Cells collected after each induction were centrifuged, fixed and checked by CyAn ADP Analyzer. The data were analysed by flowJo software, and the population distribution GFP (y axis) and side scatter (x axis) levels are shown in cytometry images (cells were gated by forward and side scatter). The proportions of fluorescent cells in each sample are shown in the graph. The dark grey bars represent the samples after the 1st induction and the light grey bars represent the samples after the 2nd induction.

To analyse the data shown in Figure 5.6 and Figure 5.7, the proportion of fluorescent cells after the first induction and second induction are represented by "F_1" and "F_2", respectively. The switch-on and switch-off reaction efficiencies are represented by "X" and "Y", respectively. It was assumed that: 1) the proportion of cells changing from non-fluorescent to fluorescent or changing from fluorescent to non-fluorescent remains constant for a given pulse length; 2) the *attB×attP* recombination efficiency equals to the proportion of fluorescent cells after the first induction ($X=F_1$), since the reaction started with pure BP substrate and all of the RL product was regarded to be generated through *attB×attP* recombination. Using these assumptions, the *attR×attL* recombination efficiency can be calculated according to the recombination process and equation shown in Figure 5.8.

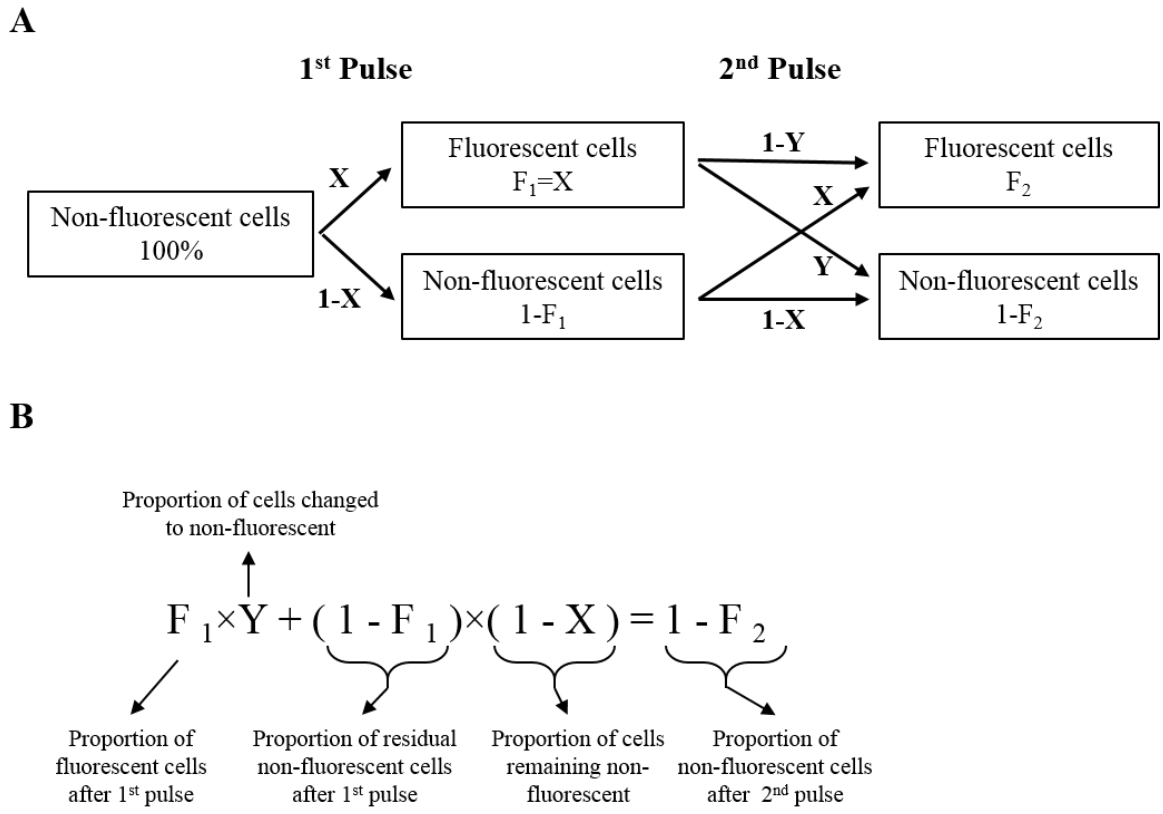


Figure 5. 8 Recombination process with chromosomal single-signal controlled module and the equation used to estimate switch-on and switch-off reaction efficiencies. (A) Schematic representation of a cycle of switch-on and switch-off reaction using a single repeating signal. After the first induction, cells changed from complete non-fluorescent (BP) to fluorescent (RL) with a switch-on efficiency of 'X', resulting in a proportion of cells ($F_1=X$) fluorescent. After the second pulse, some of the residual non-fluorescent cells switched to fluorescent with an switch-on efficiency of 'X', and some of the fluorescent cells switched to non-fluorescent with a switch-off efficiency of 'Y', resulting in the proportion of fluorescent cells changed from 'F_1' to 'F_2'. The states of the cell population are shown in the boxes, and the changes caused by each induction are shown near the line arrows. (B) Equation used to estimate the efficiency of *attR×attL* recombination. After the second pulse, the fluorescent cells (F_1) changed to non-fluorescent with an efficiency of 'Y', the non-fluorescent cells ($1-F_1$) retained non-fluorescent with an efficiency of '1-X'. As 'F_1', 'F_2', and 'X' can be achieved from the figure, the switch-off efficiency 'Y' can be calculated using this equation.

Using the equation described in Figure 5.8, the switch-on reaction efficiency (X) and the switch-off reaction efficiency (Y) were calculated for different pulse lengths (5-30 min), chromosomal constructs (M1 or M2), and Int expression plasmids (pZJ7m or pZJ94) and shown in Table 5.2.

Table 5. 2 Recombination efficiencies of chromosomal devices controlled by a single signal

Pulse length (min)	M1/pZJ7m		M1/pZJ94		M2/pZJ7m		M2/pZJ94	
	X	Y	X	Y	X	Y	X	Y
5	33%	-	22%	-	52%	-	21%	-
10	52%	51%	46%	29%	68%	49%	60%	26.2%
15	48%	68%	50%	55%	64%	57%	66%	52%
30	33%	78%	36%	69%	55%	66%	60%	80%

'-' indicates the calculated efficiency is a negative number, the best switch-on efficeicnies and the best switch-off efficiencies are highlighted using boxes and underlines, respectively.

In *vivo* time courses with all of these modules showed that a specific range of pulse lengths (10-15 min) was required for the most efficient switch-on reaction (~50% for module 1 and ~65% for module 2, boxes in Table 5.2), while a longer pulse length led to higher switch-off reaction (>60% with 30 min, underlines in Table 5.2). The decrease of the switch-on recombination efficiency observed for longer induction times (>10 min for reactions with pZJ7m; >15 min for reactions with pZJ94) might be because Gp3 starts to be produced shortly after the substrate sequence is recombined to the RL (ON) state. However, a higher maximum switch-on efficiency was achieved with module 2 (with *tetR* delay) than that with module 1 (without *tetR* delay), which indicates that the addition of *tetR* is working to delay the Gp3 expression.

It was expected that the system could be switched from ON to OFF by the second induction pulse. According to the calculated recombination efficiencies, short induction times (5 min and 10 min) appear to be too short to trigger stable and efficient switch-off recombination. The switch-off (*attR*×*attL*) was more efficient with a longer induction time (maximum switch-off efficiency reached 80% after 30 min induction with DS941-M2/94, Table 5.2). However, this was difficult to observe through checking the final proportion of fluorescent cells because not all cells were in ON state after the first induction pulse.

5.7 Conclusion and discussion

In this chapter, the behaviours of chromosomal state-based latch and the chromosomal binary counting modules were explored. Experiments showed that host cells containing these devices fell into either GFP ON or GFP OFF group, indicating that the mixed cell population observed with the plasmid-borne devices was no longer a problem after integrating substrate sequence on the chromosome as a single copy.

5.7.1 Transposons as a tool to deliver a single copy substrate sequence into the *E.coli* chromosome

The substrate sequence was successfully moved into the chromosome of *E.coli* using ISY100 transposition. However, their insertion occurs at random locations throughout the genome, potentially resulting in insertional mutagenesis or inappropriate activation of nearby genes in host cells. If a fixed location is required, site-specific recombination-based chromosomal integration (St-Pierre *et al*., 2013) or the recently reported approach coupling the CRISPR/Cas9 system to lambda Red recombineering (Pyne *et al*., 2015) could be used as delivery systems. Since there was not any detrimental effect on cell growth and viability observed, only one strain for each chromosomal module was tested and introduced in this chapter.

To deliver the transposon into the target strain, it was placed on a suicide plasmid. Using strain MFDpir as a donor, this plasmid was conjugated into a recipient cell in which the plasmid cannot replicate. The conjugation reaction had a higher efficiency than transformation in mobilizing the suicide plasmid. It should be noticed that transforming the ligation product of the suicide delivery plasmid (step 2, Fig. 5.1) into MFDpir (Δ*recA*) strain directly was less efficient than transforming strain π3 (*rec*$^+$). This might because of the constitutively expressed transposase from the plasmid, as it was reported that the transposase influences the cell growth, but the *recA*$^+$ strain had better tolerance of this influence (Urasaki *et al*., 2002). To solve this problem, it might be better to place the transposase gene under the control of an inducible promoter, and activate transcription when the transposition reaction is required.

5.7.2 Efficient and reliable recombination reaction with the chromosomal state-based latch (two signals controlled system)

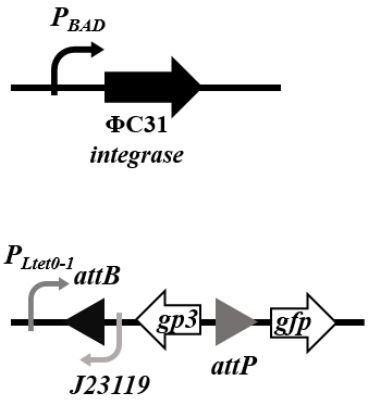
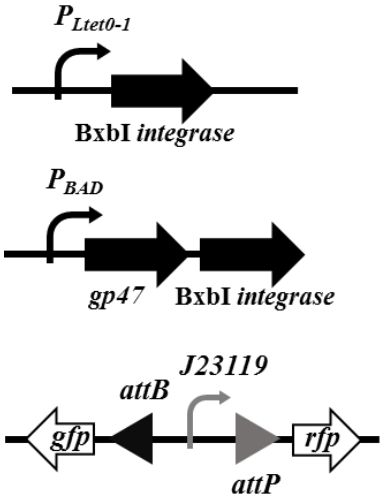
The substrate sequence of the state-based latch was successfully integrated into the chromosome of *E. coli* DS941Z1. It was confirmed that fluorescence microscopy and flow cytometry could be used accurately to determine the state of the chromosomal switch, while colony PCR confirmed that the inversion reaction was taking place but could not be used to check the recombination efficiencies so precisely.

Highly efficient *attB*×*attP* recombination (>95%) was achieved with the state-based latch on the chromosome (Fig. 5.5), but *attR*×*attL* recombination was less efficient (~80%, Fig. 5.5) under the induction of two different input signals. This might be due to reduced expression of Gp3 from a single gene copy on the chromosome, compared to expression from a multicopy plasmid.The state-based latch used had been optimised for *attB*×*attP* recombination on a plasmid borne device, by reducing the Gp3 RBS strength to decrease Gp3 expression. It appeared that Gp3 expression levels were still sufficient for the *attR*×*attL* recombination though a weaker RBS for Gp3 was used. However, different from the optimisation process used on the plasmid-borne device, for chromosomal *attR*×*attL* recombination might require increasing Gp3 RBS strength in order to increase the Gp3 to Integrase ratio.

As tuning the Gp3 expression through manipulating the sequence on the chromosome is time-consuming, instead, the Int expression level and degradation rate were adjusted to attempt improving the ratio of Gp3 to Int. This was done by using a weaker start codon and adding a fast degradation tag to Int. More efficient *attR*×*attL* reactions were achieved with these new Int expression plasmids (pZJ7m and pZJ94, Fig. 5.4). However, the way in which the degradation tag of Int influenced the recombination efficiency was not confirmed. It was reported that adding the degradation tag decreased the steady state concentration of protein and increased the degradation rate of protein (Arpino *et al*., 2013). Therefore, it was possible that the *attR*×*attL* recombination benefited from a higher ratio of Gp3 to Int since the steady concentration of Int was decreased by the degradation tag. Another possibility is that the ratio of Gp3 to Int remained high during the protein degradation stage after removal of arabinose and aTc because of the fast degradation of Int relative to Gp3. To confirm the reason, the kinetics of the reaction need to be analysed in a time course experiment or the expression condition of Int should be checked by experiments (such as Western blot assay).

The chromosomal state-based latch has similar properties to the rewritable recombinase addressable data module (Bonnet *et al.*, 2012). The architectures, operation, and performances of these two devices are listed in the table below (Table 5.3).

Table 5. 3 Comparison between the state-based latch and the rewritable recombinase addressable data module

Device		State-based latch	Recombinase addressable data module
Architecture			
Integrase/ RDF		ϕC31 integrase/ Gp3	BxbI integrase/ Gp47
Host strain		DS941Z1	DH5αZ1
Incubation condition		L-Broth (37 ℃)	M9 (37 ℃)
Doubling time		30 min	90 min
attB× *attP*	Input/ pulse length	arabinose/ 1 h	arabinose/ 4 h
	Efficiency	> 95%	> 95%
attR× *attL*	Input/ pulse length	arabinose+aTc/ 1 h	aTc/ 4 h
	Efficiency	~ 80%	> 90%

The same efficiency (greater than 95%) *attB*×*attP* recombination was achieved compared to the previously constructed rewritable recombinase addressable data module (Bonnet *et al.*, 2012). For the state-based latch, though the RDF *gp3* gene was integrated into the substrate sequence and placed on the chromosome as a single copy, efficient *attR*×*attL* recombination (about 80% efficiency) was observed. It should be possible to

achieve higher *attR*×*attL* recombination efficiency through increasing the Gp3 expression level. The chromosomal state-based latch constructed in this study suggested the potential of ϕC31 integrase for efficient writing and rewriting information into DNA, and suggested that it could be used to construct combinatorial data storage together with other orthogonal integrases.

The state-based latch performed reliably when cells were cultured in rich medium, with a significately shorter doubling time than used by Bonnet *et al*. The state-based latch described here worked efficiently with a short input pulse (one hour), suggesting its advantage in transient signal detection or saving of inducers during industrial production. Most importantly, placing the RDF gene into the substrate sequence provides the possibility of controlling the system with a single input signal, making it a candidate to be developed into a binary counting module. The performances of the device controlled by a single signal and its improvements are discussed below.

5.7.3 Ways to improve the single-signal controlled systems

The invertible substrate sequence of state-based latch (without *tetR*) was integrated into the chromosome of strain DS941 (*tetR*$^-$). In this strain, the $P_{Ltet0\text{-}1}$ promoter would be active all the time, resulting in automatic expression of the Gp3 as soon as the substrate goes into the RL (ON) state. The first Int pulse (triggered by the arabinose pulse) was expected to switch the device from BP (OFF) to RL (ON), turning on the expression of Gp3. The second Int pulse (again triggered by arabinose) together with Gp3 was expected to switch the device from RL (ON) to BP (OFF). In this way, the device could be controlled by a single repeating pulse.

The performance of the chromosomal state-based latch induced by a single input signal indicated that no optimal length of pulse could trigger an efficient switch-on reaction (Fig. 5.6). A short pulse cannot active sufficient Int expression, and it appeared that a long pulse allowed switching back to BP due to the rapid accumulation of Gp3. Even with an optimal length of input pulse, the state of the chromosomal state-based latch did not change any more at the cell population level after the first induction pulse (e.g. DS941-M1/pZJ94, with a 15 min pulse length, Fig. 5.9A), indicating that each pulse led to equal amount exchange between fluorescent cells and non-fluorescent cells.

The same limitation happened to the plasmid-borne devices, and was overcome by introducing the *tetR* gene into the substrate sequence (section 4.5.3). This method was also applied to the chromosomal system to delay Gp3 expression. With this change, an increase switch-on efficiency from ~50% to ~65% (Table 5.2) was achieved. However, the switch-off efficiency was decreased more than 10% when using pZJ7m as Int expression plasmid (after 15 min and 30 min pulse induction with DS941-M2/pZJ7m and DS941-M1/pZJ7m, Table 5.2). It was possible that the expression level of Gp3 was decreased by adding *tetR* to repress the $P_{Ltet0-1}$ promoter for *gp3* transcription, resulting in less efficient switch-off reaction. No obvious decrease of switch-off efficiency was seen when using pZJ94 (Int with tag) Int expression plasmid, which might be because the ratio of Gp3 to Int was still sufficient for switch-off reaction in this circumstance.

Strain DS941-M2/PZJ94 gave the best performance that 60% of cells switched from OFF to ON with the first 30 min arabinose pulse, and 80% of cells switched from ON to OFF with the second 30 min arabinose pulse. This module showed obvious change of fluorescent state after the first and second induction, and it was predicted that the state can be changed one more cycle before approaching a steady state at the cell population level according to its recombination efficiency (Fig. 5.9B). To improve the efficiency of the chromosomal binary counting module, especially the *attB*×*attP* recombination efficiency, the Gp3 expression should be further delayed, or the expression and degradation rates of Int, Gp3, and Tet repressor should be adjusted. Once *attB*×*attP* and *attR*×*attL* reaction efficiencies are improved to 80%, three cycles of switch-on switch-off behaviour can be achieved, and more than five cycles can be achieved if the efficiencies reach 90% (Fig. 5.9B and C).

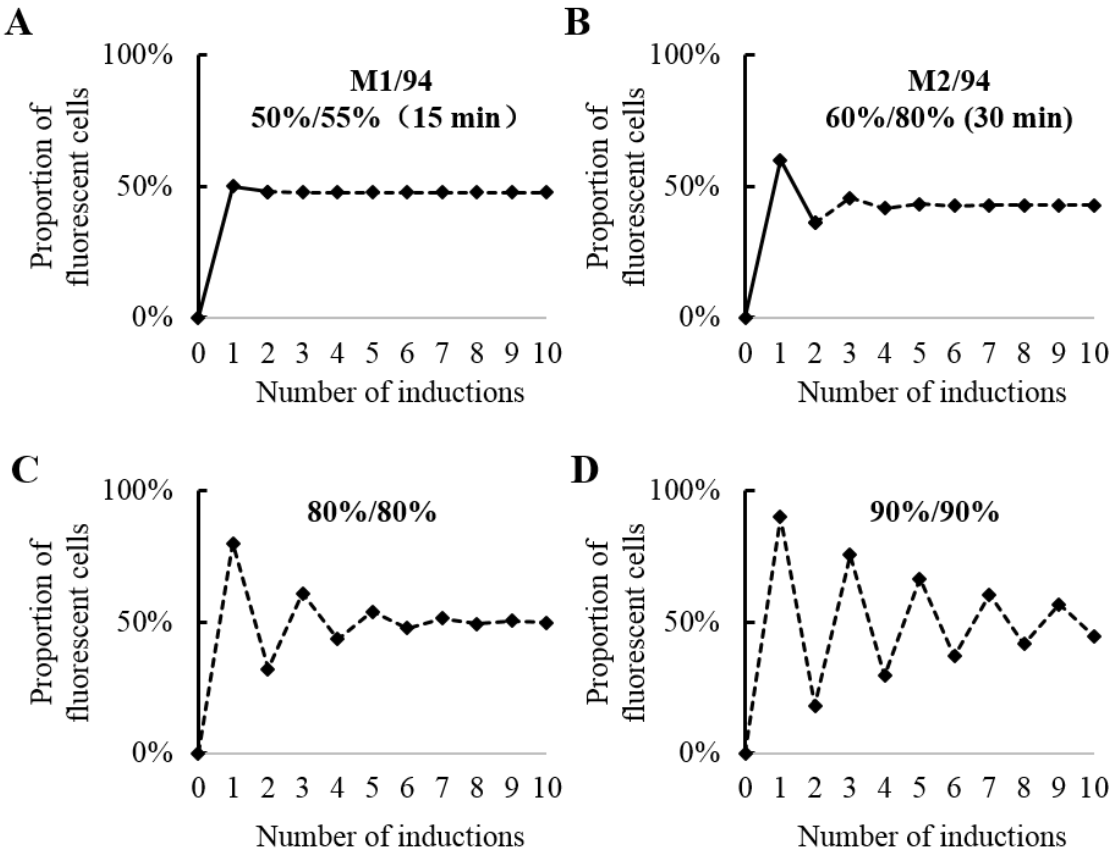


Figure 5. 9 Predicted performance of modules with different reaction efficiencies. (A)Strain DS941-M1/pZJ94 was induced by 15 min arabinose pulse. (B) Strain DS941-M2/pZJ94 was induced by 30 min arabinose pulse. The proportions of fluorescent cells (on solid line) after the 1st and 2nd induction were obtained from experimental test, and the proportions of fluorescent cells (on dash line) after more times of induction were predicted according to the estimated recombination efficiency (A and B). (C) Predicted cell fluorescent state of an optimised module with 80% efficiency for a*ttB*×*attP* and *attR*×*attL* reaction. (D) Predicted cell fluorescent state of an optimised module with 90% efficiency for a*ttB*×*attP* and *attR*×*attL* reaction. Proportion of fluorescent cells were calculated using the equation in Fig. 5.8.

It is worth mentioning that the expected performance of cells is based on the assumptions that the same length of pulse will lead to fixed *attB*×*attP* and *attR*×*attL* recombination efficiencies, and the *attB*×*attP* recombination efficiency equals to the proportion of fluorescent cells after the first induction. Though these assumptions are reasonable, replicated experiments are necessary to account for data reliability, and more times of pulses should be applied to confirm the prediction. Those calculations for the *attR*×*attL* recombination efficiencies (Table 5.2) also need to be tested by isolating a pure cell population containing only the RL (ON) substrate and subjecting to the same input pulse. Another interesting approach to further investigate the swiching performance might be following the fate of individual cells using microfludic chip.

6 Final discussion

Engineered genetic counters could be employed in a variety of applications, such as recording the age of an organism, executing cell death after a specified number of cell generations, or recording the number of occurrences of a specific environmental event. To record the occurrences of a particular event or process, a counter requires several characteristics, including a fast response time, capability to count large numbers, distinct outputs for each number, and no limitation of event lengths. Currently, none of the engineered biological counters can fulfill all of these requirements. The overall aim of this study was to investigate design principles for a new genetic counting module that could be further developed to a counter containing most of these characteristics.

Inversion recombination systems based on large serine integrases have the potential to meet the design requirements mentioned above. For example, the rewritable memory module from Bonnet *et al*. (2012), demonstrated the possibility of storing one bit of information in DNA using a set of Int-RDF. In contrast to a counter using one recombinase to store each event in a linear fashion, a combinatorial counter with N rewritable memory modules would be expected to count from 0 to 2^N-1, resulting in an increase of information storage capacity per recombinase. Another example is the toggle flip-flop designed by Subsoontorn and Endy (2012), which requires both the rising edge and the falling edge of an input signal to complete the transition from one state to the other. Thus, it is insensitive to the duration of the input pulse. However, these designs still have their own deficiency. The rewritable memory module cannot be used to record multiple repeats of a single event, while the toggle flip-flop needs two sets of Int-RDF to record one bit of information, and is too complex to construct. Therefore, new designs need to be tested to boost the development of engineered genetic counters.

Three recombinase-based genetic devices were constructed in this study. A set-reset latch (architecture shown in Fig. 4.1) was first constructed to confirm the application of ϕC31 integrase inversion recombination system to the engineered memory device. The output of this device is a function only of the two input signals. Then, a state-based latch (architecture showing in Fig. 4.4) was constructed to use the current state of the device as a control element for the output state. Based on its design principle, the main construction objective, a binary counting module (architecture shown in Fig. 4.16) was finally developed. It can switch between two states in response to a single repeating

signal and the states can indicate whether an odd number or an even number of input signal has been encountered by the device. Both the state-based latch and the binary counting module were tested on plasmids background and on host chromosome. The results presented in previous chapters are summarised and discussed in this chapter, including the initial characterization of ϕC31 integrase-Gp3 pair, construction of plasmid-borne genetic devices and chromosomal devices, and tests of their performances (Fig. 6.1).

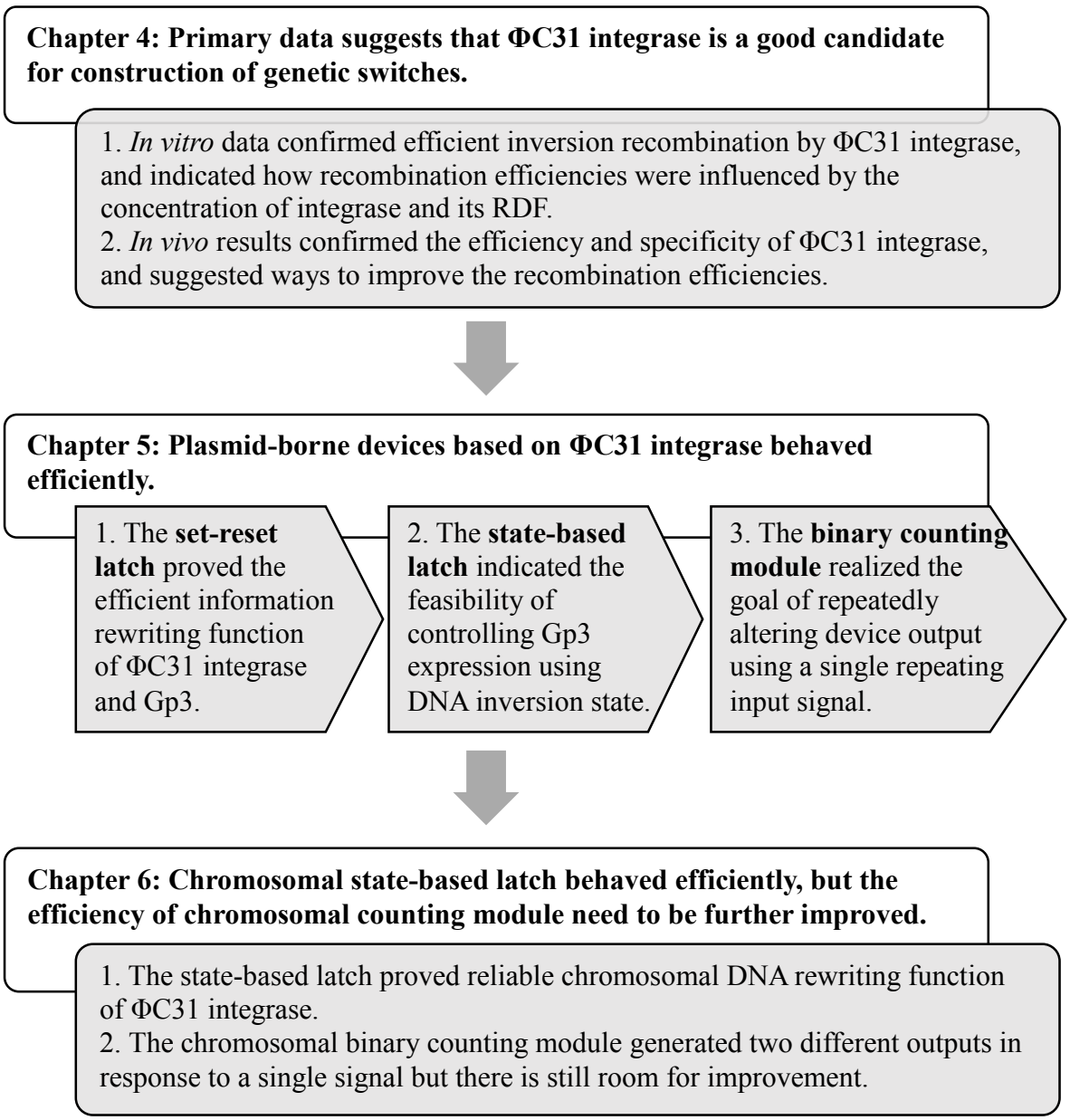


Figure 6. 1 Summary of the results presented in previous result chapters

6.1 Primary data suggested ϕC31 integrase was a good candidate for construction of genetic devices

In chapter 3, *in vitro* reactions showed that both *attB*×*attP* and *attR*×*attL* recombination efficiencies changed as the concentration of ϕC31 integrase and its RDF gp3 were varied. A Gp3 to ϕC31 integrase ratio greater than 1:2 was found necessary to activate *attR*×*attL* recombination and to inhibit *attB*×*attP* recombination (section 3.2.1.2). These *in vitro* data supported the conclusion that Gp3 acts with Int stoichiometrically rather than catalytically or by binding to recombination site during the recombination (Khaleel *et al*., 2011).

The highest efficiency of reaction observed *in vivo* in this study was higher than that observed *in vitro* (section 3.3.3.1). Nearly 100% *attB*×*attP* recombination was achieved *in vivo*, while the highest *attB*×*attP* recombination efficiency was 70% *in vitro*. The *in vivo attR*×*attL* recombination went to 70% completion, and it might be further improved by increasing the ratio of Gp3 to Int, while the highest *attR*×*attL* recombination efficiency was 50% *in vitro*. Possible reasons for the lower recombination efficiency *in vitro* include inactive protein occupying the binding sites on substrate DNA, protein becoming inactive in the reaction buffer over time, or non-optimal recombination condition. To test these hypothesises, experiments could be done in which fresh protein is added after the reaction has reached completion. If the recombination sites are blocked by inactive protein, they may not be recombined by adding new protein. However, if it is because the lack of active proteins, adding new proteins into the reaction may result in more substrate DNA being recombined. To determine the optimal reaction conditions, the effects of parameters, such as pH, temperature, metal ions, and other agents, on Int activity can be investigated (Zhang *et al*., 2008).

The *in vitro* data collected in this study and other existing data (Khaleel *et al*., 2011; Thorpe *et al*., 2000) on the kinetics of recombination catalysed by ϕC31 integrase (with or without Gp3) could be used to help build up a model to further understand the recombination process. Work is ongoing in the lab to build such a mathematical model (Alexandra Pokhilko, unpublished). Efficient models would be necessary in construction of synthetic circuits since it becomes increasingly impractical to test even a small portion of possible component combinations as engineered biological systems become more complex (Purnick and Weiss, 2009). The important role of mathematical modelling has

been described for some of the engineered circuits introduced in chapter 1 (Bonnet *et al*., 2012; Elowitz and Leibler, 2000; Friedland *et al*., 2009; Gardner *et al*., 2000). The construction and optimising process described in chapter 4 also indicated how time consuming optimising such genetic circuits can be. Thus, a mathematical model of the recombination reaction (and devices based on large serine integrases) might help this optimisation process.

During *in vivo* tests of the inversion reactions, several methods were used to tune the parameters of parts in order to improve the recombination efficiencies. The substrate plasmid multimerisation problem was mitigated by adding a *cer* recombination site (section 3.3.2.2), the expression level of Int was regulated by changing its RBSs and start codon (section 3.3.2.4), and the substrate plasmid concentration in cells was adjusted by changing its origin of replication (section 3.3.3). In most cases, positive results were obtained with these methods, and the initial test of inversion recombination mediated by Int showed efficient recombination results, paving the way for the later construction of integrase-based devices.

In addition to ϕC31 integrase, more and more efficient integrases have been discovered and engineered (Xu *et al*., 2013), all these integrases will be potential candidates for construction of efficient and large capacity genetic devices. For instance TG1 integrase, which has recently been tested *in vitro* and showed higher recombination efficiency than ϕC31 integrase (Femi Olorunniji, unpublished results), and 34 new integrases identified using bioinformatics (Yang *et al*., 2014).

6.2 Plasmid-borne devices behaved efficiently

In chapter 4, three genetic devices: 1) the set-reset latch (section 4.2), 2) the state-based latch (section 4.3), and 3) the binary counting module (section 4.4) constructed on a plasmid backbone, were reported. They were engineered to perform efficiently *in vivo*. Here, the performance of these devices are summarized and some possible ways to further improve their efficiencies are discussed.

6.2.1 The set-reset latch proved that ϕC31 integrase could efficiently set and reset information in DNA

The bidirectional inversion recombination mediated by ϕC31 integrase was first tested in an engineered set-reset latch. In this device, the Int expression was controlled by a P_{BAD} promoter (arabinose inducible promoter) and the Int and Gp3 expressions were controlled by a $P_{Ltet0\text{-}1}$ promoter (aTc inducible promoter). The invertible substrate sequence can be set to one direction through *attB*×*attP* recombination with over 75% efficiency, and reset to the opposite direction through *attB*×*attP* recombination with over 70% efficiency by input signals arabinose and aTc, respectively. This is the first time that ϕC31 integrase and its RDF gp3 have been used in this way. More efficient reactions should be obtainable via optimisation of the Int and RDF expression levels and degradation rates. This was not done since the set-reset latch was not the major objective of this study.

6.2.2 The state-based latch demonstrated that expression of RDF could be controlled by the DNA inversion state

After efficient bidirectional inversion recombination was demonstrated by the set-reset latch, a more complex device, the state-based latch, was constructed and optimised. In contrast to the set-reset latch, RDF gp3 expression in the state-based latch is determined not only by an inducible promoter, but also by the state of the invertible substrate sequence. This design was intended as a stepping stone towards a system where Gp3 expression was controlled by the DNA inversion state alone. In this way, the input signal used to activate Int expression would be the only input signal to trigger the whole system.

The state-based latch was initially tested in strain DS941Z1 expressing the Tet repressor and therefore inhibiting expressing Gp3. To achieve an efficient state-based latch, the switch was optimised in multiple steps, including changing gene order and site-directed mutagenesis to prevent read through transcription, and tuning gene expression levels. Finally, this device drove recombination to RL state with 95% efficiency (at DNA level) after the set operation, and back to BP state with 90% efficiency (at DNA level) after the reset operation (section 4.3.3). The state-based latch was the most efficient recombinase addressable device for encoding information into plasmid-borne DNA.

Although using Int and its RDF to set and reset an inversion switch is conceptually simple, a practical demonstration was more difficult than anticipated, since recombination becomes bidirectional in the presence of the RDF. It is thought that free Int inverts DNA by recombining *attB* and *attP*, and the Int-RDF complex inverts DNA by recombining *attR* and *attL* (Purcell and Lu, 2014). To further improve the performance, it might be necessary to find new Int-RDF pairs or engineer the existing pairs to obtain more efficient *attR*×*attL* recombination. In addition, Int-RDF fusion proteins constructed and tested in our group (unpublished results) will potentially mediate more efficient *attR*×*attL* recombination, since no free Int molecule is present to mediate *attP*×*attB* recombination.

6.2.3 The binary counting module realised the goal of altering outputs using a single repeating input signal

In the absence of the Tet repressor, where RDF expression was controlled solely by the DNA inversion state, it was thought that the state-based latch might switch from BP state to RL state and back again in response to repeating pulse of Int expression. However, the state-based latch failed to exhibit this behaviour in the DS941 (*tetR*$^-$) strain. This was thought to be because the short interval between the Int degradation and Gp3 expression. To overcome this problem, a time delay for Gp3 expression was introduced into the system.

By introducing state-based *tetR* expression to delay Gp3 expression, *attB*×*attP* recombination efficiency reached 80%-95%, *attR*×*attL* recombination efficiency reached 60%-90% (at DNA level) under the control of the same input signal (section 4.4.4), and the reliability of this device was proved in a multicycle reaction process. The new device, called the binary counting module, behaved like a single binary digit that can switch between two distinct states in response to a single signal. Such a device was first constructed based on inversion site-specific recombination system of serine integrase. Since using DNA to store information has many advantages (stable, heritable, and easily detectable) (Ham *et al*., 2008) and it is demonstrated that orthogonal serine integrases can be layered to record large amount information (Yang *et al*., 2014), it is reasonable to believe that the integrase-based binary module is a key step towards construction of a powerful biological binary counter.

A limitation of using state-based *tetR* expression as time delay is that production of Tet repressor during the *attR*×*attL* recombination might interfere with Gp3 expression,

preventing efficient *attR*×*attL* recombination. To further improve the performance of the binary counter, other technologies could be tried to extend the interval between Int degradation and Gp3 expression. The Int degradation rate could be increased by addition of a degradation tag, such as the C-terminal peptide tag recognised by tail-specific Tsp proteases (Keiler and Sauer, 1996). The transcription of *gp3* could be delayed by placing one or more genes in front of *gp3*, since the transcription delay for a gene depends upon its distance from the transcription initiation site (Arpino *et al*., 2013; Gotta *et al*., 1991). Another approach to delay the expression of Gp3 at both transcription and translation levels would be to introduce an intein, which can be transcribed and translated together with the target protein (Gogarten *et al*., 2002). The functional protein would be produced after the intein's self-excision and splicing of the target protein exteins. All of these approaches have the risk of changing the Gp3 expression level, and a tuning process would probably be needed to achieve optimal Gp3 expression.

6.3 Making single copy memory devices

The mixed state observed for single cells with plasmid-borne devices was addressed by integrating the invertible sequence onto the host chromosome as a single copy. In strain DS941Z1 (*tetR*$^{+}$), the chromosomal state-based latch was able to drive the recombination to RL state with greater than 95% efficiency and drive the recombination back to BP state with about 80% efficiency at the single cell level, using different input signals (section 5.5).

In strain DS941 (*tetR*$^{-}$), the performance of the chromosomal binary counting module was better than that of the state-based latch when triggered by a single signal (section 5.6). When triggering the chromosomal binary counting module with an optimal length of pulse for different number of times, different proportions of fluorescent cells was achieved, and the highest switch-on efficiency of 60% and switch-off efficiency of 80% was achieved (Fig. 5.7). However, the performance of the chromosomal binary counting module was not as good as the plasmid-borne one, since the *gp3* and *tetR* genes were integrated into the chromosome as a single copy, resulting in the reduction of their expression and recombination efficiencies. It should be possible to further improve recombination efficiencies of these chromosomal devices by optimising the expression level and degradation rates of recombination related proteins (Int, Gp3, and TetR).

Apart from suboptimal recombination efficiency, another possible factor that can lead to incomplete switching behaviour is cell-to-cell variability. Early studies suggest that isogenic microbial populations growing in homogenous conditions generate cell-to-cell differences in physiological parameters such as growth rate, gene expression level, and resistance to stress (Lidstrom and Konopka, 2010). In this study, experiments showed that preculture length prior to addition of inducers affected the recombination behaviours (section 4.4.3). This indicates that cells in different growth phases response to induction differently. Even in an isogenic microbial culture, cells with different growth rates are likely to generate different states when encountering the same pulse of input. Using the chromosomal binary counting module as an example (Fig. 5.6 and Fig. 5.7), a specific length of pulse might not be sufficient to trigger recombination in slow growing cells, but too long so that it triggers unwanted back-switching in fast growing cells, leading to mixed cell populations as observed.

In nature, bacteria use quorum-sensing systems to synchronize particular behaviours on a population-wide scale, and some systems have been well characterised and described (Waters and Bassler, 2005). One of these quorum-sensing systems has been applied in an engineered network to generate synchronized oscillations (Danino *et al*., 2010). Inspired by this, cell-to-cell difference observed in this study might be addressed by introducing a quorum-sensing system to synchronize behaviour throughout the culture.

Given the time-consuming process of chromosome manipulation, the conditionally amplifiable bacterial artificial chromosomes (BACs) could be an effective substitute for the host chromosome to carry the invertible substrate sequences. These conditionally amplifiable BACs are stably maintained as single copy vectors, or give high yield when conditionally induced (Wild *et al*., 2002). Compared to plasmid vectors, BACs can carry extremely large DNA molecules and are stable in cell culture. Compared to the host cell chromosome, BACs are easy to manipulate. However, large BACs are likely to contain sequences that lead to unexpected gene expression or some random homologous recombination (Liu *et al*., 2011). Therefore, whether the BAC-based devices could be used to carry integrase based genetic devices as expected needs to be clarified with practical experiments.

After improving the efficiency of a genetic module based on one kind of integrase, the next step is to test other integrases under the same architecture. In this study, another

integrase (Bxb1) was tested in the state-based latch architecture simply by replacing the ϕC31 integrase gene, its RDF gene, and the *att* sites with Bxb1 equivalents, and keeping all other sequences the same (section 4.3.5). However, the new system did not behave well as expected, and it is likely that regulatory signals will have to be tuned for each different recombinases even under the same architecture. It is also possible that other integrase and RDF genes and *att* sites contain transcriptional signals (promoters) that interfere with designed function. Once more modules using orthogonal recombinases are constructed, they can be layered using the recombinase expressed from an upstream module to mediate recombination of a downstream module, to execute more complex computational functions, such as a multibit ripple counter. One such hypothetical example is illustrated in Figure 6.2, which uses two sets of integrase-RDF to record two bits information, and can count from zero to three. Addition of another integrase-RDF pair would give a three-bit counter that could count up to seven. Results presented in this thesis suggest that a delay on expression of each Int and RDF might be required for proper function of this device.

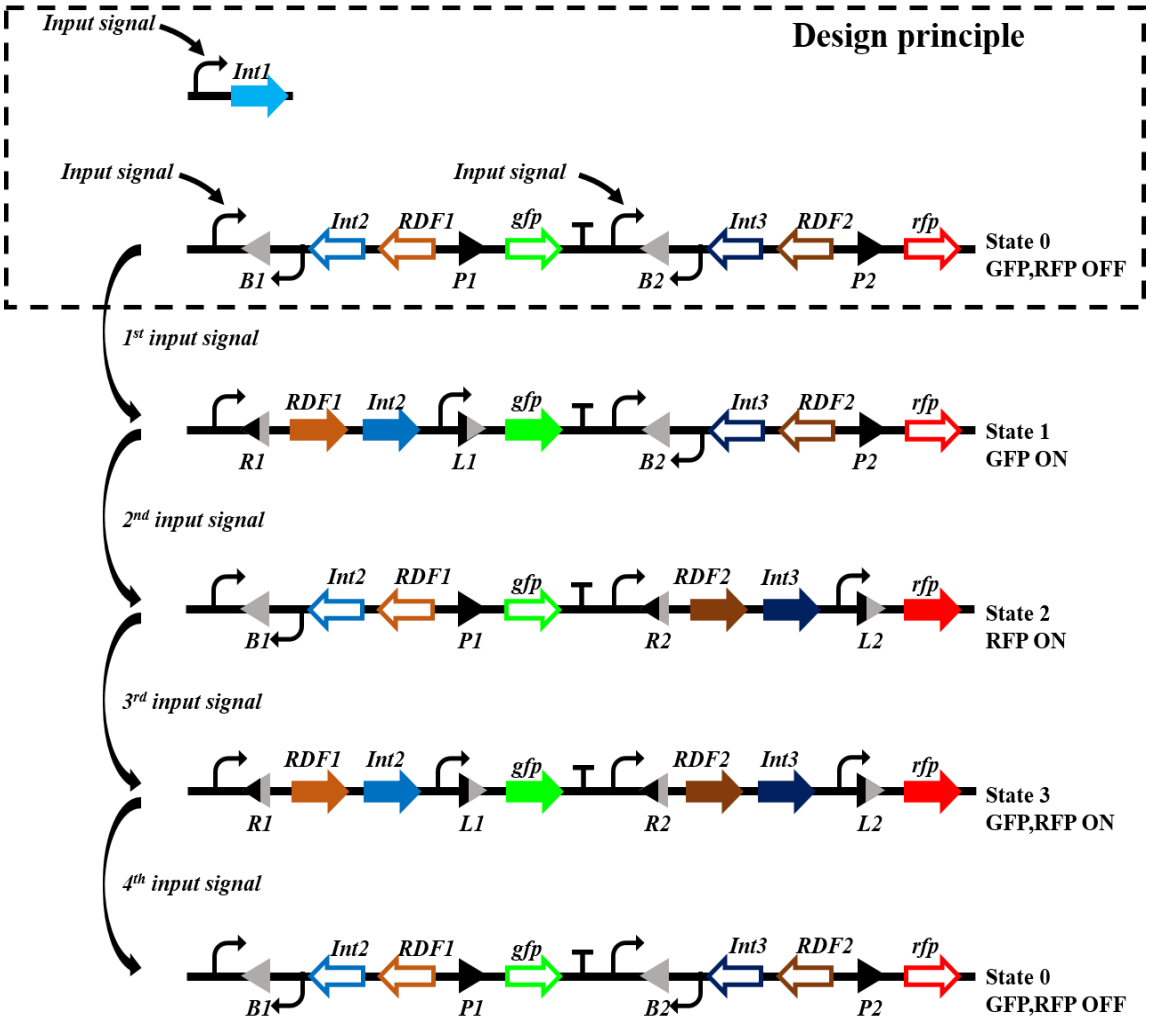


Figure 6. 2 A schematic diagram showing the hypothetical architecture of a two bit binary counter. Different copies of the same inducible promoter are used to control the expression of integrase and RDF genes. The first integrase Int1 is controlled directly by this promoter. Its corresponding RDF (RDF1), the second integrase Int2, and an inverted constitutive promoter are placed between inverted repeat recombination sites for Int1. The inducible promoter and a reporter gene *gfp* are placed outside of the recombination sites. The RDF2, Int3 and an inverted promoter are placed between inverted repeat recombination sites for Int2. The inducible promoter and a reporter gene *rfp* are placed outside of the recombination sites. The first input pulse turns on Int1 expression, resulting in *B1*×*P1* recombination and switching on *gfp* expression. The circuit state is changed from 0 to 1. The second input pulse turns on Int1, RDF1 and Int2 expression, resulting in the *R1*×*L1* and *B2*×*P2* recombination, switching off *gfp* expression and switching on *rfp* expression. The circuit state is changed from 1 to 2. Similarly, the circuit state is changed to 3 and back to 0 with the third and fourth input pulses. By introducing a third module which can be recombined by Int3, a 3-bit device is created that can count up to seven.

Although much work remains to be done, large serine integrases hold promise in the development of memory devices. As one of the exciting memory devices, the synthetic binary counter could be applied in a variety of applications in the future development. For example, by using a cell cycle factor as the input signal, a counter could count the number of times a cell has divided. This operation could be further applied to isolate cells of a specific age, which might be useful in stem cell engineering.

References

Abremski, K., and Gottesman, S. (1982). Purification of the bacteriophage λ xis gene product required for λ excisive recombination. *J. Biol. Chem.*, 257, 9658–9662.

Abremski, K., and Hoess, R. (1984). Bacteriophage P1 site-specific recombination. *J. Biol. Chem.*, 259, 1509–1514.

Ajo-Franklin, C. M., Drubin, D. A., Eskin, J. A., Gee, E. P. S., Landgraf, D., Phillips, I., and Silver, P. A. (2007). Rational design of memory in eukaryotic cells. *Genes Dev.*, 21, 2271–6.

Alén, C., Sherratt, D. J., and Colloms, S. D. (1997). Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination. *EMBO J.*, 16, 5188–5197.

Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P., Givskov, M., and Molin, S. (1998). New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.*, 64, 2240–2246.

Andrianantoandro, E., Basu, S., Karig, D. K., and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, 2, 1–14.

Arpino, J. a J., Hancock, E. J., Anderson, J., Barahona, M., Stan, G.-B. V, Papachristodoulou, A., and Polizzi, K. (2013). Tuning the dials of Synthetic Biology. *Microbiology*, 159, 1236–53.

Atkinson, M. R., Savageau, M. A., Myers, J. T., and Ninfa, A. J. (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell*, 113, 597–607.

Aubel, D., and Fussenegger, M. (2010). Watch the clock-engineering biological systems to be on time. *Curr. Opin. Genet. Dev.*, 20, 634–643.

Austin, S., Ziese, M., and Sternberg, N. (1981). A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell*, 25, 729–736.

Ball, C. a., and Johnson, R. C. (1991). Multiple effects of Fis on integration and the control of lysogeny in phage λ. *J. Bacteriol.*, 173, 4032–4038.

Bibb, L. A., Hancox, M. I., and Hatfull, G. F. (2005). Integration and excision by the large serine recombinase ΦRv1 integrase. *Mol. Microbiol.*, 55, 1896–1910.

Blakely, G., May, G., McCulloch, R., Arciszewska, L. K., Burke, M., Lovett, S. T., and Sherratt, D. J. (1993). Two related recombinases are required for site-specific recombination at dif and cer in *E. coli* K12. *Cell*, 75, 351–361.

Bohn, C., Collier, J., and Bouloc, P. (2004). Dispensable PDZ domain of *Escherichia coli* YaeL essential protease. *Mol. Microbiol.*, 52, 427–435.

Bonnet, J., and Endy, D. (2013). Switches, switches, every where, in any drop we drink.

Mol. Cell, 49, 232–333.

Bonnet, J., Subsoontorn, P., and Endy, D. (2012). Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U. S. A.*, 109, 8884–8889.

Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013). Amplifying genetic logic gates. *Science*, 340, 599–603.

Breüner, A., Brøndsted, L., and Hammer, K. (1999). Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1. *J. Bacteriol.*, 181, 7291–7297.

Brophy, J. A. N., and Voigt, C. A. (2014). Principles of genetic circuit design. *Nat. Methods*, 11, 508–520.

Brown, W. R. A., Lee, N. C. O., Xu, Z., and Smith, M. C. M. (2011). Serine recombinases as tools for genome engineering. *Methods*, 53, 372–379.

Buchholz, F., Angrand, P. O., and Stewart, A. F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.*, 16, 657–662.

Burrill, D. R., and Silver, P. A. (2010). Making cellular memories. *Cell*, 140, 13–18.

Cameron, D. E., and Collins, J. J. (2014). Tunable protein degradation in bacteria. *Nat. Biotechnol.*, 32, 1276–1281.

Canton, B., Labno, A., and Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.*, 26, 787–793.

Caryl, J. A., Smith, M. C. A., and Thomas, C. D. (2004). Reconstitution of a staphylococcal plasmid-protein relaxation complex in vitro. *J. Bacteriol.*, 186, 3374–3383.

Chavez, C. L., Keravala, A., Chu, J. N., Farruggio, A. P., Cuéllar, V. E., Voorberg, J., and Calos, M. P. (2012). Long-Term Expression of Human Coagulation Factor VIII in a Tolerant Mouse Model Using the ΦC31 Integrase System. *Hum. Gene Ther.*, 23, 390–398.

Chen, A. H., Lubkowicz, D., Yeong, V., Chang, R. L., and Silver, P. A. (2015). Transplantability of a circadian clock to a noncircadian organism. *Sci. Adv.*, 1, 1–6.

Chen, Y., Liu, P., Nielsen, A. A. K., Brophy, J. A. N., Clancy, K., Peterson, T., and Voigt, C. A. (2013). Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods*, 10, 659–64.

Chen, Y. Y., Jensen, M. C., and Smolke, C. D. (2010). Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc. Natl. Acad. Sci.*, 8531–8536.

Cheng, A., and Lu, T. K. (2012). Synthetic Biology : An Emerging Engineering Discipline. *Annu. Rev. Biomed. Eng.*, 14, 155–178.

Cheo, D. L., Titus, S. a., Byrd, D. R. N., Hartley, J. L., Temple, G. F., and Brasch, M. a. (2004). Concerted assembly and cloning of multiple DNA segments using in vitro site-specific recombination: Functional analysis of multi-segment expression clones. *Genome Res.*, 14, 2111–2120.

Colloms, S. D., Mcculloch, R., Grant, K., Neilson, L., and Sherratt, D. J. (1996). Xer-mediated site-specific recombination in vitro. *EMBO J.*, 15, 1172–1181.

Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C. M., Osbourn, A., Keasling, J. D., and Rosser, S. J. (2013). Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res.*, 42, 1–10.

Colloms, S. D., Sykora, P., Szatmari, G., and Sherratt, D. J. (1990). Recombination at ColE1 cer requires the *Escherichia coli* xerC gene product, a member of the lambda integrase family of site-specific recombinases. *J. Bacteriol.*

Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B., and Bacchetti, S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.*, 11, 1921–1929.

Craig, N. L. (1988). The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.*, 22, 77–105.

Danino, T., Mondragón-Palomino, O., Tsimring, L., and Hasty, J. (2010). A synchronized quorum of genetic clocks. *Nature*, 463, 326–30.

Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011). Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.*, 39, 1131–1141.

Demarre, G., Guérout, A.-M., Matsumoto-Mashimo, C., Rowe-Magnus, D. A., Marlière, P., and Mazel, D. (2005). A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate *Escherichia coli* host strains. *Res. Microbiol.*, 156, 245–255.

Dewitt, S. K., and Adelberg, E. A. (1962). The Occurrence of a Genetic Transposition in a Strain of Escherichia Coli. *Genetics*, 47, 577–585.

Dudás, A., and Chovanec, M. (2004). DNA double-strand break repair by homologous recombination. *Mutat. Res.*, 566, 131–167.

Ellis, T., Adie, T., and Baldwin, G. S. (2011). DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr. Biol. (Camb).*, 3, 109–118.

Elowitz, M. B., Hsing, W., and Leibler, S. (2002). Combinatorial Synthesis of Genetic Networks. *Science.*, 296, 1466–1470.

Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional

regulators. *Nature*, 403, 335–8.

Elvin, C. M., Thompson, P. R., Argall, M. E., Hendry, P., Stamford, N. P., Lilley, P. E., and Dixon, N. E. (1990). Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene*, 87, 123–126.

Feng, X., and Colloms, S. D. (2007). In vitro transposition of ISY100, a bacterial insertion sequence belonging to the Tc1/mariner family. *Mol. Microbiol.*, 65, 1432–1443.

Ferrières, L., Hémery, G., Nham, T., Guérout, A.-M., Mazel, D., Beloin, C., and Ghigo, J.-M. (2010). Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.*, 192, 6418–27.

Finnegan, D. J. (1992). Transposable elements. *Curr. Opin. Genet. Dev.*, 2, 861–867.

Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M., and Smith, M. C. M. (2014). New applications for phage integrases. *J. Mol. Biol.*, 426, 2703–2716.

Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G., and Collins, J. J. (2009). Synthetic gene networks that count. *Science*, 324, 1199–202.

Fung, E., Wong, W. W., Suen, J. K., Bulter, T., Lee, S., and Liao, J. C. (2005). A synthetic gene-metabolic oscillator. *Nature*, 435, 118–122.

Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. a, Roehner, N., Wilson, M. L., Adam, L., Anderson, J. C., Bartley, B. a, Beal, J., Chandran, D., Chen, J., Densmore, D., Endy, D., Grünberg, R., Hallinan, J., Hillson, N. J., Johnson, J. D., Kuchinsky, A., Lux, M., Misirli, G., Peccoud, J., Plahar, H. a, Sirin, E., Stan, G.-B., Villalobos, A., Wipat, A., Gennari, J. H., Myers, C. J., and Sauro, H. M. (2014). The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat. Biotechnol.*, 32, 545–50.

Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403, 339–342.

Gellert, M., and Nash, H. (1987). Communication between segments of DNA during site-specific recombination. *Nature*, 325, 401 - 404.

Ghosh, P., Kim, A. I., and Hatfull, G. F. (2003). The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol. Cell*, 12, 1101–1111.

Ghosh, P., Wasil, L. R., and Hatfull, G. F. (2006). Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol.*, 4, 0964–0974.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases.

Nat. Methods, 6, 343–345.

Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E. (2002). Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.*, 56, 263–287.

Gotta, S. L., Miller, O. L., and French, S. L. (1991). rRNA transcription rate in *Escherichia coli*. *J. Bacteriol.*

Gregory, M. A., and Smith, M. C. M. (2003). Integration Site for Streptomyces Phage ϕBT1 and Development of Site-Specific Integrating Vectors. *Society*, 185, 5320–5323.

Grindley, N. D. F., Whiteson, K. L., and Rice, P. A. (2006). Mechanisms of site specific recombination. *Annu. Rev. Biochem.*, 75, 567–605.

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

Ham, T., and Lee, S. (2006). A tightly regulated inducible expression system utilizing the fim inversion recombination switch. *Biotechnol. Bioeng.*, 94, 1–4.

Ham, T., Lee, S. K., Keasling, J. D., and Arkin, A. P. (2008). Design and construction of a double inversion recombination switch for heritable sequential genetic memory. *PLoS One*, 3, 9.

Hartley, J. L., Temple, G. F., and Brasch, M. a. (2000). DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Res.*, 1788–1795.

Hirano, N., Muroi, T., Takahashi, H., and Haruki, M. (2011). Site-specific recombinases as tools for heterologous gene integration. *Appl. Microbiol. Biotechnol.*

Hoang, T. T., Kutchma, a J., Becher, a, and Schweizer, H. P. (2000). Integration-proficient plasmids for *Pseudomonas aeruginosa*: site-specific integration and use for engineering of reporter and expression strains. *Plasmid*, 43, 59–72.

Holden, N., Blomfield, I. C., Uhlin, B. E., Totsika, M., Kulasekara, D. H., and Gally, D. L. (2007). Comparative analysis of FimB and FimE recombinase activity. *Microbiology*, 153, 4138–4149.

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genet. Res.*, 89, 282–304.

Hu, G., Goll, M. G., and Fisher, S. (2011, September). ϕC31 integrase mediates efficient cassette exchange in the zebrafish germline. *Dev. Dyn.*,240, 2101-2107.

Inniss, M. C., and Silver, P. A. (2013). Building synthetic memory. *Curr. Biol.*, 23, R812–R816.

Keiler, K. C., and Sauer, R. T. (1996). Sequence determinants of C-terminal substrate recognition by the Tsp protease. *J. Biol. Chem.*, 271, 2589–2593.

Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., and Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.*, 3, 4.

Khaleel, T., Younger, E., McEwan, A. R., Varghese, A. S., and Smith, M. C. M. (2011). A phage protein that binds ϕC31 integrase to switch its directionality. *Mol. Microbiol.*, 80, 1450–1463.

Kirchmaier, S., Lust, K., and Wittbrodt, J. (2013). Golden GATEway cloning-a combinatorial approach to generate fusion and recombination constructs. *PLoS One*, 8, e76117.

Kitney, R., and Freemont, P. (2012). Synthetic biology – the state of play. *FEBS Lett.*

Knight, T. (2003). Idempotent Vector Design for Standard Assembly of Biobricks Idempotent Vector Design for Standard Assembly of Biobricks. *Structure*, 1–11.

Kotula, J. W., Kerns, S. J., Shaket, L. a, Siraj, L., Collins, J. J., Way, J. C., and Silver, P. a. (2014). Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1321321111

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.*, 58, 401–465.

Kuhstoss, S., and Rao, R. N. (1991). Analysis of the integration function of the streptomycete bacteriophage ϕC31. *J. Mol. Biol.*, 222, 897–908.

Kuhstoss, S., Richardson, M. A., and Rao, R. N. (1991). Plasmid cloning vectors that integrate site-specifically in *Streptomyces* spp. *Gene*, 97, 143–146.

Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996). A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J.*, 15, 5470–5479.

Landy, A. (1989). Dynamic, Structural, and Regulatory Aspects of λ Site-Specific Recombination. *Annu. Rev. Biochem.*, 913–949.

Leypold, C. F., Marian, D., Roman, C., Schneider, S., Schubert, P., Scholz, O., Hillen, W., Clark, T., and Lanig, H. (2004). How does Mg^{2+} affect the binding of anhydrotetracycline in the TetR protein? *Photochem. Photobiol. Sci.*, 3, 109–119.

Li, M. Z., and Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods*, 4, 251–256.

Lidstrom, M. E., and Konopka, M. C. (2010). The role of physiological heterogeneity in microbial population behavior. *Nat. Chem. Biol.*, 6, 705–712.

Lou, C., Liu, X., Ni, M., Huang, Y., Huang, Q., Huang, L., Jiang, L., Lu, D., Wang, M., Liu, C., Chen, D., Chen, C., Chen, X., Yang, L., Ma, H., Chen, J., and Ouyang, Q. (2010). Synthesizing a novel genetic sequential logic circuit: a push-on push-off

switch. *Mol. Syst. Biol.*, 6, 350.

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.*, 25, 1203–1210.

Marsischky, G., and LaBaer, J. (2004). Many paths to many clones: A comparative look at high-throughput cloning methods. *Genome Res.*, 14, 2020–2028.

Martín, C., Mazodier, P., Mediola, M. V, Gicquel, B., Smokvina, T., Thompson, C. J., and Davies, J. (1991). Site-specific integration of the *Streptomyces* plasmid pSAM2 in *Mycobacterium smegmatis*. *Mol. Microbiol.*, 5, 2499–2502.

McCulloch, R., Burke, M. E., and Sherratt, D. J. (1994). Peptidase activity of *Escherichia coli* aminopeptidase A is not required for its role in Xer site-specific recombination. *Mol. Microbiol.*, 12, 241–251.

Meselson, M. S., and Radding, C. M. (1975). A general model for genetic recombination. *Proc. Natl. Acad. Sci. U. S. A.*, 72, 358–361.

Miyazaki, R., and van der Meer, J. R. (2013). A new large-DNA-fragment delivery system based on integrase activity from an integrative and conjugative element. *Appl. Environ. Microbiol.*, 79, 4440–4447.

Moon, T. S., Clarke, E. J., Groban, E. S., Tamsir, A., Clark, R. M., Eames, M., Kortemme, T., and Voigt, C. a. (2011). Construction of a genetic multiplexer to toggle between chemosensory pathways in *Escherichia coli*. *J. Mol. Biol.*, 406, 215–27.

Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012). Genetic programs constructed from layered logic gates in single cells. *Nature*, 000, 1–5.

Morita, K., Yamamoto, T., Fusada, N., Komatsu, M., Ikeda, H., Hirano, N., and Takahashi, H. (2009). In vitro characterization of the site-specific recombination system based on actinophage TG1 integrase. *Mol. Genet. Genomics*, 282, 607–616.

Muñoz-López, M., and García-Pérez, J. L. (2010). DNA transposons: nature and applications in genomics. *Curr. Genomics*, 11, 115–128.

Nafissi, N., and Slavcev, R. (2014). Bacteriophage recombination systems and biotechnical applications. *Appl. Microbiol. Biotechnol.*, 98, 2841–851.

Nissim, L., and Bar-Ziv, R. H. (2010) A tunable dual-promoter integrator for targeting of cancer cells. *Mol Syst Biol.*, 6,444

Podhajska, a J., Hasan, N., and Szybalski, W. (1985). Control of cloned gene expression by promoter inversion in vivo: construction of the heat-pulse-activated att-nutL-p-att-N module. *Gene*, 40, 163–168.

Ptashne, M. (2004). A genetic switch : phage lambda revisited. *Cold Spring Harbor Laboratory Press Cold Spring Harbor NY* (Vol. 3).

Purcell, O., and Lu, T. K. (2014). Synthetic analog and digital circuits for cellular

computation and memory. *Curr. Opin. Biotechnol.*, 29, 146–155.

Purnick, P. E. M., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.*, 10, 410–22.

Pyne, M. E., Moo-Young, M., Chung, D. a., and Chou, C. P. (2015). Coupling the CRISPR/Cas9 system to lambda Red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Appl. Environ. Microbiol.*, AEM.01248–15.

Reddy, P., Peterkofsky, a, and McKenney, K. (1985). Translational efficiency of the *Escherichia coli* adenylate cyclase gene: mutating the UUG initiation codon to GUG or AUG results in increased gene expression. *Proc. Natl. Acad. Sci. U. S. A.*, 82, 5656–5660.

Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.*, 26, 51–56.

Roberts, M. a J., Cranenburgh, R. M., Stevens, M. P., and Oyston, P. C. F. (2013). Synthetic biology: Biology by design. *Microbiol. (United Kingdom)*, 159, 1219–1220.

Rowley, P. A., Smith, M. C. A., Younger, E., and Smith, M. C. M. (2008). A motif in the C-terminal domain of ϕC31 integrase controls the directionality of recombination. *Nucleic Acids Res.*, 36, 3879–3891.

Rutherford, K., and Van Duyne, G. D. (2014). The ins and outs of serine integrase site-specific recombination. *Curr. Opin. Struct. Biol.*, 24, 125–31.

Rutherford, K., Yuan, P., Perry, K., Sharp, R., and Van Duyne, G. D. (2013). Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res.*, 41, 8341–56.

Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, 27, 946–50.

Sauer, B. (1998). Inducible gene targeting in mice using the Cre/lox system. *Methods*, 14, 381–392.

Shao, Z., Zhao, H., and Zhao, H. (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.*, 37, 1–10.

Siuti, P., Yazbek, J., and Lu, T. K. (2013). Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.*, 31, 448–52.

Skerra, A. (1994). Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene*, 151, 131–135.

Smith, M. C. a, Till, R., Brady, K., Soultanas, P., Thorpe, H., and Smith, M. C. M. (2004). Synapsis and DNA cleavage in ϕC31 integrase-mediated site-specific recombination. *Nucleic Acids Res.*, 32, 2607–17.

Smith, M. C. M., Brown, W. R. a, McEwan, A. R., and Rowley, P. A. (2010). Site-specific recombination by ϕC31 integrase and other large serine recombinases. *Biochem. Soc. Trans.*, 38, 388–94.

Smith, M., and Thorpe, H. (2002). Diversity in the serine recombinases. *Mol. Microbiol.*, 44, 299–307.

Smolke, C. D. (2009). Cell biology. It's the DNA that counts. *Science*, 324, 1156–1157.

Stanton, B. C., Nielsen, A. A. K., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. A. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.*, 10, 99–105.

Stark, W. M., Boocock, M., and Sherratt, D. (1992). Catalysis by site-specific recombinases. *Trends Genet.*, 8, 432–439.

Stirling, C. J., Colloms, S. D., Collins, J. F., Szatmari, G., and Sherratt, D. J. (1989). xerB, an *Escherichia coli* gene required for plasmid ColE1 site-specific recombination, is identical to pepA, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. *EMBO J.*, 8, 1623–1627.

Stirling, C. J., Stewart, G., and Sherratt, D. J. (1988). Multicopy plasmid stability in *Escherichia coli* requires host-encoded functions that lead to plasmid site-specific recombination. *Mol. Gen. Genet.* 8, 2–6.

St-Pierre, F., Cui, L., Priest, D. G., Endy, D., Dodd, I. B., and Shearwin, K. E. (2013). One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.*, 2, 537–541.

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature*, 456, 516–9.

Subsoontorn, P., and Endy, D. (2012). Design and Analysis of Genetically Encoded Counters. *Procedia Comput. Sci.*

Summers, D. K., and Sherratt, D. J. (1984). Multimerization of high copy number plasmids causes instability: CoIE1 encodes a determinant essential for plasmid monomerization and stability. *Cell*, 36, 1097–1103.

Summers, D. K., and Sherratt, D. J. (1988). Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the cer site. *EMBO J.*, 7, 851–858.

Thomas, C. M., and Smith, C. A. (1987). Incompatibility group P plasmids: genetics, evolution, and use in genetic manipulation. *Annu. Rev. Microbiol.*, 41, 77–101.

Thorpe, H. M., and Smith, M. C. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U. S. A.*, 95, 5505–5510.

Thorpe, H. M., Wilson, S. E., and Smith, M. C. (2000). Control of directionality in the

site-specific recombination system of the *Streptomyces* phage ϕC31. *Mol. Microbiol.*, 38, 232–241.

Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S., and Calos, M. P. (2001). Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage ϕC31 Integrase. *Mol. Cell. Biol.*, 21, 3926–3934.

Urasaki, A., Sekine, Y., and Ohtsubo, E. (2002). Transposition of cyanobacterium insertion element ISY100 in *Escherichia coli. J. Bacteriol.*, 184, 5104–5112.

van de Putte, P., and Goosen, N. (1992). DNA inversions in phages and bacteria. *Trends Genet.*, 8, 457–462.

Van Duyne, G. D., and Rutherford, K. (2013). Large serine recombinase domain structure and attachment site binding. *Crit. Rev. Biochem. Mol. Biol.*, 48, 476–91.

Venken, K. J. T., He, Y., Hoskins, R. a, and Bellen, H. J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science*, 314, 1747–1751.

Wang, B., Kitney, R. I., Joly, N., and Buck, M. (2011). Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat. Commun.*, 2, 1–9.

Waters, C. M., and Bassler, B. L. (2005). Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.*, 21, 319–346.

Wild, J., Hradecna, Z., and Szygbalski, W. (2002). Conditionally amplifiable BACs: Switching from single-copy to high-copy vectors and genomic clones. *Genome Res.*, 12, 1434–1444.

Win, M. N., and Smolke, C. D. (2008). Higher-order cellular information processing with synthetic RNA devices. *Science*, 322, 456–460.

Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M., and Brown, W. (2013). Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol.*, 13, 1–17.

Yang, L., Nielsen, A. A. K., Fernandez-Rodriguez, J., McClune, C. J., Laub, M. T., Lu, T. K., and Voigt, C. A. (2014). Permanent genetic memory with >1-byte capacity. *Nat. Methods*, 11, 1261–1266.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33, 103–19.

Ye, L., Chang, J. C., Lin, C., Qi, Z., Yu, J., and Kan, Y. W. (2010). Generation of induced pluripotent stem cells using site-specific integration with phage integrase. *Proc. Natl. Acad. Sci. U. S. A.*, 107, 19467–19472.

Zhang, L., Ou, X., Zhao, G., and Ding, X. (2008). Highly efficient in vitro site-specific recombination system based on *Streptomyces* phage Highly efficient in vitro site-specific recombination system based on Streptomyces ϕBT1 integrase. *J. Bacteriol.*

Zhang, L., Zhao, G., and Ding, X. (2011). Tandem assembly of the epothilone biosynthetic gene cluster by in vitro site-specific recombination. *Sci. Rep.*, 1, 1–5.

Zhang, L., Zhu, B., Dai, R., Zhao, G., and Ding, X. (2013). Control of directionality in *Streptomyces* phage ϕBT1 integrase-mediated site-specific recombination. *PLoS One*, 8, 1–8.

Zuo, Y., and Deutscher, M. P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.*, 29, 1017–1026.