



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Synthetic Logic Circuits encoded on
Toehold Strand-Displacement Switchable
CRISPR guide RNAs.**



Pascoe James Harvey

Thesis presented for the degree of Doctor of Philosophy

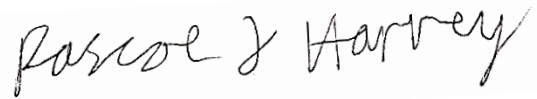
Institute of Quantitative Biology, Biochemistry and Biotechnology

The University of Edinburgh

November 2018

Declaration

I declare that this thesis was composed by myself and that the work presented is my own, except where otherwise stated. This thesis has not been submitted for any other degree or professional qualification.

A handwritten signature in black ink that reads "Pascoe J Harvey". The signature is written in a cursive style with a large initial 'P' and 'H'.

Pascoe J Harvey

Contents

Declaration	i
Table of figures	vi
Abbreviations	viii
Acknowledgements	xi
Abstract	xii
Lay abstract	xiv
Chapter 1: Introduction	1
1.1 RNA folding and secondary structure	1
1.1.1 Structure from comparative sequence analysis	1
1.1.2 Lowest free energy folding prediction	3
1.1.3 Ensemble and RNA-RNA interaction prediction	3
1.1.4 Software packages	4
1.2 Logic in biological systems	5
1.2.1 History of the Boolean Logic Gate	9
1.2.2 Signal transfer in logic gates	9
1.2.3 Latency	12
1.3 CRISPR Cas9 system	14
1.3.1 Modifications of the gRNA	18
1.4 RNA synthetic biology	26
1.4.1 Toehold switches	26
1.5 Objectives for constructing a NAND gate	30
1.6 Targets for inactivation within the gRNA.....	30
1.7 RNA-RNA interaction initiation	35
Chapter 2: Construction & characterisation of crgRNA	37
2.1 Introduction	37
2.1.1 Overview of an RNA-mediated NAND gate	39
2.2 Results	45
2.2.1 Experimental setup	45
2.2.2 Selecting target for <i>cis</i> -repressing element	47
2.2.3 Design and synthesis of crgRNA.....	48

2.2.4	Initial testing of crgRNA variant K	52
2.2.5	Optimization of dCas9 and crgRNA expression levels	54
2.2.6	Testing crgRNA variants K, W and Y with optimized dCas9 and crgRNA expression levels	58
2.2.7	Testing dCas9 dependency of reporter repression.....	61
2.2.8	Optimization of induction levels for crgRNA Y variant	63
2.2.9	Effect of Temperature on repression.....	65
2.2.10	Time series measurements	67
2.2.11	Engineering the dCas9 Expressing Plasmid for Lower Expression	73
2.3	Discussion.....	81
2.3.1	Proof of Concept	81
2.3.2	Methodological Insights.....	81
2.3.3	dCas9 expression.....	82
2.3.4	<i>Cis</i> -repressing elements at the 5' and 3' terminus	82
2.3.5	Design insight	84
2.3.6	Characterisation	84
Chapter 3: Optimisation of crgRNA and orthogonality		87
3.1	Introduction	87
3.2	Results	91
3.2.1	Library Design.....	91
3.2.2	<i>In vivo</i> Library testing	94
3.2.3	Toehold orthogonality	101
3.2.4	Stage 1: Toehold generation.....	106
3.2.5	Stage 2: Toehold ranking and screening.....	106
3.2.6	Stage 3: Pairwise Interaction Matrix.....	107
3.2.7	Toehold set optimisation	112
3.2.8	Orthogonality index over time.....	114
3.2.9	Result verification	117
3.2.10	Testing universalizability of Lp crgRNA	120
3.2.11	<i>In vivo</i> Orthogonality.....	120
3.2.12	Multiple crgRNA targets.....	124
3.2.13	Control Experiments	126

3.3	Discussion.....	128
3.3.1	No recruitment of dCas9 by asRNA	128
3.3.2	<i>Cis</i> -repressing element mechanism of inactivation.....	128
3.3.3	Orthogonality Algorithm	129
3.3.4	Observed orthogonality <i>in vivo</i>	130
3.3.5	Application of system to multiple genes	131
Chapter 4: Discussion.....		132
4.1	Summary	132
4.2	Future work.....	134
4.2.1	Logic circuit layering.....	134
4.2.2	Reduced background repression	138
4.2.3	Potential network size.....	138
4.2.4	Future work: autoregulation.....	139
4.3	Contrast with other systems	140
4.3.1	Contrast with a gRNA antisense system	140
4.3.2	Contrast with riboswitched gRNAs: study I	144
4.3.3	Contrast with riboswitched gRNAs: study II	146
Chapter 5: Methods and Materials.....		149
5.1	Materials	149
5.1.1	List of plasmids.....	152
5.1.2	<i>E. coli</i> strains.....	153
5.2	Molecular Biology Methods.....	153
5.2.1	Storage of <i>E. coli</i> strains	153
5.2.2	Preparation of chemically competent cells	154
5.2.3	Transformation into <i>E. coli</i>	154
5.2.4	Plasmid preparation.....	155
5.2.5	Polymerase Chain Reaction (PCR).....	155
5.2.6	Restriction digest	156
5.2.7	Agarose gel electrophoresis.....	156
5.2.8	Gel extraction.....	156
5.2.9	Plasmid construction.....	156
5.3	Data Gathering Methods.....	157

5.3.1	End point reading.....	157
5.3.2	Time course.....	158
5.4	Data Processing.....	158
5.4.1	Controlling for row bias	158
5.4.2	Propagation of noise	159
5.4.3	Normalisation.....	159
Appendix 1: Toehold orthogonality script.....		161
Appendix 2: Nucleotide sequences.....		187
	Sequences from results chapter 1	187
	5' variants.....	187
	3' variants.....	188
	Sequences from results chapter 2	190
	Variant library and toehold change	190
	Orthogonality	192
	Multiple targets.....	193
	dCas9 expression.....	194
References		197

Table of figures

Figure 1.1.....	2
Figure 1.2.....	7
Figure 1.3.....	8
Figure 1.4	11
Figure 1.5.....	13
Figure 1.6.....	15
Figure 1.7.....	22
Figure 1.8.....	25
Figure 1.9.....	29
Figure 1.10.....	34
Figure 1.11	36
Figure 2.1.....	43
Figure 2.1.....	44
Figure 2.2.....	46
Figure 2.3.....	51
Figure 2.4.....	53
Figure 2.5.....	56
Figure 2.5.....	57
Figure 2.6.....	59
Figure 2.7.....	62
Figure 2.8.....	64
Figure 2.9.....	66
Figure 2.10.....	69

Figure 2.11	71
Figure 2.11.....	72
Figure 2.12.....	74
Figure 2.13.....	76
Figure 2.14.....	79
Figure 2.14.....	80
Figure 3.1	89
Figure 3.1.....	90
Figure 3.2.....	93
Figure 3.3.....	96
Figure 3.4.....	98
Figure 3.5.....	100
Figure 3.6.....	103
Figure 3.7.....	105
Figure 3.8	109
Figure 3.9.....	111
Figure 3.10.....	113
Figure 3.11.....	116
Figure 3.12.....	119
Figure 3.13.....	122
Figure 3.13.....	123
Figure 3.14.....	125
Figure 3.15.....	127
Figure 4.1.....	137
Figure 4.2.....	143

Abbreviations

ANOVA	Analysis of variance
asRNA	Anti-sense RNA
aTc	Anhydrotetracycline
Cas9	CRISPR associated protein 9
CFP	Cyan Fluorescent Protein
crgRNA	<i>Cis</i> -repressed guide RNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
dCas9	Catalytically dead Cas9
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
GFP	Green Fluorescent Protein
GC	Guanine, Cytosine
gRNA	Guide RNA
HIV	Human immunodeficiency virus

IEEE	Institute of Electrical and Electronics Engineers
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Lysogeny broth
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger RNA
NAND	Not AND
NUPACK	Nucleic Acid Package
OD	Optical Density
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
RBS	Ribosome Binding Site
RCF	Relative centrifugal force
RNA	Ribonucleic acid
RNase	Ribonuclease
SD	Standard deviation
SELEX	Systematic Evolution of Ligands by EXponential enrichment
SOC	Super Optimal broth with Catabolite repression
taRNA	<i>trans</i> -activating RNA
TAE	Tris(hydroxymethyl)aminomethane, acetic acid, EDTA

TALE	Transcription activator-like effector
TE	Tris(hydroxymethyl)aminomethane, EDTA
tracrRNA	<i>Trans</i> -activating crRNA
TSS	Transcription Start Site
UTR	Untranslated region
YFP	Yellow Fluorescent Protein
ZF	Zink Finger

Acknowledgements

I'd like to thank Clare Hobba and Nigel Harvey, not just for their guidance and encouragement during my PhD and thesis writing but also for equipping me with the skills and outlook needed to approach it.

I'd like to thank Susan Rosser, Emily Johnston, Tessa Moses, Christine Merrick, Dirk-Jan Kleinjan, Maryia Trubitsyna, James Bryson, Matthew Dale, Jamie Auxillos and Anita Wilkinson and the rest of the Rosser Lab for their contributions, support, and advice both in the writing of this thesis and over the course of my PhD.

I'd also like to thank you, reader, for taking the time to read any part of this thesis.

Abstract

The field of biological computing offers the potential to construct devices using complex and dynamic regulation for applications ranging from theranostics to the production of high value chemicals in bioreactors. However, building these complex regulatory systems depends on the creation of effective logic gates, which form the basis of digital systems. The low metabolic load, small genetic footprint and the low latency of expression that can be achieved with a small RNA-based regulatory system in contrast to protein regulators, emphasises the potential within RNA regulation. From an engineering perspective, the great advantage of an RNA approach is the highly predictable nature of RNA folding and the availability of established *in silico* tools.

This thesis describes a novel *de novo* mechanism for NAND gate implementation *in vivo* using two RNAs, both of which must be expressed for the repression of an output gene. To construct this regulatory system a guide RNA of the CRISPR-Cas9 system is modified through the addition of a *cis*-repressing element, which complements part of the guide RNA and represses its activity. The activity of the *cis*-repressed guide RNA (crgRNA) can be rescued by the expression of an antisense RNA, which complements the *cis*-repressing element. This allows the guide RNA to return to an active conformation and repress the target promoter through CRISPRi (in strains expressing dCas9). This represents a NAND gate, as the output is repressed (OFF) only when both input RNAs are expressed. The design and optimisation of this system was performed using modelling of system energy states and dynamics and machine learning optimisation in a process which was automated into a single pipeline for future users. This system was characterised over a range of crgRNA and dCas9 expression levels and temperatures, and in different growth phases. Eight designs were tested and the optimal variant, for which output gene expression most closely approximated the OFF (repressed) and ON (un-repressed) states required for a logic gate, was chosen. The resulting NAND gate has a 10-fold repression of the output promoter when both RNAs were present; in contrast, only

1.2 fold repression was obtained when only the crgRNA was expressed. Consequently, multiple versions of the optimal variant were synthesised, each with different sequences but the same design principles. These performed similarly when applied to the repression of different reporter genes. Finally, an *in silico* approach was used to maximise orthogonality of different versions of the optimal variant which was then demonstrated *in vivo*. This novel NAND gate design offers the ability to build large libraries of logic gates with small genetic footprints (304 bp) and the potential to be combined to produce complex regulatory networks.

Lay abstract

Complex machines require control systems or 'logics'; from elevators to cars, control systems are most often made using logic gates which make an output change depending on two inputs. This is usually in the form of electronic transistors. The rise of synthetic biology as a field has led to the need for control systems for biological machines. To this end, this thesis presents a new design of a biological logic gate and demonstrates its functionality.

The logic gate exploits an already established system in which a protein called dCas9 can be guided by an RNA known as a guide RNA to a DNA target where it represses or turns OFF the output gene inside the cell. In the system presented here the guide RNA, which is used to target the repression of the dCas9 to a particular gene, was engineered through the addition of an extra sequence that folds into the guide RNA and stops it repressing the output gene. The other main component of the logic gate is a second RNA which binds to the extra sequence preventing it from folding in to the guide RNA allowing repression of the output gene. This means for the target output gene to be repressed both input RNAs need to be expressed. This forms a NAND gate. After testing a number of designs, the optimal design variant was selected. This variant demonstrated 10-fold repression as an OFF output and between 0 and 19% repression as an ON output. Different versions of this variant were created which were able to repress different output genes in the same manner. A computational approach was taken to creating versions which were able to operate in the same cell without interacting with one another. This low level of crosstalk was demonstrated within living cells. This novel NAND gate design offers the ability to build a large number of logic gates with each gate using only a small amount of the genome. It also has the potential to be layered in to complex regulatory networks.

Chapter 1: Introduction

1.1 RNA folding and secondary structure

RNA structure is defined on three main levels: primary structure, secondary structure and tertiary structure. The primary structure is the linear sequence of nucleotides which the RNA molecule is made up of in order: this is principally defined by the template DNA from which the RNA is transcribed. The secondary structure is formed by the interactions between bases within the RNA sequence. These interactions, principally hydrogen bonds, allow the sequence to form secondary structures such as the stem and loop. Tertiary structure describes the position of each of the atoms within the structure in three dimensional space and is able to take into account geometric and spatial constraints.

1.1.1 Structure from comparative sequence analysis

When studying structure and folding in nature, for both RNA and proteins, comparative sequence analysis has been one of the most powerful techniques for identifying which nucleotides or amino acids contact with others. In this technique, the structure of a subject, for example a ribosomal RNA, is inferred by sampling a wide pool of homologues containing sufficient variation and aligning them. Where there is a strong correlation between mutations in one position and mutations in another position it is an indication that those two points in the primary structure are in contact with one another in the secondary or tertiary structure. Where the two positions are complementary (for example a G and a C) and this complementarity is maintained independent of mutations this indicates that those two parts of the sequence complement one another.

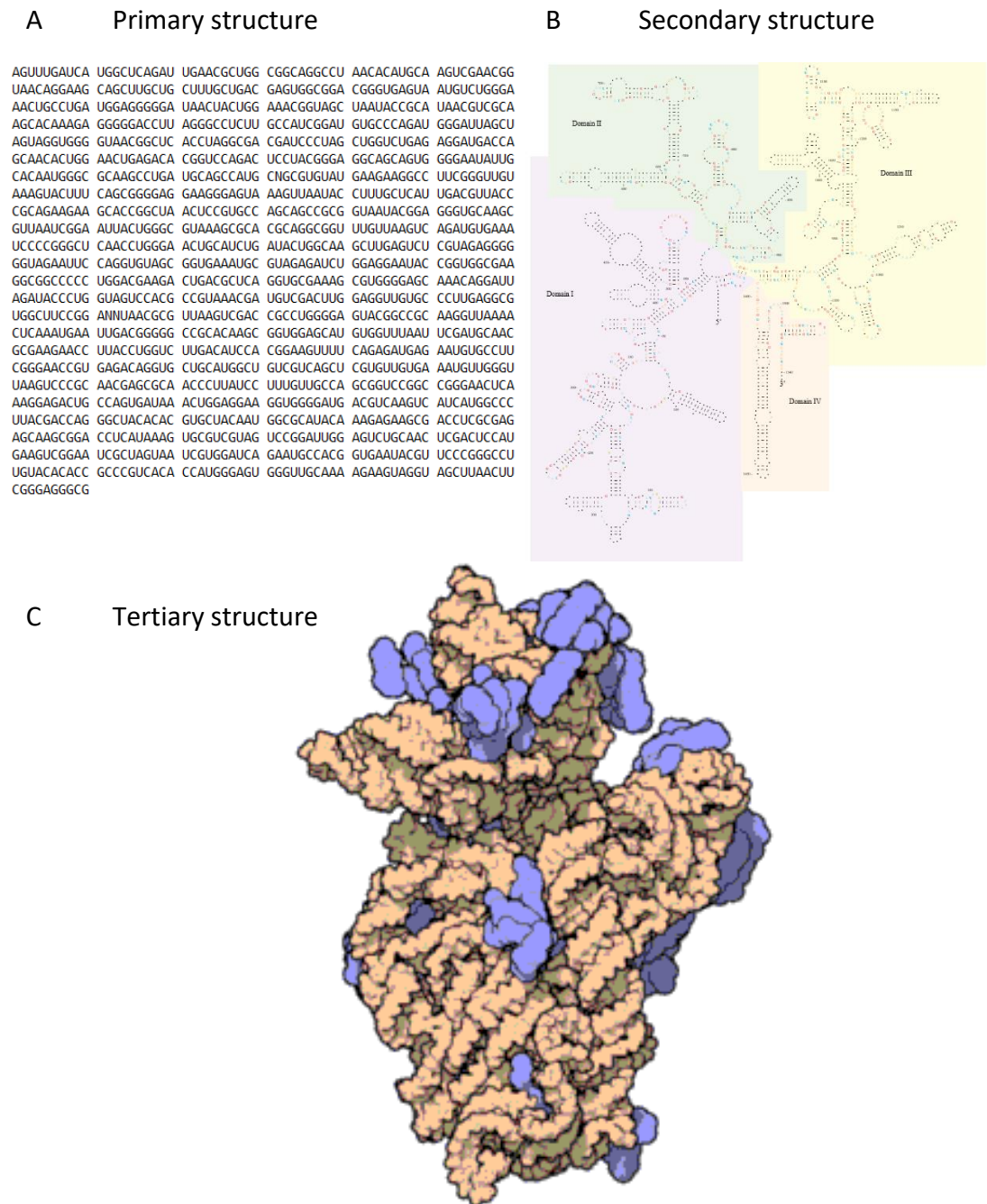


Figure 1.1 The 16S ribosomal RNA given here as an example of (A) primary structure, (B) secondary structure and (C) tertiary structure of RNA (tertiary structure includes additional poly peptides in purple) . (B,C) are reproduced under a Creative Commons Licence (CC3.0) (David S. Goodsell, 2012)

1.1.2 Lowest free energy folding prediction

Comparative sequence analysis has proved a powerful technique when investigating naturally occurring RNA structure and function but for the synthetic biologist sequence design is important and therefore *de novo* folding prediction becomes important. Most folding prediction methods have at their base a thermodynamic model derived from calorimetry (Mathews et al., 2004). This thermodynamic model can be used to predict the free energy of any arbitrary RNA structure. The most obvious approach to structure prediction would be to take an RNA sequence and generate every possible structure for that sequence then find the structure with the lowest free energy. However, this brute force approach takes too much computational time, a sequence of 100 nucleotides has more than 10^{25} possible secondary structures (Mathews, 2006). Instead a dynamic programming approach using a nearest neighbour technique is utilised. In this approach lowest possible conformational free energy is identified for sub-sequences which are increased in length from the centre (Zuker & Stiegler, 1981). While this is the most commonly used general approach to RNA secondary structure prediction, it has downsides such as the inability to predict pseudoknots. More recent approaches have been developed to include pseudoknot prediction (Reeder & Giegerich, 2004; Rivas & Eddy, 1999).

1.1.3 Ensemble and RNA-RNA interaction prediction

The RNA folding prediction approach described so far has been able to find the secondary structure with the lowest free energy for a given RNA sequence. However, in the cellular environment, RNA does not exist in a single structure but in an ensemble of structures in which the most common structure may not be that structure with the lowest free energy. Understanding this level of complexity is particularly important when working with RNA which may have two major structures with different functional properties such as in the case of a riboswitch (described below). Calculating the partition function of the ensemble also allows the calculation of base pairing probabilities and melting temperatures. These are

important when predicting how a multi-state RNA mechanism design is likely to function (Hofacker et al., 1994; McCaskill, 1990). When measuring the effect of a change in sequence on structure, changes in structure can be measured either in terms of a change in the pairing probabilities which gives granular indications as to which nucleotide pairs have changed comparing probability, or alternatively, it can be measured in terms of a change in the free energy of the ensemble which gives a single number indicating whether the changes resulted in a lower free energy ensemble (more stable, more bonds).

When considering the interaction of two RNAs, it is possible to treat the problem in a similar manner to finding a structure for a single RNA; in this approach, the two RNAs are combined and folded together. When this approach is taken, similar drawbacks are seen to predict a single structure, due to the nested nature of interactions, some interactions are excluded from predictions. One example of an interaction excluded from predictions being the kissing loop interactions where two loop sequences, each enclosed by a stem complement one another (Bernhart et al., 2006).

1.1.4 Software packages

There are a number of software packages which combine techniques for RNA structure prediction into a single unit which can be used by investigators when studying naturally occurring RNA structures or engineering new ones. The original ViennaRNA package contained a set of basic tools for predicting minimum free energy structures or partition functions of RNA sequences (Hofacker et al., 1994). The version 2.0 uses an updated energy model and an expansion in the range of tools available, specifically including tools for RNA-RNA interaction such as RNAcofold (Mathews et al., 2004). Also, increasing the range of outputs to include centroid structures and maximum expected accuracy structures as well as allowing input such as FASTA format. Each of these tools are available either on a publicly accessible server or as stand-alone routines which can easily be integrated into user scripts (Gruber, Lorenz, Bernhart, Neubock, & Hofacker, 2008; Hofacker, 2003). The

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Nucleic Acid Package (NUPACK) is a growing software suite for analysis of nucleic acid systems which expands on ViennaRNA 2.0 applications by the inclusion of design tools (Zadeh et al., 2011). Similarly to the Vienna 2.0 package, the suite can either be used on the NUPACK web server, or alternatively, it can be downloaded and compiled for stand-alone use.

1.2 Logic in biological systems

Logic gates take a number of binary inputs (usually two) and perform a logical operation to produce a single binary output (Boole, 1854). Most technological applications of logic gates are found in electronic circuitry and form the basis of modern day computing, utilising the outputs of specific gates as input into subsequent gates to form a logical network. Logic gates also feature in control systems for many machines which are not computers, such as elevators, automobiles, washing machines, etc. [†]

The Boolean logic gates are: NOT, AND, OR, NAND, NOR, XOR, XNOR and IMPLY (Figure 1.2). All of these except the NOT gate have two inputs and a single output, however they vary in the logical operation which is performed on the inputs to reach the output. The output is defined by the truth table, which takes all combinations of inputs, and for each possible combination defines which output the gate produces: it is the truth table which defines the logic gate. In Figure 1.2, the standard Boolean logic gates are shown with corresponding truth tables and symbols.

The logic gates can be broken into two categories: (i) functionally complete, and (ii) non-functionally complete. Functionally complete gates are able to perform any logical operation by combining multiple gates. The two functionally complete and

[†] In section 1.2 (Logic in biological systems) due to the established nature of some of the mathematics, Logic and Circuitry described; where not otherwise specifically referenced, the citation should be taken to be (Lehman, Leighton, & Meyer, 2017; Tanenbaum & Goodman, 2005).

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

therefore most useful gates are the NAND and NOR gates. An example of a circuit made up of these logic gates is the Full Adder which is a vital part of a computer processor. The Full Adder produces the sum of two numbers, by an operation, which sums two inputs and the carry value (Figure 1.3).









Name	Symbol	Truth table		
		INPUTs	OUTPUT	
NOT (inverter)		A	NOT A	
		0	1	
		1	0	
AND		A	B	A AND B
		0	0	0
		0	1	0
		1	0	0
		1	1	1
OR		A	B	A OR B
		0	0	0
		0	1	1
		1	0	1
		1	1	1
NAND		A	B	A NAND B
		0	0	1
		0	1	1
		1	0	1
		1	1	0
NOR		A	B	A NOR B
		0	0	1
		0	1	0
		1	0	0
		1	1	0
XOR		A	B	A XOR B
		0	0	0
		0	1	1
		1	0	1
		1	1	0
XNOR		A	B	A XNOR B
		0	0	1
		0	1	0
		1	0	0
		1	1	1
IMPLY		A	B	A IMPLY B
		0	0	0
		0	1	0
		1	0	1
		1	1	0

Figure 1.2 Boolean logic gates, symbols and truth tables. All symbols are those from the IEEE Std 91/91a-1991 standard.

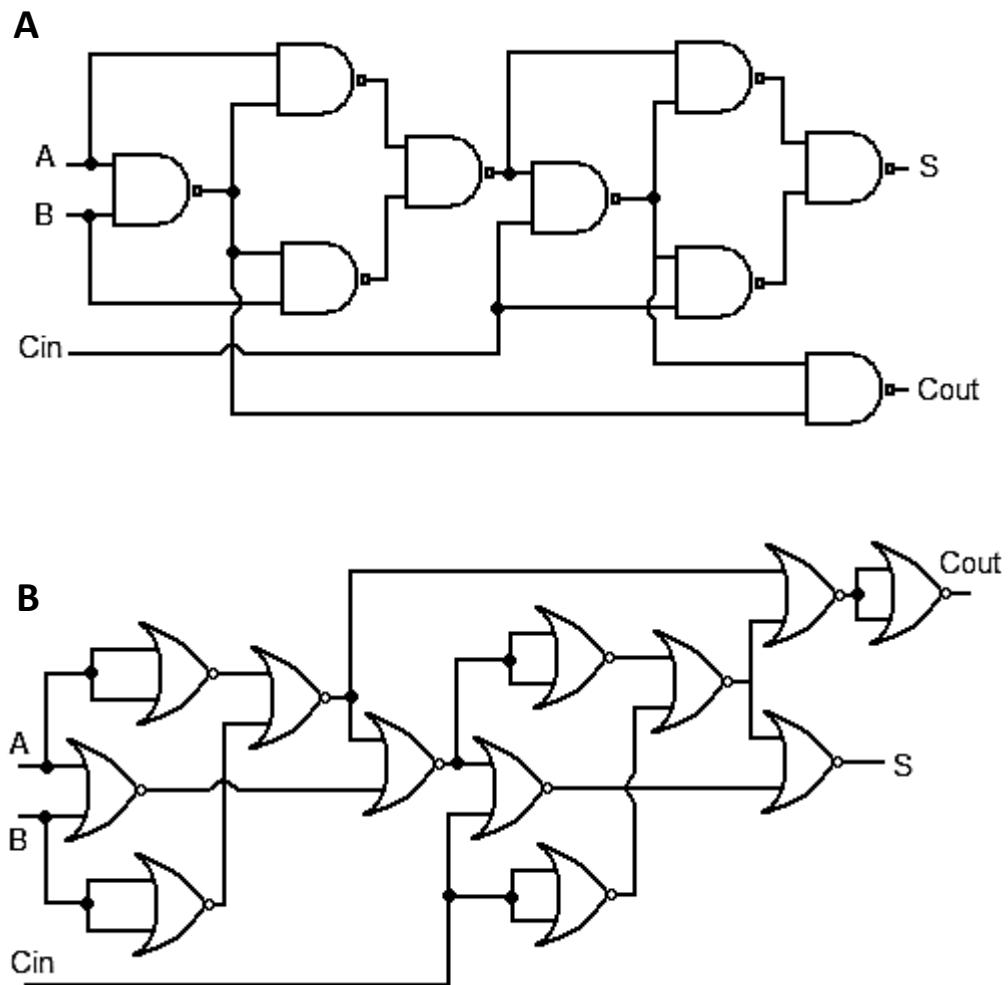


Figure 1.3 An example of a commonly used Logic circuit; The Full Adder logical circuit adds two numbers (expressed as binary input A and B). When the circuit needs to carry a value to the next position, the Cout carries the value to the Cin of the next full adder. As a demonstration of functional completeness the circuit is either implemented with the (A) NAND gate or the (B) NOR gate, both of which are functionally complete. Symbols vary slightly from IEEE standard as circuits were visualised using Logic Gate Simulator™.

1.2.1 History of the Boolean Logic Gate

In his seminal 1854 work, George Boole took the concept of algebra and expanded it, applying algebra to values of true and false (Boole, 1854). Despite this forming the basis of modern computing, ironically it was published after Ada Lovelace had published the first designs for a computer (Lovelace, Ada; Babbage, 1842). Over the course of the nineteenth century, Boolean logic was principally used in the field of Set Theory. It wasn't until 1938 that Claude Shannon applied Boolean algebra to binary switching in electronic circuits (Shannon, 1938). Over the course of the twentieth century logic gates have been built in a myriad of ways, including electromagnetic relays, vacuum tubes, pneumatic logic, optics, fluid logic, and even mechanical approaches such as marble runs, before the transistor achieved dominance (R. Stanković, 2008). In the twenty-first century in the age of synthetic biology, there is a similar expansion in the range of logic gate implementation approaches for biological systems. It is this field that the work presented in this thesis most contributes to.

1.2.2 Signal transfer in logic gates

Logic gates have a number of fundamental characteristics which need to be taken into account during the design process. Here, each of these characteristics are first described in the context of electronic systems (as an established form of logic implementation). Following this, the analogous characteristic is described in terms of one of the simplest biological logic gates: a repressor-based inverter, such as the ones used to make the Elowitz Repressilator is compared (Figure 1.4) (Elowitz & Leibler, 2000).

Logic gates are described in terms of binary inputs and outputs which can be 0 (low, False) or 1 (high, True). Physical implementations of theoretical logic gates, however, have a greater degree of complexity. In electronic logic gates, there is a range of input voltages that the circuit will accept as a 1 value, and another range of voltages which the gate will accept as a 0 value. Similarly, the output voltages constituting the binary 1 or 0, in practise also cover a voltage range. The input

voltage, together with a number of physical factors affects the output voltage. This results in a voltage transfer characteristic, in which the input position within the high or low input range of the logic gate, can alter the position of the output within the output high or low voltage range. The voltage transfer characteristic also characterizes the output voltage produced by an input voltage between the high and low ranges. In biological terms, the high (or 1) input voltage range might equate, for example, to the transcription level of a gene encoding a repressor protein. The low or 0 input voltage then equates to the levels of expression of a repressor at which the output gene ceases to be repressed. In this hypothetical biological system, the output low or 0 level voltage would equate to the basal level of expression when the output gene is repressed (Figure 1.4). In this example, the voltage transfer characteristic, would refer to the level of expression of the output reporter gene at each possible level of expression from the input promoter of the repressor.

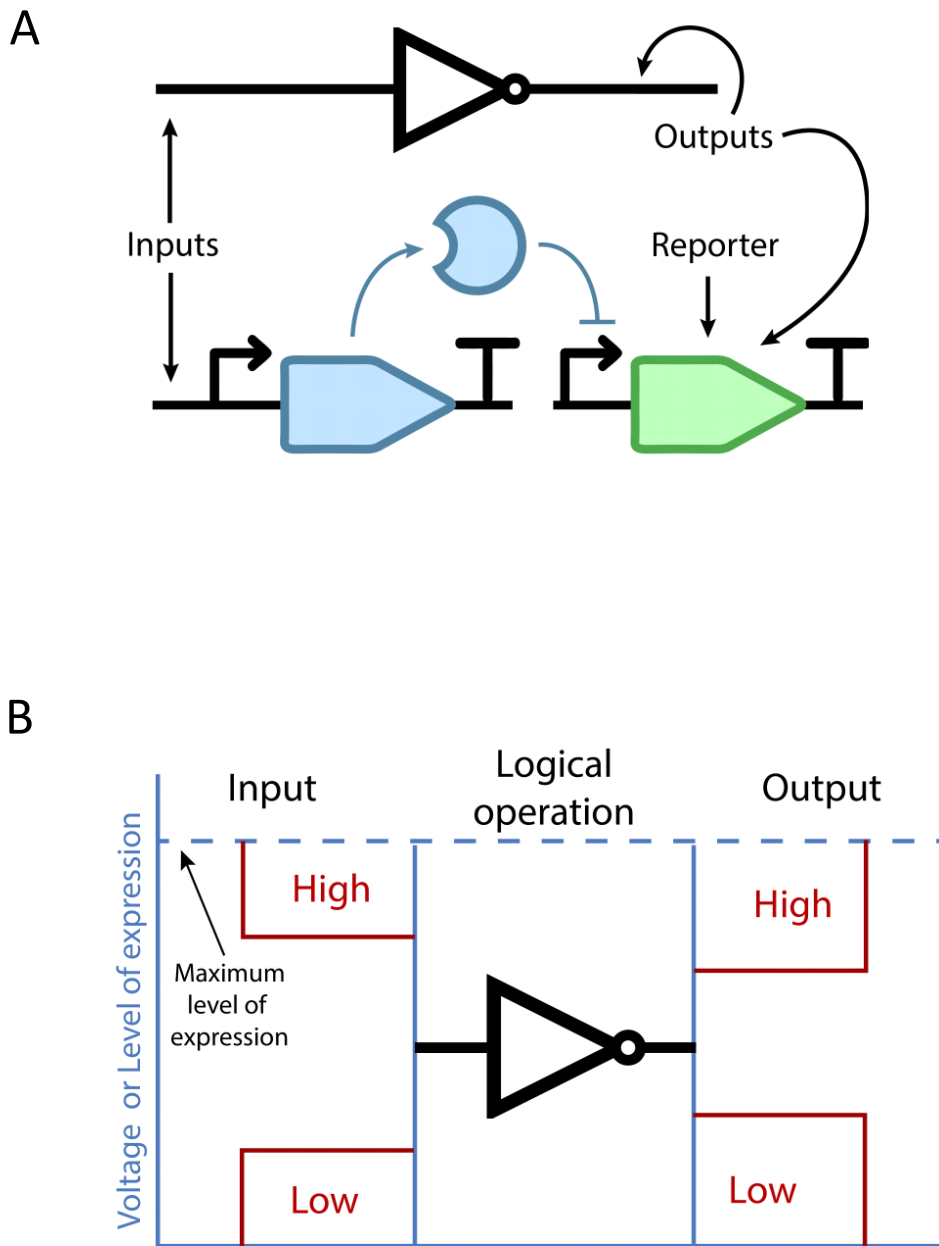
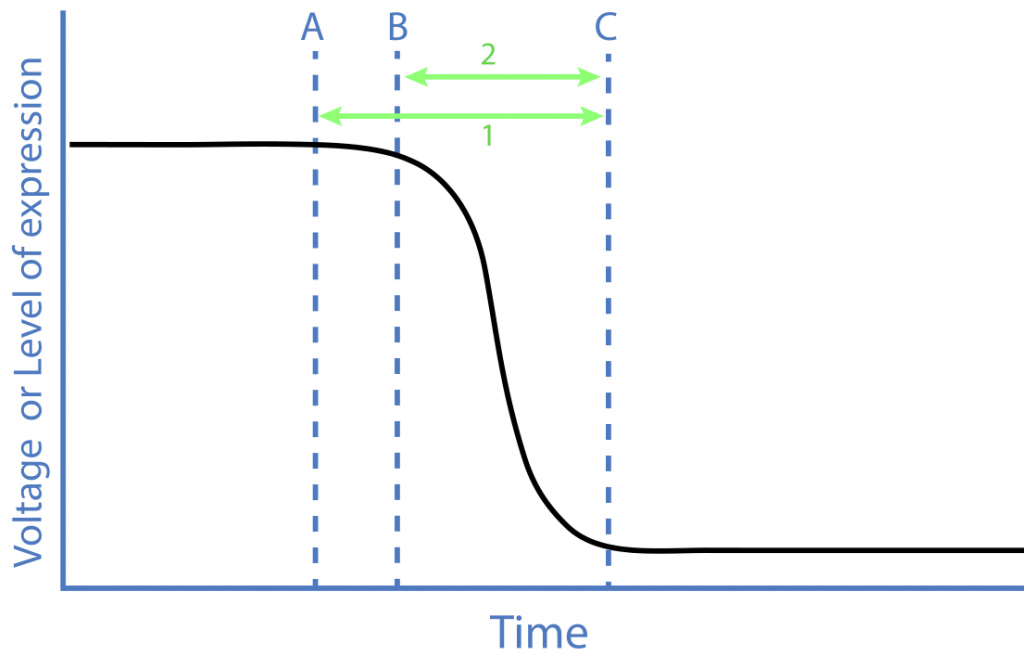


Figure 1.4 (A) The NOT gate can be implemented in a cellular environment as a genetic circuit by the use of a transcriptional repressor. **(B)** In this biological circuit the level of expression is analogous to the voltage in an electronics implementation of a NOT gate. There is a range of input levels that can be taken as High or (ON) and a range of outputs that might be expressed as High or Low. How the output voltage is affected by the input voltage is defined by the voltage transfer characteristic.

1.2.3 Latency

Latency refers to the delay between the change in the input of a gate and the change in the output in electronics: this is broken down into gate delay and fall or rise time (Figure 1.5). The gate delay is the full time between the changing of the input and the changing of the output. The fall or rise time is the period between the output starting to change and completing the transition from one output state to the other (Figure 1.5). One of the limiting factors in biological gates to date, is gate delay. There are a number of factors that can affect gate delay within a biological system. Using the repressor protein based inverter example described in Section 1.2.2, the gate delay includes transcription of the repressor coding sequence, translation of the mRNA, protein folding and accumulation time for surpassing the amount of repressor necessary for output gene repression. Subsequently the fall time can be described as the time required for the level of output reporter to equilibrate with the new rate of transcription. This period will be dependent on the degradation rate of the reporter and the dilution rate or growth phase of the host organism. In the opposite direction the gate delay would depend on the dissociation, degradation and dilution of the repressor alongside synthesis and folding of the reporter protein (Bowsher, Voliotis, & Swain, 2013). The dissociation can take particularly long in circuits using CRISPRi based repression, due to very high binding efficiency; though DNA replication can lead to an acceleration in the rate of dissociation.



A - Input becomes high, repressor expression is induced
B - reporter expression level begins to drop
C - out put is low, reporter expression level is repressed
1 - gate delay
2 - fall or rise time

Figure 1.5 Latency within a logic gate, is most commonly measured in terms of two periods. First, the gate delay which is measured between the input changing state (becoming high in this case) and the output changing state (becoming low in this case). Second, the fall or rise time which is measured between the output starting to change state and reaching the alternate state (in this case going from high to low).

1.3 CRISPR Cas9 system

Clustered Regularly Interspaced Palindromic Repeats (CRISPRs) were first discovered in Archaea and later found in bacteria (Jansen, Embden, Gaastra, & Schouls, 2002; Mojica, Díez-Villaseñor, Soria, & Juez, 2000; Mojica, Ferrer, Juez, & Rodríguez-Valera, 1995). CRISPR is primarily an adaptive immune system, where three main stages: adaptation, expression and interference allow targeted cleavage of invading foreign DNA (Figure 1.6). During the adaptation phase, foreign DNA will be processed by Cas proteins (Cas1 and Cas2 for the Cas9 system), integrating the protospacer sequence flanked by a Protospacer Adjacent Motif (PAM), which is necessary for distinguishing foreign DNA from host DNA. This sequence is integrated into an existing CRISPR RNA (crRNA) array, immediately downstream of the direct repeat. Subsequently during the expression phase, the entire crRNA array will be expressed alongside *trans*-activating crRNA (tracrRNA) and the effector protein such as Cas9 (Figure 1.6). The crRNA array can complex with tracrRNAs and in the presence of Cas9, RNase III enables cleavage to form the end complex of Cas9 and a single guiding RNA (gRNA). During interference, this complex can then target the respective protospacer sequence, complementary to the spacer within the gRNA, and cleave the sequence if the appropriate PAM is present.

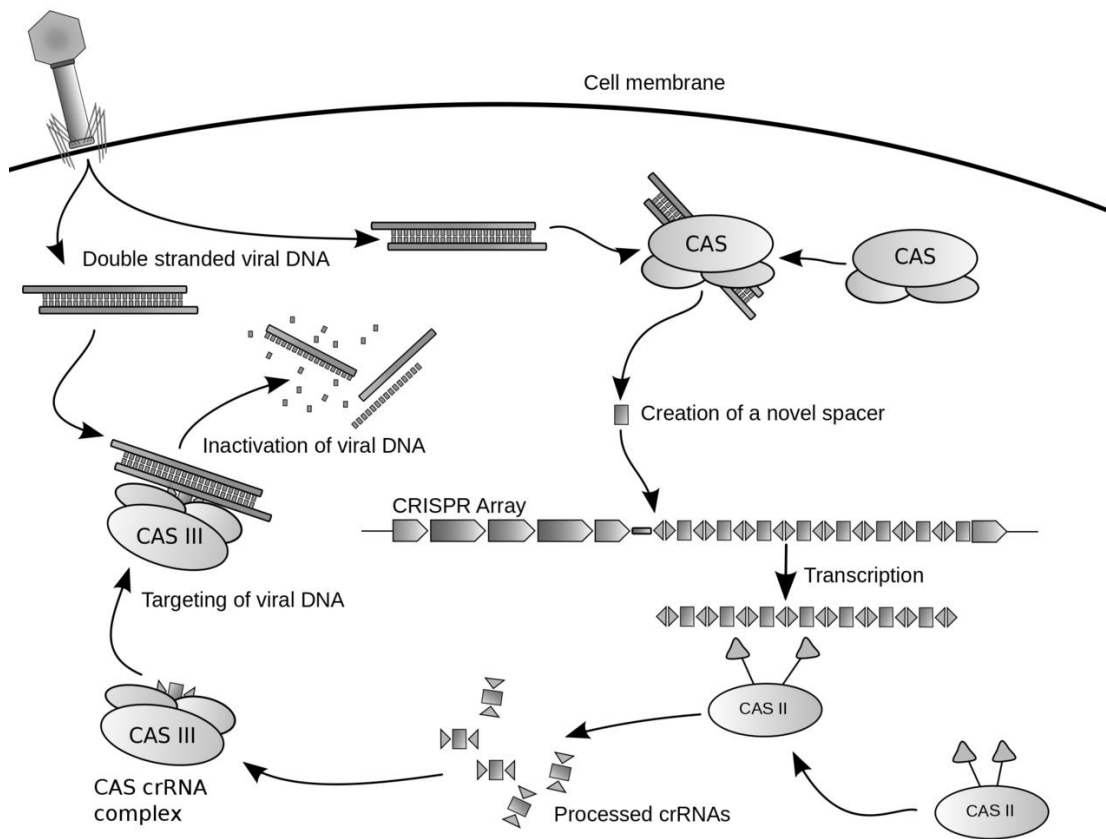


Figure 1.6 Mechanisms of CRISPR system. Wild type CRISPR system integrates sections of invading nucleotide either from plasmids or from viruses into the spacer regions of the CRISPR array where they are transcribed and used to target degradation of future invading plasmids or viruses. Figure by J. Atmos (2018), reproduced under Creative Commons License 3.0.

One of the best characterised CRISPR systems to date is the *Streptococcus pyogenes* Cas9 system (SpCas9). Being a type II CRISPR, the Cas9 protein is able to act without ancillary proteins, when coupled with a gRNA, to enable targeted cleavage. Further the SpCas9 system utilises a relatively simple PAM sequence 'NGG', permitting the targeting of many sequences, with NGG occurring approximately every 8 bp when targeting either DNA strand. The system has been further simplified by the fusion of the crRNA and tracrRNA into a single transcript to form a functional gRNA, which does not require processing (Mali et al., 2013).

The Cas9 nuclease itself possesses two nuclease domains, HNH and RuvC, which each cleave different strands to create a blunt double strand break within the gRNA base paired region. An interesting development, which underpinned this body of work, was the generation of a catalytically dead Cas9 nuclease (dCas9). This was achieved by single amino acid substitutions within the HNH domain and RuvC domain, with the subsequent dCas9 protein able to bind to but no longer cleave targeted loci.

While the basic mechanism of dCas9 dependent transcriptional repression appears simple, there is underlying complexity around the specificity and kinetics of complex formation with DNA and the degree of transcriptional repression. The main classes of context which can affect these are the level of homology between the spacer of the gRNA and the DNA sequence as well as the nucleotide composition of the spacer sequence. Which orthologue of Cas9 is utilised as well as the host organism also needs to be considered. From the perspective of off-target effects and specificity, complementarity is more important at the 3' end of the spacer which appears to act as the seed region during complex formation with the DNA target. Mismatches and bulges in the 5' end of the spacer still allow the formation of the gRNA-dCas9-DNA complex (Gilbert et al., 2014). In terms of DNA binding, mismatches between the gRNA and DNA template appear to be tolerated to a similar degree and in a similar pattern for both Cas9 and dCas9 (Bikard et al., 2013a; Gilbert et al., 2014), concordant with the binding kinetics they appear to share

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

(Richardson, Ray, DeWitt, Curie, & Corn, 2016). Cleavage activity of Cas9 on the other hand has different tolerances for bulges and mismatches, with a higher degree of complementarity being required for cleavage than binding (Bikard et al., 2013a; Knight et al., 2015).

This allows for Cas9 to be used in a comparable manner to dCas9 for regulation without cleavage, when a gRNA is used containing the appropriate truncation or mismatches. The extra sensitivity to mismatches seen for Cas9 cleavage is likely a result of the large conformational change that occurs during cleavage (Knight et al., 2015; Sternberg, LaFrance, Kaplan, & Doudna, 2015).

Catalytically inactive Cas9 is well suited for transcriptional regulation in synthetic gene circuits due to the high degree of transcriptional repression that can be achieved with the system and the ability to bind to and repress transcription from arbitrary DNA sequences. This allows the system to interact with a wide range of synthetic and endogenous promoters. Previous systems such as Zinc Fingers (ZFs) and transcriptional activator-like proteins (TALEs) required a long protein coding sequence for each additional transcriptional regulation target. In contrast, only one copy of the 4.2 kbp dCas9 protein needs to be expressed for the repression of multiple genes, when expressed alongside the corresponding gRNAs. The ability to use one gRNA to repress the expression of another gRNA is conducive to the generation of interconnected *de novo* networks. Here interactions can either equate to NOT gates: when the input gRNA is expressed (ON) the output gRNA target gene of the gate is repressed (OFF). Alternatively, if there are multiple input promoters expressing gRNAs to repress the output gene this functions as a NOR gate, with an arbitrary number of inputs. In this system, any input gRNA being expressed (ON) will lead to the output gene being repressed (OFF). To exploit this, Nielsen & Voigt, (2014) generated and validated 5 orthogonal promoter/gRNA pairs, which were used to create networks with up to 4 layers and interface with inducible promoters as well as endogenous genes in a predictable manner. In a more recent

paper, circuits regulating metabolic pathways were constructed using T7 polymerase promoters repressed by dCas9(Cress et al., 2016).

1.3.1 Modifications of the gRNA

1.3.1.1 Aptamer dependent gRNA functionality

An RNA aptamer is an RNA sequence which binds a specific ligand; these can be found naturally, or alternatively, selected *in vitro* via Systematic Evolution of Ligands by EXponential enrichment (SELEX). Generally, binding of a specific ligand to the aptamer stabilizes the RNA molecule or causes conformational changes in its structure. In nature, such aptamer sequences can be found in messenger RNA (mRNA) transcripts, where this change in structure is mechanistically linked to a change in the level of expression of the protein the mRNA encodes, for example by occluding the RBS. This is known as a riboswitch, with a classic example being the *BtuB* riboswitch in which binding of coenzyme B12 to the riboswitch regulates the rate of translation of the *BtuB* gene (Nahvi et al., 2002). This concept has also been applied to other RNA functionalities such as aptazymes: a combination of aptamer and ribozyme, where aptzyme cleavage becomes dependent on the presence of the aptamer ligand (Zhong, Wang, Bailey, Gao, & Farzan, 2016). The predictable nature of RNA binding and folding as well as the number of available aptamers presents the opportunity to engineer gRNAs as ligand-responsive switches, adopting different conformational states and functionality depending on presence or absence of small molecule or protein ligands.

1.3.1.2 Aptamer based recruitment

Aptamers have been used in several ways to increase the range of CRISPR/Cas applications and improve CRISPR/Cas efficacy for a number of pre-existing applications. Aptamers have been included in loop sequences of the gRNA to recruit effectors. for example, Konermann et al., 2015 fused two MS2 aptamers within loop sequences of the gRNA, allowing the gRNAs to recruit the VP64 *transactivation* domain, connected by a glycine/serine rich linker to the MS2 binding protein. This

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

has been applied, with a clinical focus, to up-regulate expression of latent HIV1 for therapeutic purposes (Zhang et al., 2015).

The inclusion of aptamers in the gRNA can also be used to facilitate live cell genome imaging. Aptamers added to the gRNA transcript recruit one of three possible fluorophores which can be added to the media. Fluorophores pass through the cell membrane to become part of the cytoplasm. The fluorophore which is recruited to the riboprotein complex is then in turn targeted to a specific region of the genome by the gRNA for imaging. The presence of two aptamer insertion sites in the gRNA allows the creation of combinations of colours, when using different pairings from the aptamers available, each of which corresponds to a different fluorophore. By recruiting different combinations of pairs of fluorophores with different emission spectra, then 6 different combinations of colours can be used to image the genome during cell division (Ma et al., 2016; Wang, Su, Zhang, & Zhuang, 2016).

Liu et al. (2016) utilised a series of aptamer-containing gRNAs, to engineer gRNA “signal conductors”, with functionality dependent on the presence or absence of ligand signals. In this paper, they demonstrated functionality of such signal conductors with either a small molecule (tetracycline, theophylline) or protein (β -catenin) as signals. In the ON state, the spacer sequence of the gRNA is complemented by an additional functional module inserted in the 3' end of the gRNA transcript, which also contains an aptamer sequence (Figure 1.7Aii). This means the gRNA does not direct the dCas9 complex to the DNA target site leaving transcription unrepressed. The ligand complexing with the ligand-binding loop of the aptamer induces a conformational change and binding between the complementary sequence and the antisense stem within the 3' aptamer module. This strand displacement releases the spacer sequence, and in this OFF state, the spacer guides the dCas9 complex to the DNA target site where transcription is repressed (Figure 1.7Aiii).

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

In order to optimise the system, and maximise the dynamic range achievable, a number of different antisense stem lengths were tested. Irrespective of the $\Delta\Delta G$ of ligand binding, Liu et al. (2016) found that an antisense region of 15 nucleotides maximised ligand-dependent range of transcriptional modulation. If the antisense region was increased to 18 bases, the stable interactions between the spacer and aptamer module resulted in reduced transcriptional repression activity, with low dependence on ligand concentration. Conversely, when the antisense region was reduced to 11 bases, the gRNA exhibited “leakiness”, with a degree of activity retained in the absence of ligand, which is hypothesised to be due to the lower stability of the hairpin (Liu et al., 2016).

Following this optimisation process, Liu et al (2016) used their signal conductors to construct logic gates, attempting all the symmetric Boolean logic gates. Whilst these gates certainly have some functional value, they also have great limitations. The two inputs of the logic gates are in the form of ligands to modulate two ligand dependent gRNAs, and in all cases (except the XOR logic gate) the combinatorial effect of the two inputs is evaluated through two complexes binding to/near the promoter, as opposed to one. Consequently, the effect of having one input in the ON state instead of two, approximates to being halfway between having no inputs ON, and both ON (an output value which might approximate to 0.5 rather than 1 or 0).

After this earlier work, Tang et al. (2017) engineered a system in which ligand-binding enabled a ribozyme to cleave itself from the transcript, taking with it the antisense *cis*-repressing element. Tang et al. (2017) were able to engineer gRNAs using either theophylline or guanine as aptazyme ligands. A “blocking sequence” in the transcript complemented a section of the gRNA sequence to repress activity (Figure 1.7B). Tang et al. (2017) explored a number of approaches to engineering the aptazyme into the gRNA including complementation of the spacer region, separating the gRNA into separate crRNA and tracrRNA molecules, and using the blocking sequence to prevent the complementation between the crRNA and

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

tracrRNA necessary to create a functional unit. Tang et al. (2017) found that the most effective of these three approaches was using the blocking sequence to complement the spacer region. They demonstrated functionality of a theophylline-regulated aptazyme gRNA, for theophylline-dependent cleavage of DNA using Cas9, and also a guanine-dependent aptazyme, for guanine-dependent activation of a promoter via dCas9-VP64-p65-Rta activity.

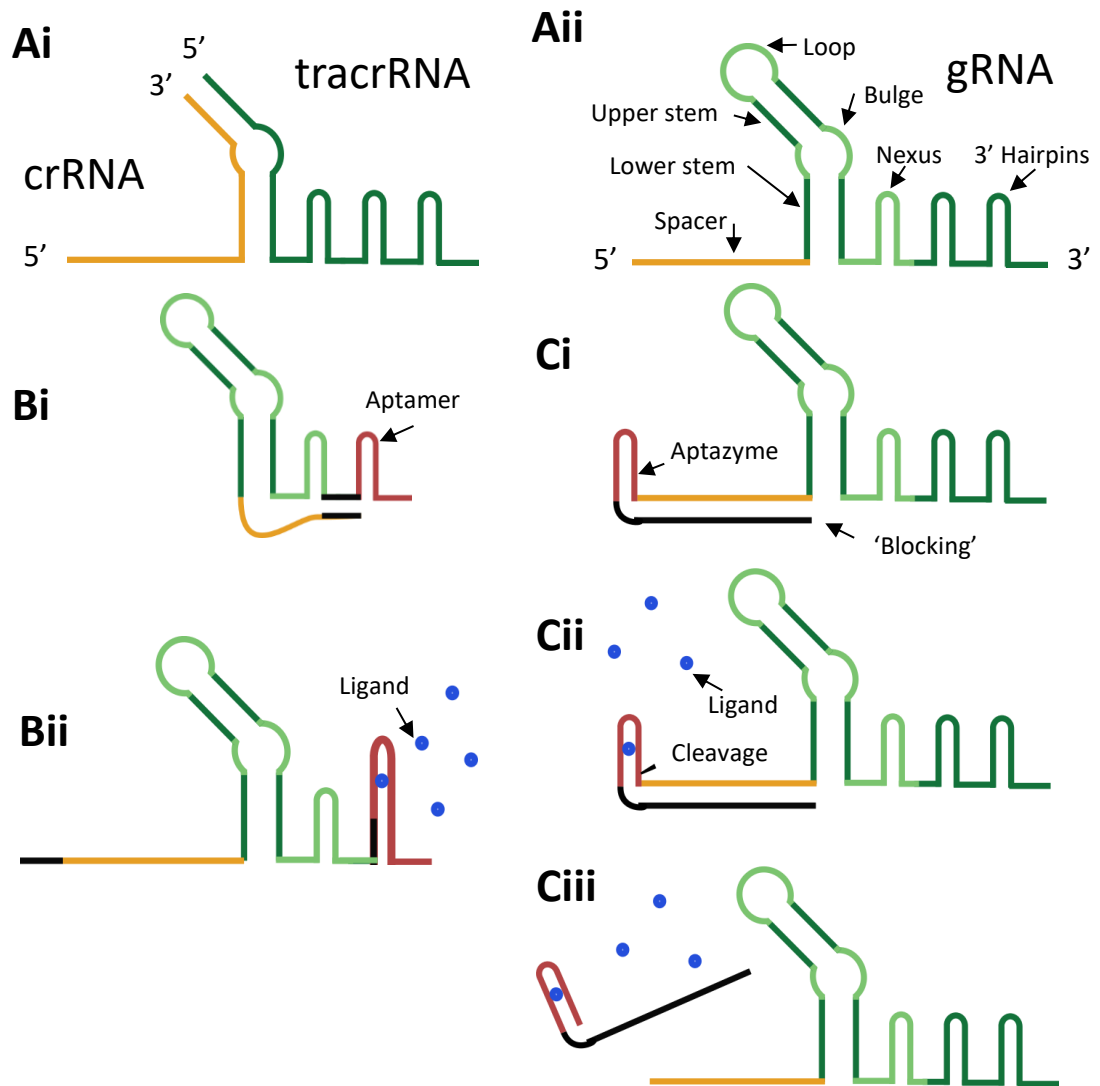


Figure 1.7 Model of (A) original structures (B) aptamer and (C) aptazyme-regulated gRNA system. (Ai) in the wt system, crRNA and tracrRNA combine with Cas9 to make a functional complex, the (Aii) gRNA is produced by combining the crRNA and tracrRNA in to a single transcript. The functional modules of the gRNA are labeled. (Bi) gRNA in its active, native conformation. (Bii) module added to the 3' end of the transcript complements the spacer region functionally inactivating it. (Biii) ligands bind to the aptamer in the 3' module resulting in a change in conformation releasing the spacer and rescuing functionality. (Ci) inactive system, spacer is functionality sequestered by blocking sequence. (Cii) when the ligand is present in the system this complexes with the aptazyme leading to a change in structure resulting in the, now active, aptazyme cleaving. (Ciii) after cleavage the blocking sequence dissociates from the gRNA allowing it to return to an active state.

1.3.1.3 Sequestration of gRNA by antisense RNA

As described in Section 1.3.1.2, Liu et al (2016) demonstrated dCas9 repression as the evaluation of two logical inputs by adding an extra dependency (repression was dependent on the gRNA and the aptamer ligand). Using a different approach, Lee et al., (2016) were able to form logic gates by making repression dependent on two RNAs. In this system they explored direct sequestration and degradation of a gRNA by an antisense RNA (asRNA) which complements the gRNA. This means repression is only intended to occur when the gRNA is present and the asRNA is absent.

Lee et al., (2016) used three rounds of improvement, initially the asRNA complemented the spacer region of the gRNA. As well as inhibiting complex formation between the dCas9 and the gRNA, an Hfq binding scaffold was included in the asRNA transcript (first MicF then Spot42). Hfq stabilises the asRNA and promotes interactions with other RNAs as well as recruiting RNase E to degrade the RNA heteroduplex (Morita & Aiba, 2011). This initial method of sequestration and degradation led to a de-repression of 15% which increased to 43%, 55% and finally 95%. Through the extension of the complementary spacer, exchange of Hfq scaffold from MicF to Spot42 which was found to have a higher affinity, and finally swapping the target of the asRNA from the spacer to an engineered 3' linker sequence added to the gRNA transcript. This allowed the investigators to further decrease the ΔG of heteroduplex formation between the asRNA and the gRNA. The strategy used by Lee et al (2016) expresses system components from two plasmids; first, dCas9 is expressed from a vector, available from Addgene - Plasmid #44249 (pdCas9-bacteria), which has become the standard for expressing dCas9 in bacterial hosts. Secondly, a single high copy ColE1 plasmid was used to express both the gRNA and asRNA RNAs. This is a different plasmid to the Addgene - Plasmid #44249 (pdCas9-bacteria). The details of this setup are described here as they were used as the basis for the design of the experimental system used in this thesis.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

The Lee et al., (2016) system has two input RNAs, which combine in a logical operation to yield a repressed or non-repressed output (analogous to OFF/ON or 0/1). When represented as a logic gate, this system equates to an IMPLY gate as the output is only repressed when the gRNA input is ON and the asRNA input is OFF (Figure 1.8). The IMPLY gate however is asymmetric and appears rarely in computational design. Lee et al. have therefore presented their mechanism in terms of gene regulation more generally, rather than emphasising use as a logic gate.

gRNA	asRNA	Mechanism	output
0	0	<p>promoter transcription dCas9</p>	1
0	1	<p>asRNA</p>	1
1	0	<p>Transcriptional repression gRNA</p>	0
1	1		1

Figure 1.8 Mechanism of antisense gRNA system with truth table of each mechanistic state. The Logical evaluation performed by the mechanism equates to an IMPLY gate.

1.4 RNA synthetic biology

When constructing genetic circuitry within synthetic biology, the requirements of a system can vary dramatically in terms of regulation, sensing, latency, orthogonality and interaction with host systems. Early genetic circuits principally used protein-based gene repression, including two papers seminal to the field of synthetic biology; the Elowitz repressilator and the Gardner toggle switch (Elowitz & Leibler, 2000; Gardner, Cantor, & Collins, 2000). However the diversity in RNA functions and predictable nature of RNA folding and structure, have led to an increase in focus on RNA-based regulation in genetic circuitry.

One of the advantages of utilising RNA from a synthetic biology perspective, is the wide range of functional mechanisms available. RNA can interact with small molecules or other RNA to modulate rates of translation, interfere directly with transcription, catalyse reactions including RNA cleavage and RNA synthesis, and form cofactors for enzymatic reactions (Dutta & Srivastava, 2018). Another great advantage is the predictable nature of RNA folding and structure, and the software tools available for RNA structure/function analysis and design such as the Vienna 2.0 RNA folding suite and NUPACK (Lorenz et al., 2011; Zadeh et al., 2011).

1.4.1 Toehold switches

There have been a range of mechanisms described by which an RNA molecule can change the rate of translation of an mRNA in nature, such as the *hok/sok* system and pseudoknot-dependent *repBA* system of the *IncB* plasmid (Gulyaev, Franch, & Gerdes, 1997; Praszkie & Pittard, 2002). Inspiration from these natural systems led to the creation of a range of synthetic translation-regulating RNA mechanisms. The first of these was the *cis*-repressed RNA/*trans*-activating RNA (taRNA) system created by Isaacs et al., 2004. In this system the Shine-Dalgarno (SD) of an mRNA is occluded by a 5' untranslated region (UTR), which forms a hairpin with the SD, preventing ribosome binding; this is referred to as *cis*-repressed mRNA (note, the 'crRNA' acronym used by the authors is not used here due to a conflict with crRNA 'CRISPR RNAs'). Translation of the transcript is activated by the taRNA, which forms

a heteroduplex with the *cis*-repressing element of the *cis*-repressed mRNA, displacing the hairpin and exposing the SD for ribosome binding (Figure 1.9). This system has two major limitations. First, complementing the SD imposes sequence constraints on the *cis*-repressing element, and hence on the taRNA. This results in low orthogonality as there is a high degree of sequence similarity between the *cis*-repressing elements and taRNAs from different pairs, resulting in crosstalk. This also impacts the range in SDs that can be used. The second limitation is that the same structure which occludes the SD also prevents the taRNA from interacting with the *cis*-repressed mRNA. Consequently, heteroduplex formation is dependent on either the taRNA complexing with the nascent *cis*-repressing element, or spontaneous hairpin melting. The result is in a low percentage rescue of function by the taRNA, and therefore a low taRNA-dependent fold change in expression (Isaacs et al., 2004).

The basic concept that the *cis*-repressed mRNA/taRNA system was based on (that a hairpin can reversibly inactivate a functional module of an RNA sequence, and that a different RNA can displace the hairpin to rescue transcript functionality), has the potential to be applied in a variety of ways if previously highlighted drawbacks can be mitigated. Green et al 2014 sought to create a system in which both of these drawbacks were mitigated. The biggest innovation was the inclusion of a 'toehold' sequence at the 5' end of the hairpin, which was also complemented by the 'trigger RNA' (note the change in nomenclature from taRNA to trigger RNA). Exposed nucleotides of the toehold are available to form a seed interaction with the trigger RNA, and this transient complex allows for strand displacement to spread up the hairpin, displacing the *cis*-repressing element from the SD (Figure 1.9). In the dynamic equilibrium between dissociated RNA molecules and heteroduplex formation, the toehold increases the forward rate constant, by reducing the activation energy required for heteroduplex formation. It also reduces the reverse reaction rate, by generating a heteroduplex with a lower enthalpy than the enthalpy of *cis*-repressing element hairpin formation. This results in a higher

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

equilibrium constant, and a higher proportion of the *cis*-repressed mRNA pool expressed within a cell complexing with the trigger RNA. The second drawback of sequence constraints was mitigated by changing the mode of repression so the SD sequence and AUG codon are loop and bulge sequence respectively and the hairpin forms in the sequence either side of the SD and with the spacer sequence between the SD and AUG. This reduces the sequence constraints on the *cis*-repressing element and the trigger RNA; the increase in available sequence space allows a higher degree of orthogonality.

The system was later expanded and improved by the multiplexing of toehold switch hairpins to allow multiple inputs to a single modified mRNA or 'gate RNA' thus creating an OR gate (Green et al., 2017). The authors also expanded the logical applications of the system by subdividing the taRNA into multiple component input RNAs which need to bind together to form a functional complex, able to *trans*-activate the toehold switch. As all input RNAs are required for *trans*-activation this is an AND gate. NOT gates were also included by adding the antisense of an input RNA to sequester it (Green et al., 2017).

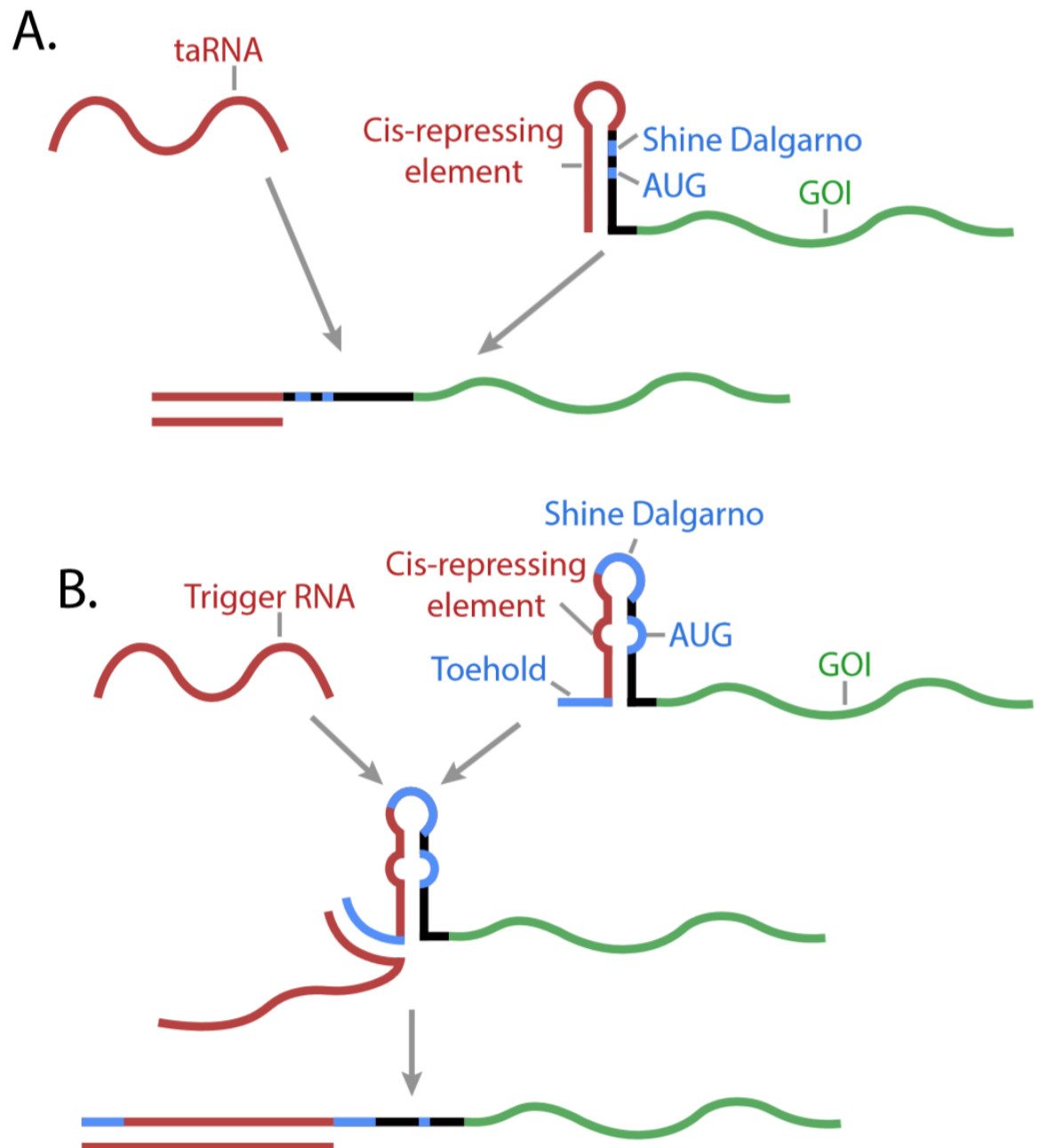


Figure 1.9 (A) Mechanism of interaction between *cis*-repressed mRNA and *taRNA* to expose *cis*-repressed mRNA SD for translation initiation created by Isaacs et al., 2004. (B) Mechanism dramatically improved upon by Green et al 2014. The toehold forms a seed interaction with the trigger RNA increasing the probability of strand displacement and heteroduplex formation. Neither the SD nor the AUG are directly complemented, reducing constraints and increasing the available sequence space allowing for a higher degree of orthogonality

1.5 Objectives for constructing a NAND gate

When constructing and optimising the NAND gate presented in this thesis there were three main objectives that the majority of challenges and design decisions could be described in terms of. These main objectives are introduced here. For the first objective, a NAND gate has a high output unless both inputs are ON. The equivalent for the gate developed here, being that the output promoter is unrepressed unless both the input RNAs are being expressed. Hence, the first objective is that neither of the two input RNAs lead to repression of the output when expressed individually. In a similar manner, the second objective is that the expression of both input RNAs leads to a high level of repression of the output promoter, as would be seen in a NAND gate. While the first two objectives are in terms of making the RNA mechanism operate in the manner of a NAND gate, the third objective is in terms of orthogonality of a number gates expressed within the same host. To maximise the utility of this NAND gate to practitioners it is important that multiple versions of a gate may be generated targeting different promoters and that these versions when expressed in the same cell operate without crosstalk between them. Hence the third main objective is that the design should be amenable to creating a range of versions, with a maximum orthogonality between different instances.

1.6 Targets for inactivation within the gRNA

The NAND gate presented in this thesis will rely on the reversible inactivation of a gRNA by the inclusion of a *cis*-repressing element in the gRNA transcript. In this section the information to be considered when selecting a target for inactivation of a gRNA is introduced. The gRNA is made up of seven principal modules (Figure 1.7) each of which has a different role in complex formation, both between the gRNA and the Cas9 Protein, and between the resulting complex and the target DNA. Cas9 undergoes substantial conformational changes during complex formation with the gRNA (Jinek et al., 2014). When designing how to reversibly disrupt the gRNA a

rational place to start would be looking at which part of the gRNA sequence forms the seed interaction with the Cas9 protein. Unfortunately, this is not known, and whilst there are speculations, a mechanistic path of complex formation has also not yet been elucidated. There are two types of data available on which a design may be based. The first is structural data showing which parts of the gRNA sequence are in contact with which domains of the Cas9 protein (Nishimasu et al., 2014). The second is mutational data showing which sequences within the gRNA are required for catalytic activity, by mutating the sequence and observing the effect on *catalytic* activity (Briner et al., 2014).

The crystal structures for the gRNA-Cas9-DNA complex show very few contacts between the Cas9 protein and the 3' hairpins of the guide RNA (Nishimasu et al., 2014) (Figure 1.10). This implies that the 3' hairpins are unlikely to be a suitable target for disruption by the *cis*-repressing element, because disruption may not prevent complex formation with dCas9, or the resulting repression of the output promoter, contravening the first objective. The first of the two 3' hairpins make a small number of contacts with the C-terminal domain and RuvC nuclease domain. This region is also therefore ruled out, due to the low potential for disruption of transcriptional repression (first objective).

In contrast with the 3' hairpins, the nexus displays extensive contacts with the C-terminal domain and the arginine rich bridge helix. This is consistent with mutation data in which changes to complementarity within the nexus have led to elimination of catalytic activity for Cas9 (Briner et al., 2014). This meant the nexus was not ruled out as a potential target for repression.

The lower stem, bulge, and upper stem formed by the crRNA repeat and tracrRNA antirepeat also show a great many contacts with the Cas9 protein. Mutations which disrupt complementarity within the lower stem or bulge can both lead to the elimination of catalytic activity (Briner et al., 2014). The upper stem on the other

hand can be entirely removed while maintaining catalytic activity, meaning the bulge and lower stem remain as potential targets for disruption (Briner et al., 2014).

The spacer region of the gRNA can be divided into two parts, the seed region (3' end of the spacer) and non-seed region (5' end of the spacer). The seed region is where duplex formation between the gRNA and target DNA begins. It also has lower levels of mismatch tolerance (Bikard et al., 2013; Gilbert et al., 2014). The 5', 10 nucleotide, non-seed RNA sequence lies in the cavity formed between the RuvC and HNH nuclease domains. The seed region is pre-ordered through extensive contacts with the ribose-phosphate backbone into an A form conformation (Jiang, Zhou, Ma, Gressel, & Doudna, 2015).

Due to the lack of specific contacts within the non-seed region of the spacer, and contacts within the seed region being specifically in the ribose phosphate backbone, a complementary *cis*-repressing element targeted to the spacer region would not necessarily stop the gRNA from complexing with Cas9. Complementation would likely only lead to minor topological changes in the phosphate backbone at the point where it contacts the Cas9. This would initially suggest the spacer region would not be a good target for the *cis*-repressing element, however, as the goal is to prevent transcriptional repression of the output promoter. Therefore, complementation between the *cis*-repressing element and the seed region (5' end of the gRNA spacer), displacing or preventing contact between the seed region and the target DNA may provide the most promising avenue of investigation, particularly given the small number of mismatches with the seed region required to impair DNA binding (Pattanayak et al., 2013).

In summary, having looked at the structural roles of each functional module of the gRNA as well as the available mutation-functionality data to consider the potential efficacy of a *cis*-repressed element targeted to each functional module of the gRNA, four potential candidates remain. The spacer region, the lower stem, the bulge and finally the nexus.

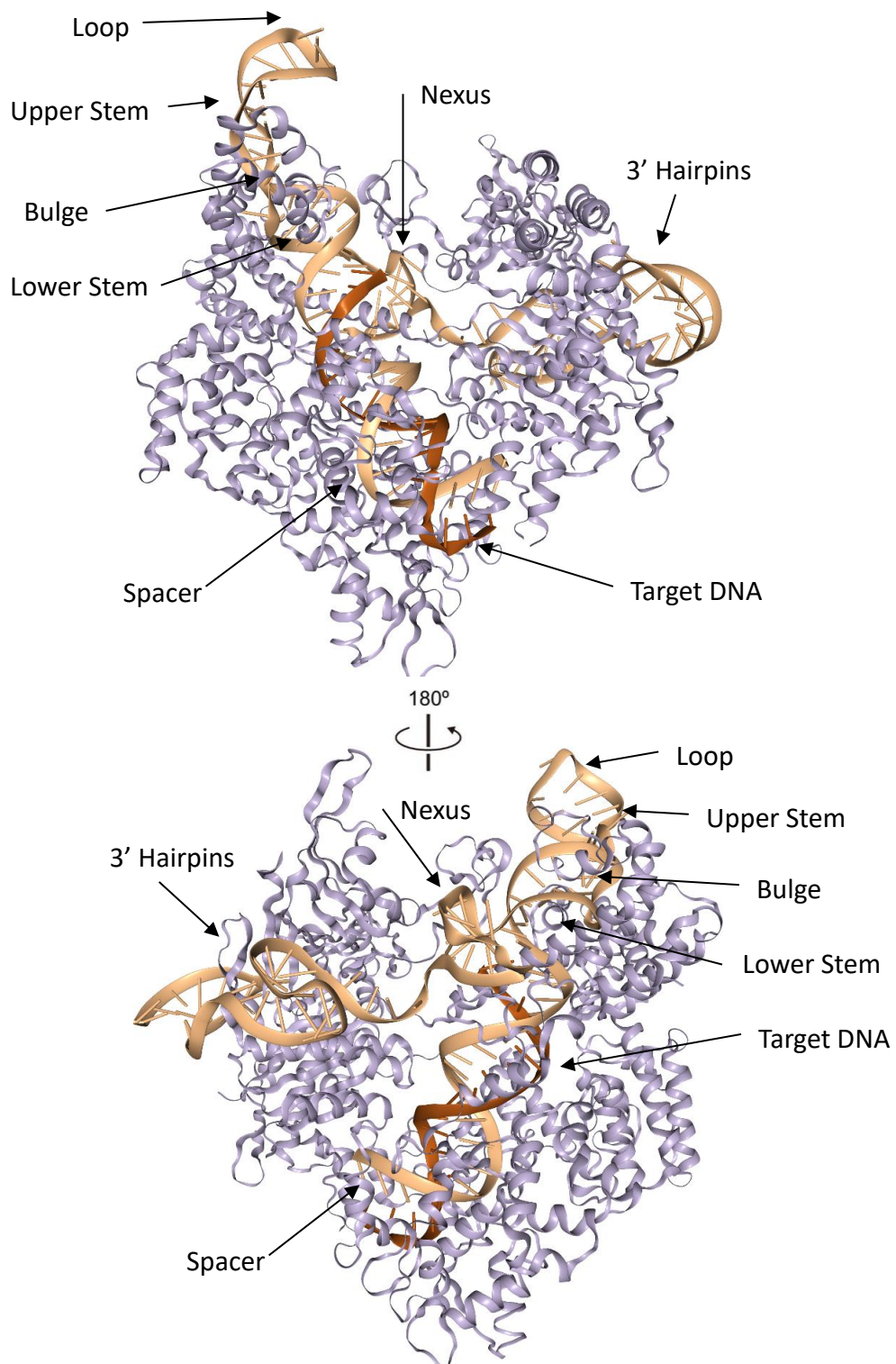


Figure 1.10 Crystal structure of gRNA complexed with Cas9 and the target DNA. (Nishimasu et al., 2014)

1.7 RNA-RNA interaction initiation

There are a number of ways to initiate the interaction between an asRNA and an RNA hairpin (Figure 1.11) such as a 'kissing loop' interaction, toehold interaction (described in Section 1.4.1), interaction between the asRNA and the nascent crgRNA before it has folded, or final interaction between the asRNA and members of the structural ensemble in which the complementation between the *cis*-repressing element and the spacer has spontaneously melted (Andersen & Collins, 2001; Meyer, Chappell, Sankar, Chew, & Lucks, 2015). Both complementing with nascent RNA and relying on spontaneous melting are inefficient and require a very high concentration of asRNA (Chappell, Takahashi, & Lucks, 2015). The kissing loop interactions occur commonly in nature, but suffer from topological restrictions due to torsion introduced through asRNA displacement of the hairpin (Di Palma, Bottaro, & Bussi, 2015). The toeholds however, have been used and characterised in a number of *de novo* systems, can be optimised for orthogonality and reduce the activation energy for heteroduplex formation reducing the ratio of asRNA to crgRNA required (Green, Silver, Collins, & Yin, 2014). This highlights the toehold as a suitable method of interaction initiation for two input RNAs.

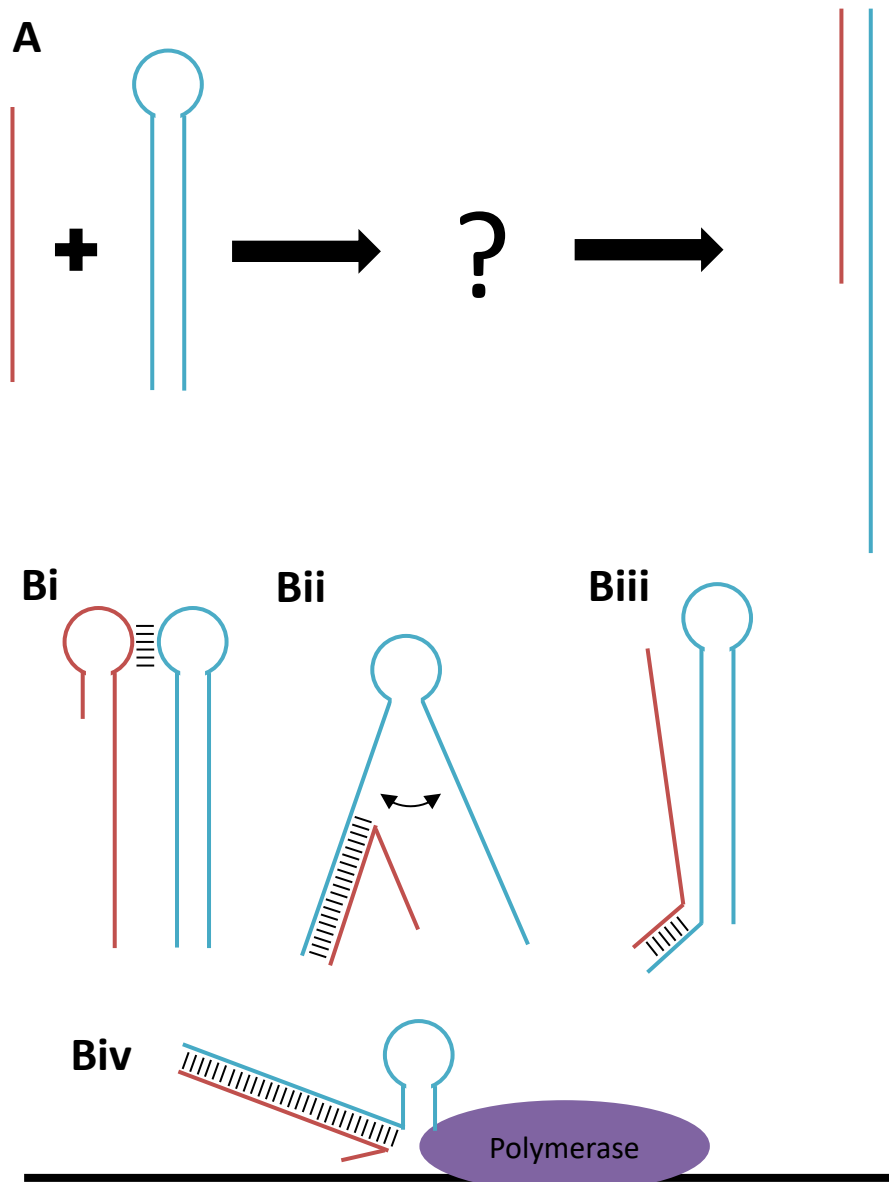


Figure 1.11 (A) there are several ways to initiate the interaction between an asRNA and an RNA hairpin to form a heteroduplex. (Bi) kissing loop interaction. (Bii) spontaneous melting of hairpin. (Biii) Toehold interaction. (Biv) Interaction between the asRNA and the nascent hairpin RNA prior to hairpin structure formation.

Chapter 2: Construction & characterisation of crgRNA

2.1 Introduction

The aims of synthetic biology are to apply engineering based approaches to the utilisation of cellular functions for technological applications. Synthetic biology circuits available to date use protein, DNA, and RNA components to achieve a range of functionality, from feedback and oscillation to bi-stability and logic. The main limitations of the currently available systems include the limited number of well-characterised, orthogonal parts and the metabolic load placed upon the host system. Other factors that must also be considered are: the human and economic resources required to synthesise and build such circuits, the amount of genetic real estate they require, and the limited number of components that can be incorporated into any one cell. Many of these problems are particularly pronounced in protein-based logic systems (Bradley, Buck, & Wang, 2016).

Current synthetic biology approaches offer the potential to perform a great range of tasks, from sensing through to actuation. To coordinate these abilities within biological devices, complex biological computing is required. Taking inspiration from the approaches adopted by engineers working in more traditional engineering media, biologists have been seeking to construct binary-based systems to regulate biological systems using Boolean logic (Bradley et al., 2016; Green et al., 2017; Liu et al., 2016). These approaches can be divided by the different mechanisms used, including protein engineering and RNA engineering, as well as by the level at which they modulate activity, from transcriptional modulation to post translation modification (Meyer, Chappell, Sankar, Chew, & Lucks, 2015; Prindle et al., 2014).

Much of the early work on cellular logic focused on the use of transcriptional activators and repressors acquired through genome mining (Stanton et al., 2014). With this approach, it was difficult to produce logic gates that were both orthogonal and functionally equivalent to one another. When orthologues have high homology

there is a greater probability of functional uniformity but also crosstalk, when orthologues are more distantly related, the inverse is true(Stanton et al., 2014). Next, a diversity of protein based systems emerged, ranging from the use of zinc finger and TALE effectors to intein based systems(Gaber et al., 2014; Schaerli, Gili, & Isalan, 2014). These systems are functional for creating individual logic gates but as the number of logic gates required increases, so the metabolic burden increases as does the genetic real estate needed. There are also temporal constraints when layering these gates as the transcription, translation and folding of protein can take time and lead to substantial dynamical error.

RNA systems on the other hand can be produced with relatively little metabolic load and due to the predictable nature of RNA folding, as well as the expansive sequence space available, large libraries of orthogonal, functionally equivalent components can be generated (Meyer et al., 2015). A common approach is to use RNA regulators to modulate the rate of translation from a constitutively expressed, engineered mRNA (Green et al., 2017). Such systems include the toehold systems or a system using Hfq based sRNA repression (Sakai et al., 2013). The key issue with such systems is that when constructing complex multi-layered logical networks, it is necessary for the output of one gate to form the input to the next. When the input is transcription of RNA and the output is modulated translation rate of protein, an extra step is required to make this output an input into the next gate.

The CRISPR Cas9 system has allowed the targeting and cleavage of specific DNA sequences using predictable, RNA sequence based targeting(Cong et al., 2013; Mali et al., 2013). The mutation of Cas9 to remove catalytic activity forming dCas9 has allowed a plethora of CRISPRi based regulatory systems to emerge (Qi et al., 2013). The advantages of these systems include the high fold change that can be achieved with a dCas9 based system and the predictability of the system, the ability to target multiple genetic loci, as well as the low genetic footprint for each additional target within a cell. Additionally, the output of a dCas9 based system, is a modulated rate of transcription. As transcribed gRNAs act as the inputs, this opens the way for

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

systems in which dCas9 based regulation is applied to the expression of gRNAs (Nielsen & Voigt, 2014).

This chapter seeks to combine the advantages of RNA based logic with those of dCas9 based systems. Specifically, to create a functionally complete Boolean NAND gate by making dCas9 based repression dependent on the expression of two RNAs.

2.1.1 Overview of an RNA-mediated NAND gate

The majority of genetically encoded logic gates fall into two categories; those that are protein-mediated and those that are RNA-mediated. The two categories each have their own advantages: protein or repressor/transcription factor based systems offer high fold changes in output, but often have problems with either orthogonality or library size. On the other hand, RNA based mechanisms, due to their highly designable nature, offer very large library sizes but often produce small fold changes in output. In the CRISPR dCas9 system there is an opportunity to combine the advantages of the two systems: an ability to produce large numbers of orthogonal logic gates, and with high output fold repressions.

For the formation of a logic gate, repression needs to be dependent on two or more inputs. In the existing CRISPR/Cas9 system, repression is dependent on the presence of both a gRNA and dCas9. While this can represent a logic gate, a large library of such gates operating orthogonally would not necessarily be possible due to the constraints from a low number of dCas9 orthologues as well as orthogonality issues between dCas9 orthologues and their respective gRNAs. Instead an approach was selected in which transcriptional repression was made dependent upon the interaction of two RNAs. One intuitive approach is to create a gRNA antisense RNA, which would complement the gRNA and thus prevent it from complexing with dCas9 or interacting with the target DNA sequence (Lee, Hoynes-O'Connor, Leong, & Moon, 2016a). Rather than yielding the more commonly used symmetric logic gate in which the two inputs are interchangeable, this would produce an IMPLY gate. An IMPLY gate is an asymmetric gate in which the output is only OFF if input A

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

(the gRNA) is ON (present) and input B (the anti-sense or asRNA) is OFF (absent). The IMPLY gate is a less commonly used logic gate which is not functionally complete.

An alternative approach is one in which the element which represses gRNA activity is included in the gRNA transcript itself, yielding a gRNA which is reversibly inactivated. Such gRNA can be reactivated when complexed with an asRNA which complements the repressing element in the gRNA transcript. This allows the rest of the gRNA transcript to return to its native conformation, enabling complexing with dCas9, to repress a target gene (Figure 2.1).

The design presented here is a system in which dCas9 mediated repression is dependent on two different RNA inputs which rely on the insertion of a '*cis*-repressing element' into the gRNA transcript (Figure 2.1). This *cis*-repressing element will sequester a region of the gRNA transcript inactivating the *cis*-repressed gRNA (crgRNA) in terms of transcriptional repression. This effect can be negated by interaction with an asRNA which interacts with the crgRNA through a toehold, passing through strand displacement to complement and sequester the *cis*-repressing element. This leads to the gRNA region of the transcript returning to a native state to produce dCas9 mediated transcriptional repression of an output gene. The experimental setup to test this system will be three plasmids; one to express dCas9, one to express a fluorescent reporter and a final plasmid to express the asRNA and crgRNA. The system will be operating as intended if, only when both the crgRNA and the asRNA are expressed is the output reporter repressed by dCas9 complexed with the crgRNA and asRNA (described further in Section 1.3.1.1).

This mechanism is a functional analogue of a NAND gate, as the output gene is only repressed in the presence of both of the input RNAs; the inactive crgRNA and the activating asRNA. Both RNAs must be present to see repression of the output. An additional advantage of this system is that, as the input and output can be expressed in terms of transcriptional flux from a targeted promoter, this system is

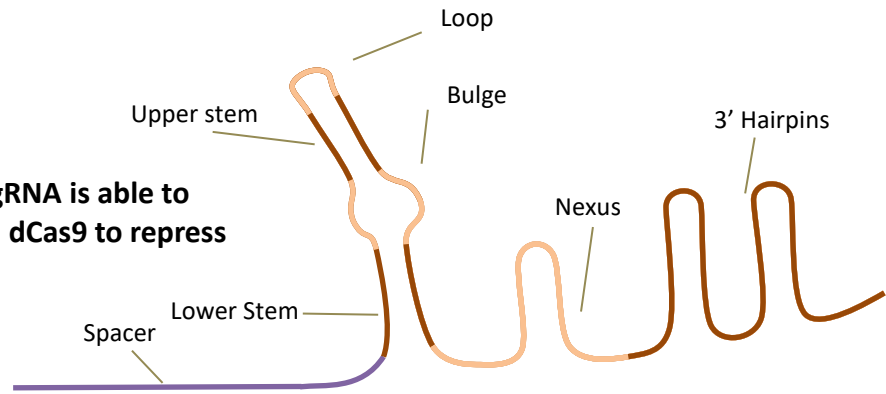
Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

amenable to layering. In addition, the RNA basis of this mechanism opens up the opportunity for large libraries of orthogonal parts.

A

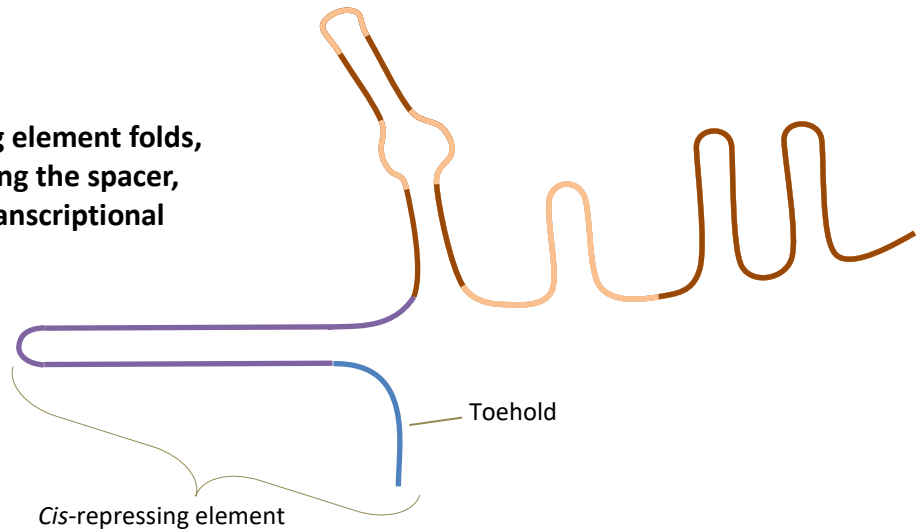
gRNA

Native state gRNA is able to complex with dCas9 to repress transcription



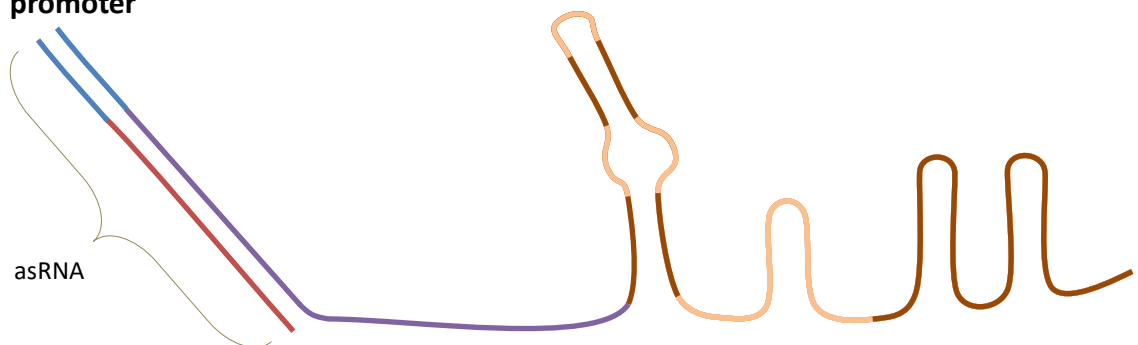
crgRNA

Cis-repressing element folds, complementing the spacer, preventing transcriptional repression



crgRNA-asRNA

complex asRNA complements the *cis*-repressing element, displacing it from the spacer, allowing the rest of the transcript to act as a native state gRNA and transcriptionally repress the target promoter



B

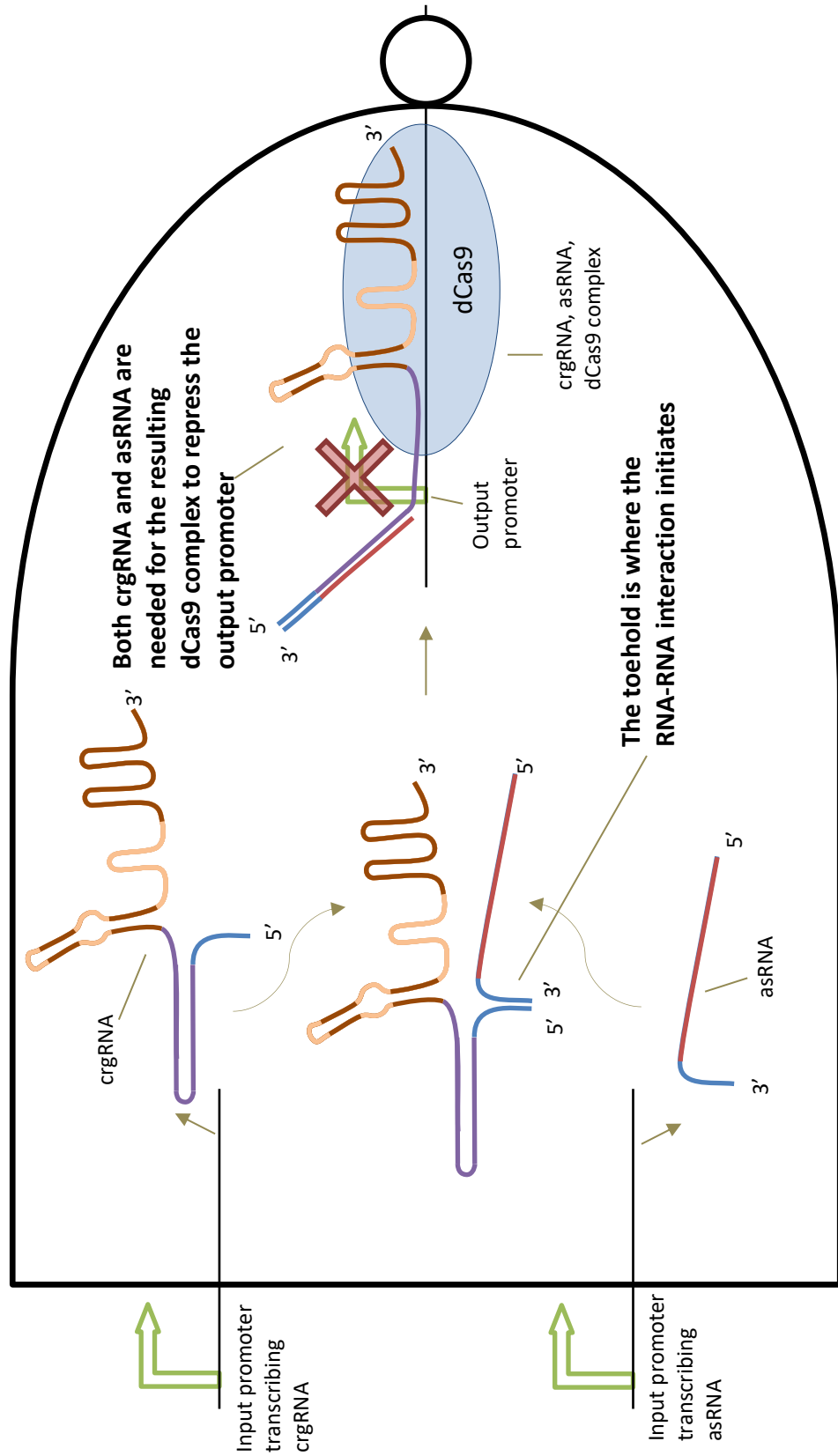


Figure 2.1 Please see figure legend on next page

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Figure 2.1: NAND gate design. The gate consists of a constitutively active output promoter that is repressed only in the presence of both the crgRNA and asRNA inputs. (A) Schematic representation of the inactive crgRNA and the active crgRNA /asRNA heteroduplex. (B) The crgRNA/asRNA system performing as a Logic gate with the two RNA transcribing promoters as the inputs and a single repressible promoter as the output. Repression of the output is dependent on the two input RNAs in the same manner as the logic gate. The inputs and output of the gate are organised in to the symbol of the NAND gate to make the inputs and output clear.

2.2 Results

2.2.1 Experimental setup

To test the NAND gate in *E. coli*, there are a number of components that need to be co-expressed (Figure 2.2). These components are expressed from three plasmids. The first component is a reporter to act as the output of the logic gate; so when the output is ON the reporter is expressed and when the output is OFF, the reporter is repressed. For this purpose mCherry was expressed from plasmid pZS2-123. This plasmid also expresses two other fluorescent proteins (YFP and CFP) which are used as controls (Cox, Dunlop, & Elowitz, 2010).

The second plasmid expresses dCas9 which mediates the repression of the output reporter gene (mCherry). This is the AddGene plasmid pdCas9-bacteria (#44249) which expresses dCas9 under the control of the anhydrotetracycline (aTc) inducible pLtetO-1 promoter and is the standard plasmid for expression of dCas9 in bacterial systems. The final two components are the two RNAs which act as inputs to the logic gate: the crgRNA and the asRNA. Both are expressed from the pBR322 plasmid. The pLlacO-1 promoter was selected as an inducible, titratable promoter to drive the expression of the crgRNA, allowing for tuning of the expression level. The j23119 promoter is used to drive the expression of the asRNA as it is a strong promoter maximising the observable effect of the asRNA. To simulate the turning OFF and ON of the asRNA input, two instances of the plasmid exist for each variant. One instance of the plasmid expressing the asRNA and the crgRNA (the +asRNA state), and the other expressing the crgRNA and a nonsense, control instead of the asRNA (the -asRNA state). The reporter plasmid also expresses CFP and YFP; YFP is used as a control to verify whether the effects observed are specific to the mCherry target. When the logic gate functions correctly the output reporter (mCherry) will only be repressed when both the crgRNA and the asRNA are expressed (crgRNA +asRNA) (Figure 2.2).

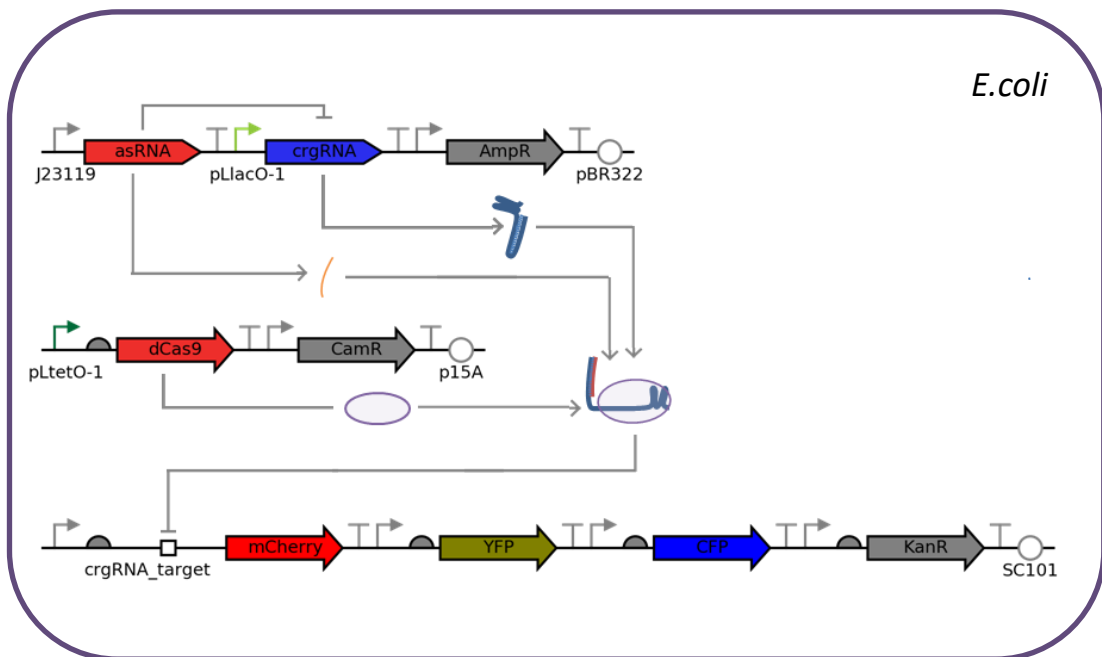


Figure 2.2 Three plasmid experimental setup. pBR322 expressed the *asRNA* and the *crgRNA*, AddGene plasmid #44249 was used to express *dCas9*. Fluorescent reporter protein (*mCherry*) and control fluorescent proteins (*CFP*, *YFP*) were expressed from pZS2-123 (Cox et al., 2010). The fluorescent reporter gene (*mCherry*) acted as a measure of the level of repression created by the system.

2.2.2 Selecting target for *cis*-repressing element

When designing the *cis*-repressing element to be included in the gRNA transcript there are a number of important variables to consider. First, which region or functional module within the gRNA should the *cis*-repressing element complement to reversibly inactivate the gRNA transcript. Second, should the *cis*-repressing element be included in the 5' or 3' end of the transcript. Third, how should interaction between the activating asRNA and the crgRNA initiate and propagate.

As described in the introduction (section 1.6), the structural roles of each of the functional modules of the gRNA as well as the available mutation-functionality data highlighted four promising candidates. The spacer region, the lower stem, the bulge and finally the nexus. Of the four candidates identified, the bulge sequence lies between the upper and lower stems. Consequently, a *cis*-repressing element, which only complements the bulge sequence (4 nt) has a small change in enthalpy (-0.67 kcal/mol). This is likely to result in the transcripts native conformation having the highest prevalence in the ensemble, negating the effect of the *cis*-repressing element. Therefore, to effectively complement the bulge sequence, the lower stem must also be displaced. As the lower stem is already a candidate, this effectively combines the bulge sequence and lower stem sequence into a single candidate target for a *cis*-repressing element to complement.

Complementing a hairpin can be problematic as the hairpin competes with the *cis*-repressing element, leading to a structural ensemble containing both conformations. This would have consequences for the first objective as the members of the ensemble in the native conformation would lead to partial repression of the output promoter this is an argument against the bulge-lower stem target. Of the, now three, potential targets for the *cis*-repressing element, the two within the dCas9 binding scaffold (nexus and bulge-lower stem) have a fixed sequence. The spacer sequence on the other hand is variable depending on the gRNA target. If the *cis*-repressing element complements part of the dCas9 binding scaffold, this sequence will be non-variable between versions, as is true for the

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

asRNA, which would complement the *cis*-repressing element. Accordingly, attempts to create functionally comparable versions which were nonetheless orthogonal would be hampered (third objective). The Spacer has already been successfully used as a target for gRNA inactivation (Lee, Hoynes-O'Connor, Leong, & Moon, 2016; Tang, Hu, & Liu, 2017). The combination of all these factors resulted in the decision to choose the spacer as the target for the *cis*-repressing element.

Having selected the spacer as the target for the *cis*-repressing element the next stage was to choose whether to fuse the *cis*-repressing element to the 5'- or 3'-end of the transcript. The shorter the sequence distance between two complementing elements, the smaller the effect of entropy and more stable the interaction (Suker & Stiegler, 1981). Consequently, it was decided to fuse the *cis*-repressing element to the 5' end of the gRNA transcript.

The Toehold was chosen as the method of interaction initiation between the asRNA and the crgRNA. Increasing the length of the toehold increases the rate of heteroduplex formation between the asRNA and the crgRNA as well as stabilises the heteroduplex through the increased number of complementary bases.

Consequently, increasing toehold length increases the proportion of molecules within the RNA structural ensemble that are in a heteroduplex and therefore active in terms of reporter repression (second objective). As the length increases however, the potential for non-cognate interactions also increases (third objective). To balance these competing considerations a 12 nt toehold was chosen, because it has been observed that the effect of increasing toehold length on heteroduplex formation tends towards saturation at a length of 12 nt (Green, Silver, Collins, & Yin, 2014).

2.2.3 Design and synthesis of crgRNA

The design decisions taken so far have resulted in a mechanism in which a 5' *cis*-repressing element complements part of the spacer region of the gRNA as well as an asRNA, which, when present, is then able to interact with the crgRNA through

complementing the 5' toehold and displacing the *cis*-repressing element from the spacer. This in turn allows the crgRNA to complex with dCas9 and repress the output promoter. A balance must be struck between the need for the *cis*-repressing element to inactivate the crgRNA and the need for the crgRNA, when expressed with the asRNA, to produce a high fold repression of the output promoter (the first and second objectives). A *cis*-repressing element that folds with a low ΔG , forming a stable structure reduces the probability of the crgRNA repressing the output promoter in the absence of the asRNA. Correspondingly, it may reduce the efficiency of asRNA binding and activation of the crgRNA leading to a low fold change in the output promoter when expressed with the asRNA. The inclusion of bulges in the *cis*-repressing element's complementation of the scaffold reduces the stability of the *cis*-repressed conformation and increases the $\Delta\Delta G$ of asRNA binding. Consequently, a library of variants with different thermodynamic stabilities was created (naming is alphabetic, in order of stability). These ranged from designs such as variant N (Figure 2.3A) where stability is reduced through the addition of several bulges, to variant J (Figure 2.3A) in which stability is increased through extending complementarity into the Cas9 binding scaffold. Stabilities ranged from -36.95 kcal/mol (N) to -60.10 kcal/mol (J).

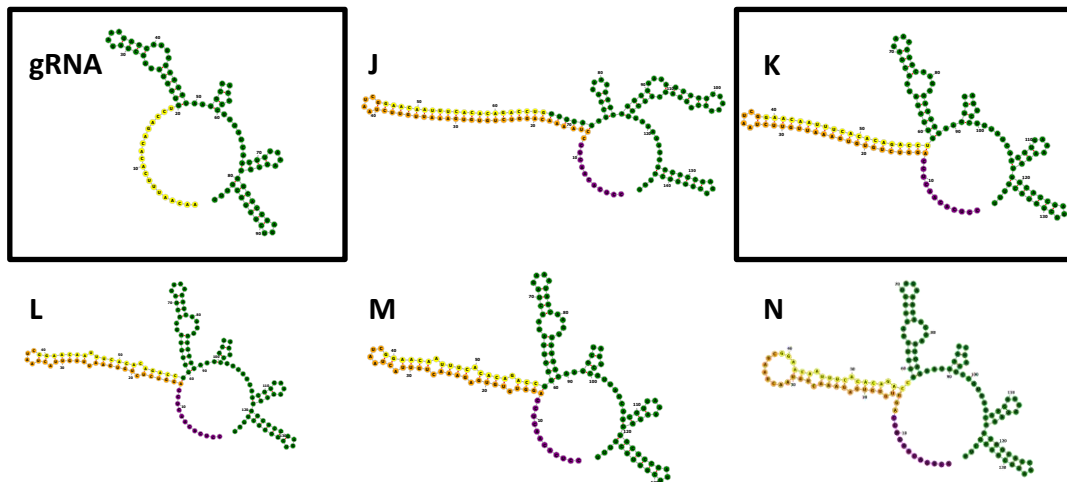
The inducible P_{Lac} promoter was chosen to express the crgRNA as it was also used in the most comparable work to this by Lee et al., 2016 to express gRNA. To avoid issues with leaky expression of the asRNA; the asRNA was expressed from a constitutive promoter, which was included or excluded from the construct to observe the resulting change in repression of the output promoter. The set of constructs, designed using these criteria (Figure 2.3A), were sent for synthesis at IDT. A problem arose when the long hairpins included in each of the constructs led to difficulties with synthesis, requiring multiple synthesis attempts leading to delayed delivery or synthesis failure. These challenges meant only the '+asRNA' and '-asRNA' versions of the K variant were received, as well as the asRNA-only control and gRNA-only control (see Figure 2.3A).

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

After discussions with staff at the synthesis facility it appeared that a crgRNA with a 3' *cis*-repressing element would be easier to synthesise. Consequently, variants in which the *cis*-repressing element was included in the 3' terminus were designed. Variants W, X and Y used *cis*-repressing elements from N, M and K respectively (Figure 2.3). To prevent torsion issues resulting from changing the *cis*-repressing element terminus, the toehold was moved from the 5' to 3' terminus of the *cis*-repressing element.

The synthesised library of constructs were delivered in the very high copy plasmid pIDTSmart-Amp (IDT Inc.). Initial attempts at transforming this very high copy plasmid, produced colonies which could not be cultured in liquid media. This may have been due to toxicity caused by the high copy number plasmid combined with the strong constitutive promoter (J23119) used to express the asRNA. Consequently, when the 3' *cis*-repressing element crgRNAs were sent for synthesis, the strong constitutive promoter in the constructs sent for synthesis was replaced with a slightly lower strength constitutive promoter (J23107, relative strength 0.70). This second synthesis round was more successful, yielding both '+asRNA' and '-asRNA' versions of two of the variants; W and Y, as well as a control in which nonsense RNA had been substituted for the 3' *cis*-repressing element of the same length. This additional control was included to observe changes in gRNA efficacy resulting from the inclusion of extra sequence in the 3' terminus. Complications which had arisen with the transformations of plasmids from both rounds of synthesis in the high-copy IDT vector, so constructs were sub-cloned into the lower copy pBR322 vector (expounded in methods and materials). This change of vector removed all observable toxicity effects and allowed the constructs to be expressed with the other two plasmids required for the experimental setup.

A -5' cis-repressing elements



B -3' cis-repressing elements

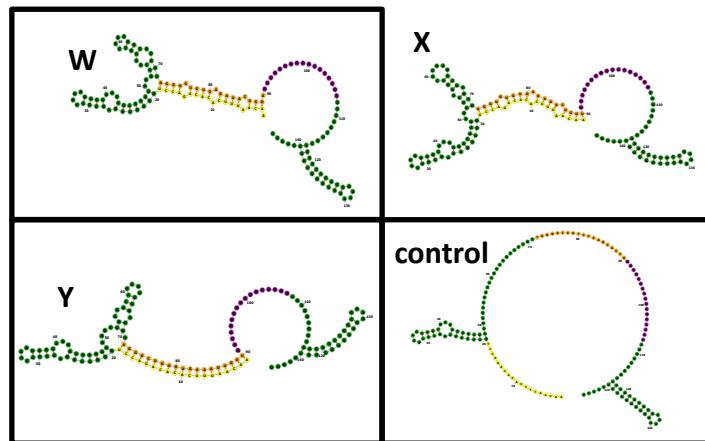


Figure 2.3 Structures of crgRNA designs. Green: dCas9 binding scaffold, Yellow: DNA binding spacer region, Orange: *cis*-repressing element, Purple: toehold. Designs are divided up in to two categories: 5' in which the *cis*-repressing element is placed upstream of the spacer region, and 3' in which the *cis*-repressing element and toehold are added between the final hairpin and the terminator. In both cases the *cis*-repressing element complements principally the spacer. Due to issues with synthesis, only those designs within boxes completed synthesis. Structures above include a control gRNA with no *cis*-repressing element and a control with nonsense RNA inserted in the site used for the 3' *cis*-repressing element to test for deleterious effects on gRNA functionality. Structures visualised using Forna (Force-directed RNA, Kerpedjiev, Hammer, & Hofacker, 2018)

2.2.4 Initial testing of crgRNA variant K

The first variant to be successfully generated and tested was variant K. To test it, crgRNA K was expressed with and without the cognate asRNA, in an *E. coli* strain also expressing dCas9 and the mCherry which is the target of the crgRNA transcriptional repression as per Figure 2.2. The results of this initial experiment are shown in Figure 2.4 as mCherry fluorescence, relative to the average fluorescence of the control strains which expressed mCherry and dCas9, but no gRNA or asRNA. If the system functioned as anticipated, crgRNA expression without asRNA would not result in reduced mCherry fluorescence, while crgRNA expression with asRNA would enable the crgRNA to repress mCherry. Expression of crgRNA K without the cognate asRNA resulted in partial repression of mCherry fluorescence; however the degree of repression was significantly greater when the crgRNA was expressed with the asRNA (two sample homoscedastic t-test $P=0.05$).

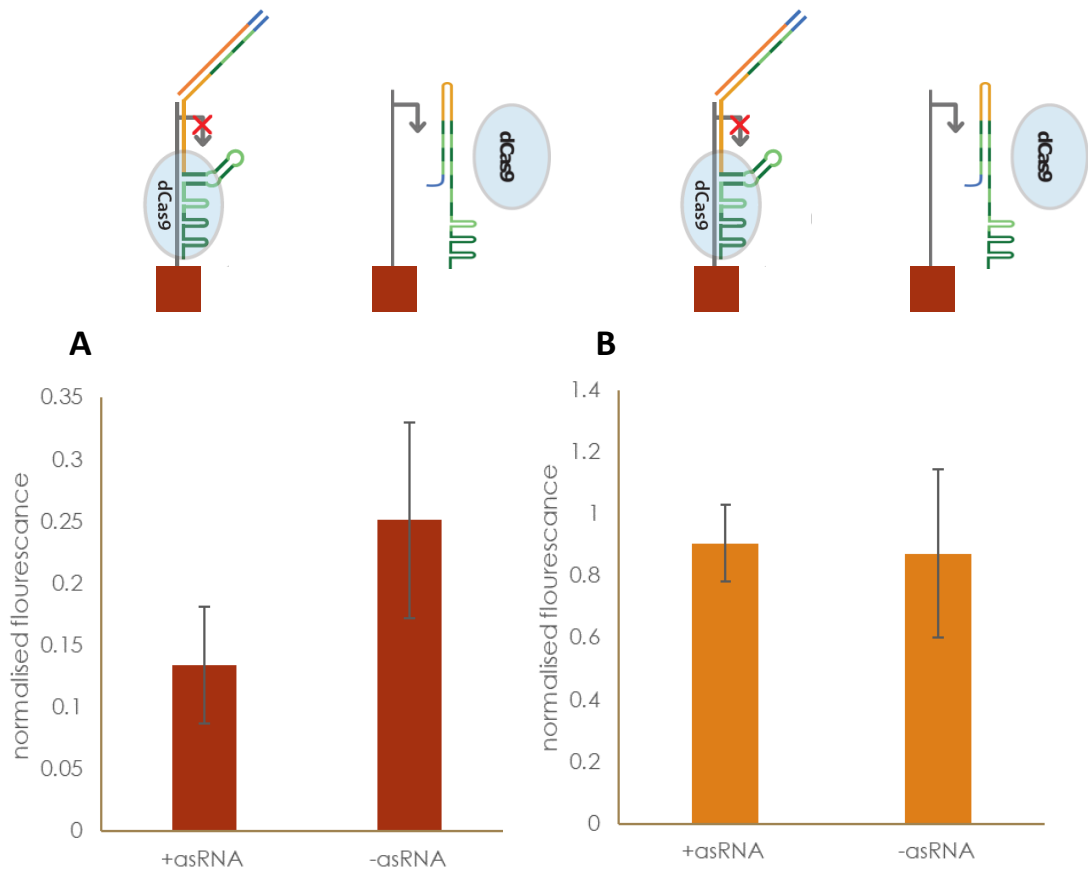


Figure 2.4. Preliminary test of K variant. (A) mCherry fluorescence (584/610-20), (B) YFP fluorescence (485/520). In both A and B, the level of fluorescence is normalised to a culture expressing the fluorescent protein without the crgRNA. A and B were read from the same cultures on different fluorescence channels. Error bars are the standard deviation of the sample. Inducer concentrations were 200 pg/ml aTc, and 250 μ M IPTG. The fold change in normalised fluorescence of mCherry was 1.86 fold (n=8). Full data processing methods found in section 5.3.1

2.2.5 Optimization of dCas9 and crgRNA expression levels

As it was observed in testing crgRNA K, that mCherry was partially repressed with expression of the crgRNA without asRNA, the efficacy of dCas9 in the presence of very small levels of aberrantly folded crgRNA was considered, and expression levels of dCas9 and crgRNA optimised.

The dCas9 plasmid used here was constructed in a study which demonstrated the high levels of repression that can be achieved with dCas9 (Qi et al., 2013). Although dCas9 expression is being induced at a low level with 200 pg/ml aTc, this is still in the upper limit of the dynamic range for dCas9-mediated repression activity (Qi et al., 2013). Therefore, if the levels of dCas9 being expressed are in excess of the level required, even a small proportion of the crgRNAs within the structural ensemble could be forming a structure that can complex with dCas9 even in the absence of the asRNA. This is true particularly if the crgRNA is being expressed at a level well above that required. Similarly, the crgRNA is induced at a level, used in another paper which may not be optimal for this system (Lee, Hoynes-O'Connor, Leong, & Moon, 2016).

As both the dCas9 and crgRNA K are under control of inducible promoters, for which expression level can be titrated with differing concentration of inducer, dCas9-mediated repression of mCherry was tested with different amounts of aTc (for dCas9 induction) and IPTG (for crgRNA induction). Again, strains were cultured for 24 hours and fluorescence from mCherry measured. The results from this experiment are shown in Figure 2.5.

As anticipated, higher levels of dCas9 and crgRNA induction led to greater repression of mCherry, both with and without asRNA expression. At all induction levels crgRNA +asRNA strains exhibited greater mCherry repression than crgRNA -asRNA confirming that the mechanism is working. However, the fold difference in repression between +asRNA and -asRNA varies. Both low (0 μ M IPTG) and high (250 μ M IPTG) expression led to low fold differences between +asRNA and -asRNA

cultures. The maximal fold change between +asRNA vs -asRNA was observed when the crgRNA was induced between the minimum and maximum, at 50 μ M IPTG. This is true at each level of dCas9 induction observed. High induction of dCas9 (200 pg/mL aTc) resulted in a low fold difference. As the level of dCas9 induction is decreased, so the fold difference increases, until it reached its maximum 3-fold when entirely un-induced. This demonstrates that in the dynamic range, observed crgRNA expressed without the activating asRNA is more sensitive to changes in level of dCas9 expression than crgRNA with the asRNA. It is notable however that the lowest level of expression in the dynamic range observed for dCas9 is still high enough to result in a 3 fold repression between the positive control and the +crgRNA, +asRNA Strain. This is likely to be due to the low level leaky expression previously observed using this system (Lee et al., 2016).

The result is that the dynamic range of dCas9 induction available from the inducible promoter only covers 25% of the dynamic range of the target promoter repression. This leaky expression of dCas9 prevents the characterisation of the logic gate at lower levels of dCas9 expression. Based on this data, the same three plasmid setup was used in successive experiments with an induction level of 50 μ M IPTG and 0 pg/mL aTc. The correlation between reduced level of dCas9 expression and higher fold change may be extrapolated to hypothesise that reducing the level of dCas9 expression further would lead to a greater fold difference between the +asRNA and - asRNA states. The very high expression levels seen in this system presently result in repression of the target promoter with the crgRNA when the asRNA is not expressed. This hypothesis was explored in Section 2.2.11 following production of different variants of the dCas9 expressing plasmids, modified for lower levels of expression.

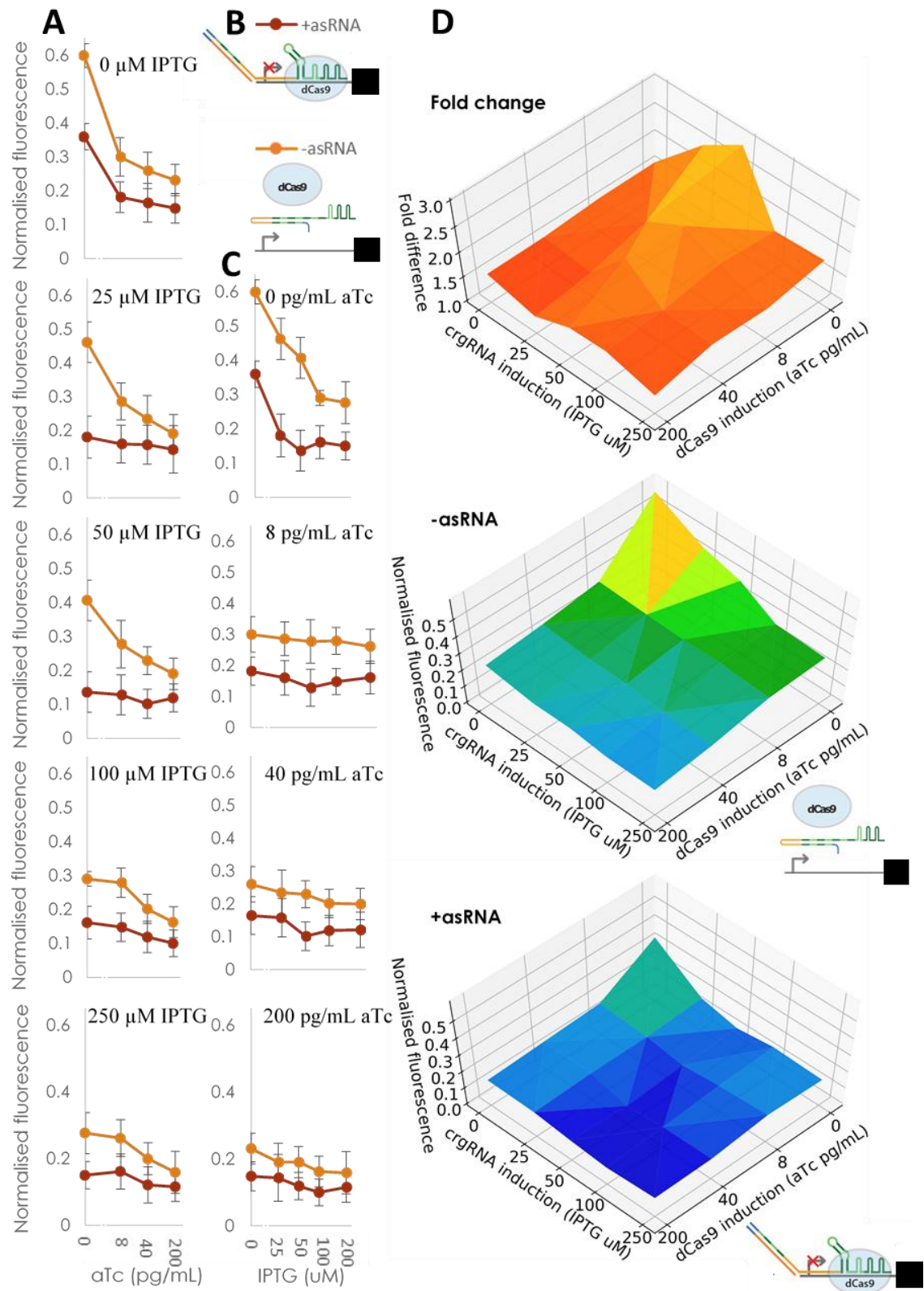


Figure 2.5 Please see figure legend on next page

Figure 2.5 Testing for optimal induction levels of dCas9 and crgRNA for asRNA dependent crgRNA repression of mCherry. Each point represents cultures grown for 24 hours in the level of the two inducers specified. Results are presented in a number of ways; in a series of slices with normalised fluorescence on the y axis, one inducer on the x-axis and the other increasing slice by slice to allow the presentation of error bars (A,B) . Data are presented with each inducer on the x axis. The more intuitive way the level of repression is represented is as the three-dimensional, triangle plots (right). Both -asRNA and +asRNA show the greatest repression at the highest level of dCas9 and crgRNA induction, and the least repression in an un-induced state. The fold difference between the +asRNA and the - asRNA varies across the induction space and is portrayed in the uppermost triangle plot (D). Error bars are standard deviation of sample. n=24.

2.2.6 Testing crgRNA variants K, W and Y with optimized dCas9 and crgRNA expression levels

Following the optimisation of dCas9 and crgRNA K induction levels, the crgRNA variants W and Y (in which the *cis*-repressing element is at the 3' end of the crgRNA, Figure 2.3) were additionally tested in parallel with crgRNA K. to control for changes in repression efficiency resulting from the inclusion of the 3' *cis*-repressing element, a gRNA control and a gRNA with a nonsense 3' insert were also included, both with the same spacer region and target as the crgRNAs. Strains were grown in LB for 24h and mCherry fluorescence measured (Figure 2.6). No statistically significant changes in optical density of cultures were observed for any of the variants either in the +asRNA or -asRNA state (Figure 2.6B), giving no evidence of toxicity from any of the three variants irrespective of asRNA presence.

There are a number of differences between the performance of the three different variants. When comparing variants W and Y, the mCherry repression in the absence of asRNA is 44% repression for W, and 21% repression for Y. Variant W displays a statistically significant ($p=0.05$) 1.1-fold difference between the +asRNA and -asRNA states. This is a smaller fold change than the 1.2-fold change ($p=0.05$) seen in the Y variant. Variation in inhibition resulting from the two different *cis*-representing elements can also be seen in the variation of change observed when the introduced asRNA negates or partially negates the effect of the *cis*-repressing element. The greater degree of inhibition seen from the Y *cis*-repressing element means a greater change in repression is seen when that effect is negated through the addition of an asRNA.

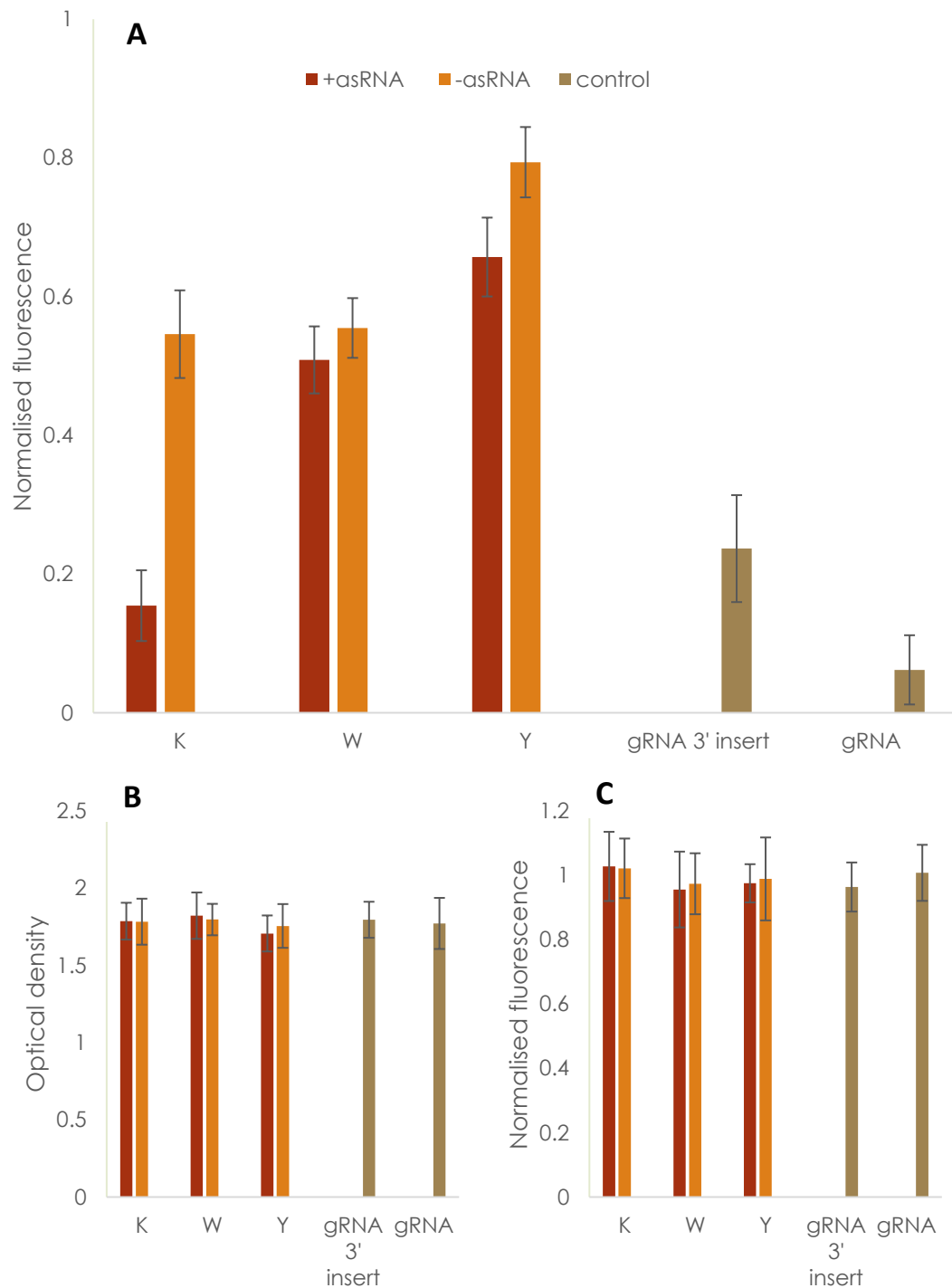


Figure 2.6 Testing the performance of all synthesised designs. (A) Normalised fluorescence of reporter mCherry showing repression resulting from crgRNA -asRNA and crgRNA +asRNA. Controls include a gRNA and a gRNA with 3' insert to observe the effect of adding the cis-repressing element and as a comparison with the crgRNA + asRNA to ascertain the level of functional rescue by the expression of the asRNA. All variants show greater repression when the crgRNA is activated by the asRNA. (B) Optical density of cultures as a measure of toxicity; no statistically significant differences between samples and controls were observed. (C) Normalised fluorescence of a control fluorescent protein (YFP) being expressed in the same cultures. No statistically significant differences between samples and controls were observed. Error bars are standard deviation n = 24. Full data processing methods found in section 5.3.1.

There is variation between the two crgRNAs with the *cis*-repressing element at the 3' end when in their active (+asRNA) state. Y +asRNA (34% repression) does not lead to the same level of repression as W +asRNA (49% repression), which in turn does not repress to the same level as the gRNA with control 3' sequence (76% repression). Both of these observations imply that the expression of the asRNA is not entirely negating the effect of the *cis*-repressing element (objective B) and so the more stable Y crgRNA is still leading to a greater degree of *cis*-repression. In addition, neither W +asRNA nor Y +asRNA (49% and 34% repression) approach the level of repression seen in the gRNA with a control 3' element (76% repression). A similar effect is seen in the K variant (which has the *cis*-repressing element at the 5' end), where the +asRNA (84% repression) strain shows less repression than the gRNA control with the same target and spacer region as the crgRNA (93% repression). However the difference between the + asRNA state and the gRNA control is smaller for the 5' variant (84%: 93% repression) when compared with the 3' variants (49%-34%: 76% repression). There is a difference in percentage rescue upon the addition of the asRNA between the 5' and 3' system variants; the K variant (5' *cis*-repressing element) expressed a percentage rescue of 81%, the 3' Y and W (3' *cis*-repressing elements) expressed rescue percentages of 19% and 9%. No significant differences were observed for the control fluorescence (YFP) (Figure 2.6C) suggesting the changes what changes are not the result of a global metabolic change in the strains.

The addition of the 3' element leads to a reduction in the level of repression seen even when the 3' element doesn't contain a *cis*-repressing element (3.8-fold change). This extra factor, reducing the activity of the inactive Y variant may explain why the observed repression exerted by inactive Y is lower than that of the K variant. Each of these factors suggested that the optimal level of expression for the crgRNA and dCas9 may be different for the two 3' variants and so a second induction level optimisation experiment was conducted.

2.2.7 Testing dCas9 dependency of reporter repression

The sequences used in each of the crgRNA designs contain sequences complementary to the sense strand of the mCherry promoter, and also to the 5' terminus of the mCherry reporter mRNA transcript. Therefore, it is necessary to check for any RNA-RNA interactions between the crgRNA and the 5' UTR of the reporter transcript. These include interactions which might either stabilise the transcript, lead to transcript degradation, alter transcription through association with the nascent RNA strand, or affect the rate of translation of the transcripts by ribosomes. Understanding the mechanisms underlying the observed modulation of reporter expression is necessary for the design, build, and test cycle.

To test for any modulation of the mCherry reporter by RNA-RNA interactions rather than dCas9 mediated interactions with DNA, strains were generated in which the crgRNA (with and without asRNA) and mCherry were expressed without dCas9. The observed reporter fluorescences are shown in Figure 2.7. A one-way ANOVA yielded no statistically significant differences between the different strains. This leaves no evidence of direct RNA-RNA interactions modulating reporter expression and no evidence of non-dCas9 mediated, crgRNA dependent modulation of reporter expression.

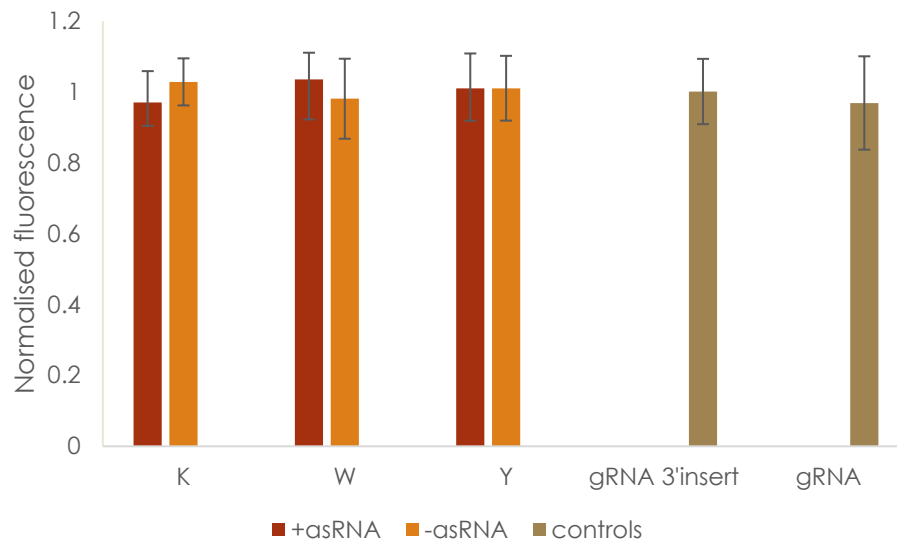


Figure 2.7 Normalised fluorescence of reporter when crgRNA are expressed without dCas9. No statistical differences were observed confirming the dCas9 dependency of crgRNA modulation of reporter expression. Error bars are standard deviation n = 24. Induction level was 50 μ M IPTG (crgRNA). Full data processing methods found in section 5.3.1.

2.2.8 Optimization of induction levels for crgRNA Y variant

Of the two crgRNA variants with 3' *cis*-repressing elements, crgRNA Y exhibited the greatest fold difference in mCherry fluorescence between +asRNA and –asRNA strains. This variant was therefore taken forward for optimisation.

The low fold repression seen in the +asRNA condition for variant Y, is problematic for the second objective (the output should be OFF when the crgRNA and asRNA are both expressed). Previous optimisation showed that crgRNA-mediated mCherry repression could be enhanced by increasing expression of either dCas9 or the crgRNA. For the fold difference between the -asRNA and +asRNA strains to increase, the +asRNA condition would have to exhibit a greater increase in repression than the –asRNA when the induction level of either the crgRNA or the dCas9 is increased. Optimisation was conducted, using higher levels of inducer than the level used for the full library in Figure 2.8. While the increase in the two induction levels led to increases in repression by the crgRNA Y in the active (+asRNA) state, they also led to increases in repression by the crgRNA Y in the inactive -asRNA state. Thus failing to provide an increase in fold change between the two states. Therefore, the original induction level of 0 pg/mL aTc and 50µM IPTG is used in succeeding experiments.

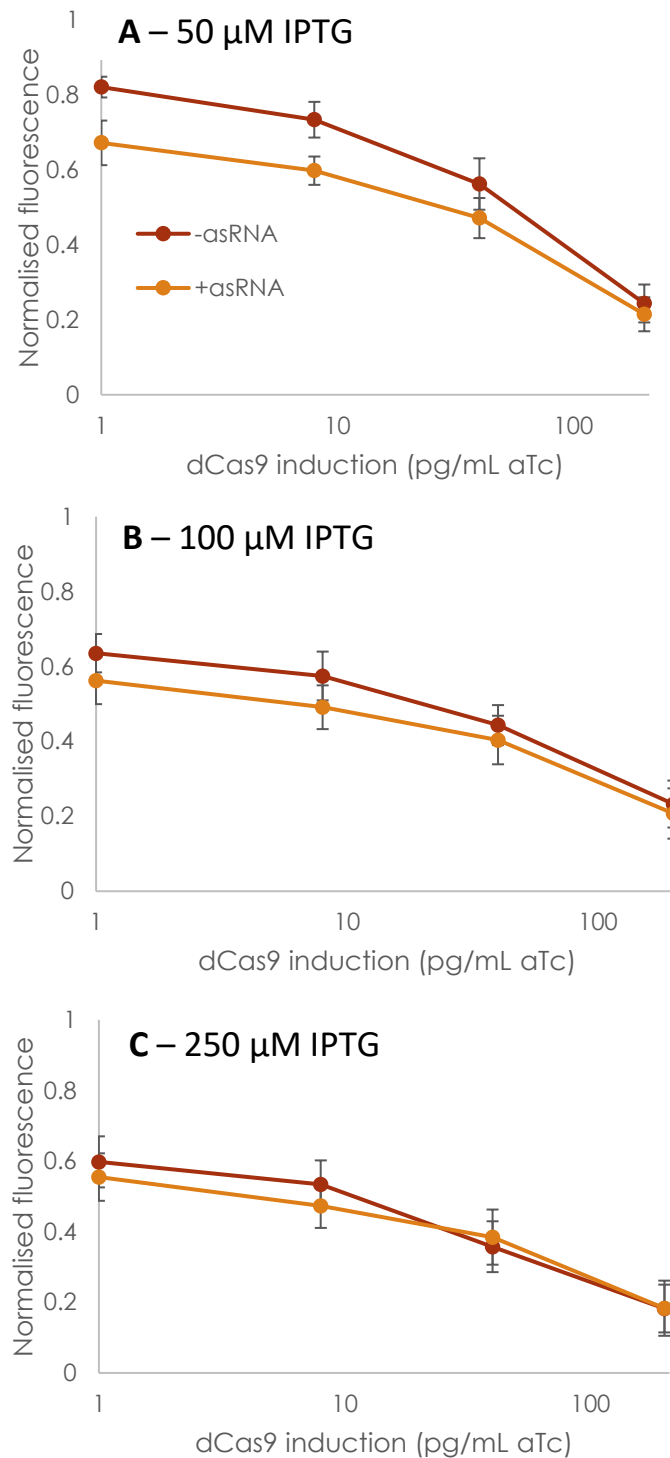


Figure 2.8 Level of repression of mCherry by the 3' Y crgRNA variant in the +asRNA and -asRNA states. Fold difference is greatest at 0 aTc and 50 μ M IPTG. Repression observed at four different aTc concentrations and three different (A,B,C) IPTG concentrations. Induction concentrations of 0 are changed to 1 so they may be presented on a log scale. Full data processing methods found in section 5.3.1. Error bars are standard deviation (n=24)

2.2.9 Effect of Temperature on repression

Changing temperature is known to result in changes in the RNA structural ensemble (Kortmann & Narberhaus, 2012), therefore the effect of change in temperature on the functionality of each of the three crgRNA variants was explored. Strains were grown at either 20 °C, 30 °C or 37 °C in LB media with 200 rpm shaking for 24 hours. The results are shown in Figure 2.9.

Results from cultures grown at 37 °C reproduced the results of the original test (also carried out at 37 °C). The results at 30 °C showed little variation relative to 37 °C. At 20 °C, the crgRNA K variant with 5' *cis*-repressing element performed in a similar manner as seen at 37 °C. The two variants with 3' *cis*-repressing elements however both resulted in complete repression of the fluorescent reporter protein at 20 °C with or without asRNA.

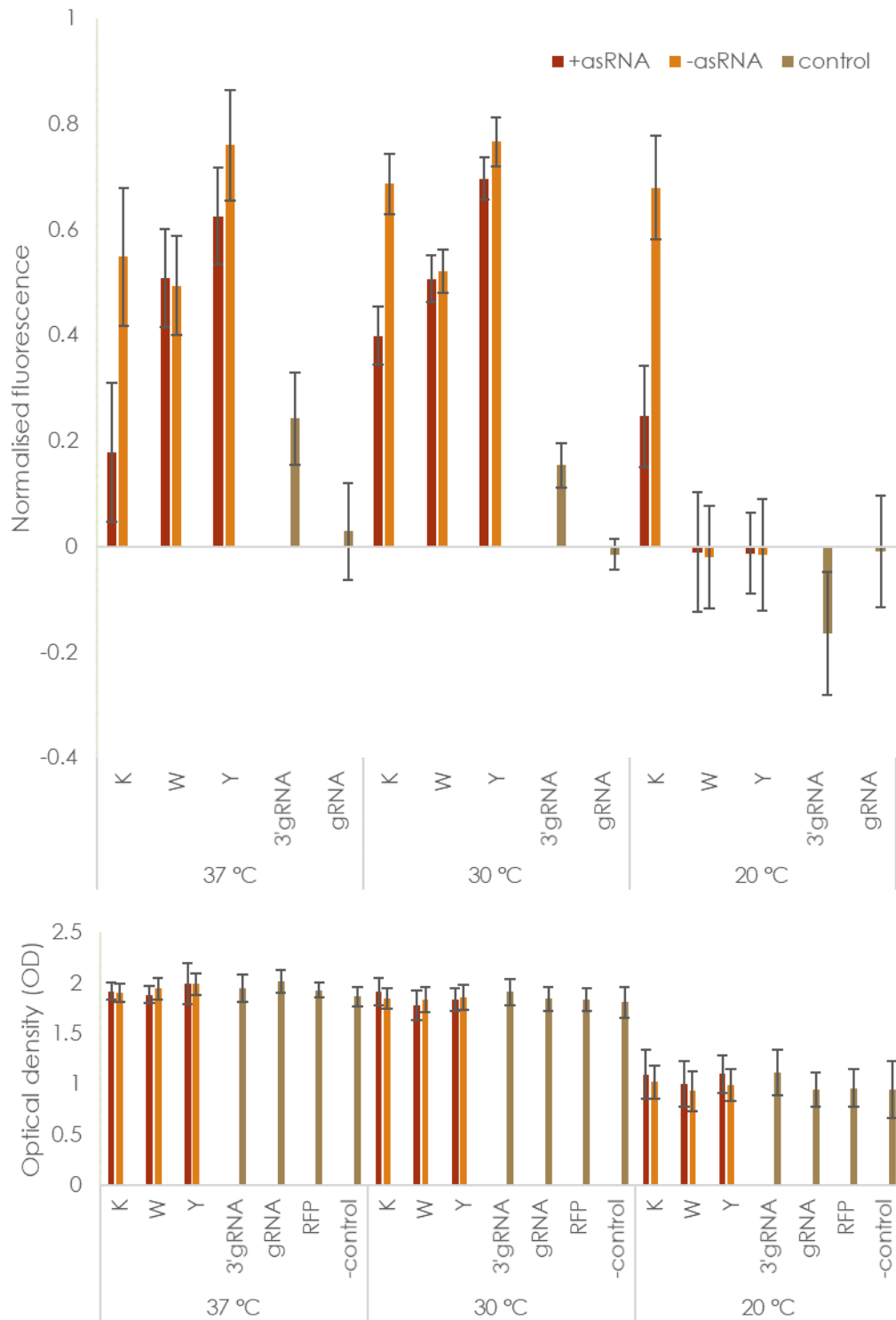


Figure 2.9 Each crgRNA variant (K,W,Y) expressed at three different temperatures. Normalised fluorescence shows repression by the crgRNA/asRNA system. Controls include a gRNA and a gRNA with 3' insert. OD of cultures is also included. Error bars are standard deviation. n = 24. Induction level of dCas9 is 0 pg/mL aTc, induction level of crgRNA is 50 μ M IPTG. Full methods found in section 5.3.1.

2.2.10 Time series measurements

In each experiment so far, fluorescence measurements have been taken after 24 hours of culturing. To explore whether the system behaves differently in each growth phase, a time series for strains expressing the K variant crgRNA both with and without the asRNA was carried out. Fluorescence and optical density was measured every 30 minutes for the 24-hour time course at 37 °C. The results are shown in Figure 2.10.

It is worth noting that during the lag phase and the early exponential phase, a small number of cells results in a very low level of fluorescence for all samples including positive and negative controls. This low level of fluorescence means that small random errors inherent within the plate reader results produce a very high margin of error in percentage terms for readings early within the time series.

The results show that the observations made in the stationary phase do not hold true for the exponential phase. In the exponential phase there still remains a difference between crgRNA functionality +asRNA and -asRNA, but both produced a lower level of fold repression than is seen at the 24-hour mark. As the cultures transition from exponential into stationary phases, the level of normalised fluorescence decreases for both the +asRNA and -asRNA states. But the decrease is more rapid for +asRNA cultures, which results in a greater fold difference between +asRNA and -asRNA states. Within the stationary phase the level of fluorescence in the -asRNA strain continues to drop. The +asRNA cultures on the other hand appear to reach maximum repression saturation point at approximately 20 hours. This contrasted with the continuing decrease in fluorescence resulting in the fold difference between +asRNA and -asRNA peaking before the 24-hour time point.

As the cultures transition from the exponential phase into the stationary phase the relative level of fluorescence appears to drop, in the +asRNA strains. This change appears to follow a sigmoidal curve reaching a saturation point at 20 hours. The -

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

asRNA cultures on the other hand continue to drop resulting in the fold difference between +asRNA and -asRNA peaking at approximately 22 hours.

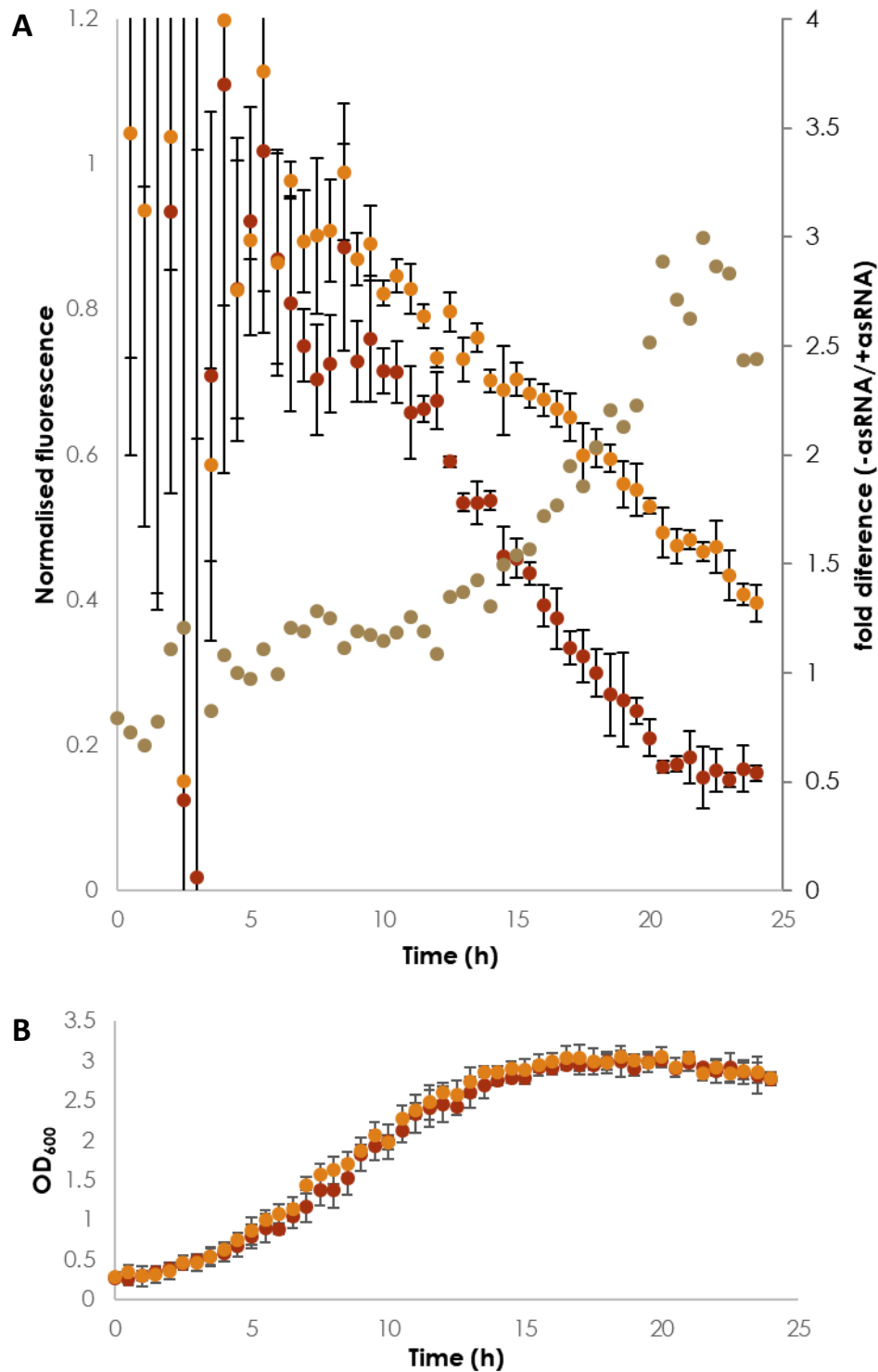


Figure 2.10 (A) Time series of repression by 5' K crgRNA variant in both +asRNA and -asRNA states. (A-primary Y axis) normalised fluorescence of mCherry expressed with crgRNA in either +asRNA (red) or -asRNA (orange) strains. Error bars are SEM (three repeats, each with 8 replicates) $n = 24$. (A-secondary Y axis) fold difference between +asRNA and -asRNA strains (brown). (B) Growth curve, OD₆₀₀ of cultures. Induction level for dCas9 is 0 pg/mL aTc. Level of induction for crgRNA is 50 μ M IPTG. Full data processing methods found in section 5.3.2.

To further characterise the growth phase dependent nature of the crgRNA functionality, time series data with was collected with a different dCas9 induction level. Having observed very low levels of repression during the exponential phase, to approach an optimum level of induction for this stage, increasing the level of dCas9 induction was explored. Again, fluorescence readings were recorded every 30 minutes. The results are shown in Figure 2.11.

No significant differences were observed in growth rate between +asRNA and – asRNA strains. Different repeats showed variation in the time point at which the fold change peaked, ranging from 17hrs to 21 hrs (Figure 2.11). This time range is earlier in the growth curve than the peak in average fold change with a lower level of dCas9 expression (Figure 2.10). A higher level of expression of dCas9 leads to fold change peaking at an earlier point in culture growth.

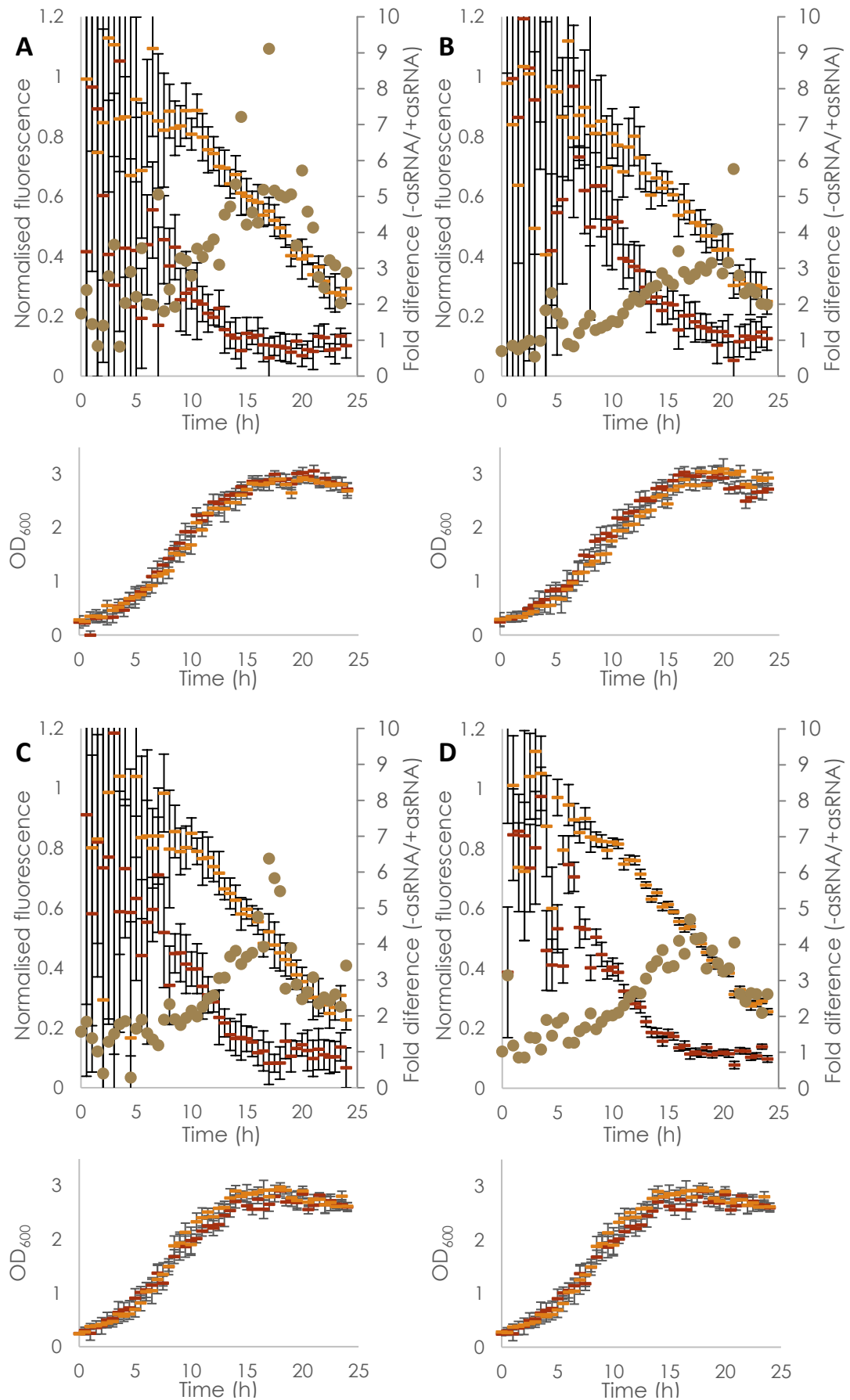


Figure 2.11 Please see figure legend on next page

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Figure 2.11 (A-D) Time series of mCherry fluorescence with repression by the crgRNA K variant either +asRNA (Red) or -asRNA (Orange) on the primary (left) Y axis and fold difference (brown circles) plotted on the secondary (right) Y axis. Each panel also includes a growth curve in which the OD₆₀₀ is plotted over time. (A-C) are repeats conducted on different days, error bars are standard deviation n= 8. (D) Is the combination of the three repeats, error bars are SEM, n=24. Each of the repeats are presented separately as well as combined to allow the observation of variation between repeats. Induction level for dCas9 was 2 ng/mL aTc, for crgRNA it was 50 μ M IPTG. Full data processing methods found in section 5.3.2.

2.2.11 Engineering the dCas9 Expressing Plasmid for Lower Expression

As highlighted in Section 2.2.8, the level of expression of dCas9 from the original AddGene plasmid #44249 may be above the optimum for this system, even when expression is entirely un-induced. Therefore, the next step was to reduce the level of expression by engineering the plasmid from which it was expressed. Two different approaches were taken to reducing the level of expression. The first was having a transcriptional attenuator between the promoter and RBS of the gene to make pdCas9-T (Figure 2.12). The attenuator used was a terminator from the Biobrick registry (BBa_B1003) with a termination efficiency of 83% (Huang, 2007). This reduces the level of expression from the promoter in both the induced and un-induced state. The fold induction for the pLtetO-1 promoter is >10 fold. Reducing the level of expression by less than 10 fold should allow characterisation of the system in an expression level range below that from the un-induced original plasmid. With the expression level from the induced pdCas9-T overlapping with the level of expression from un-induced original plasmid.

The original plasmid ribosome binding site had a predicted strength of 4662 AU (Espah Borujeni, Channarasappa, & Salis, 2014). Two alternative versions of the plasmid were made with edits to the RBS, to reduce the rate of mRNA translation (Figure 2.12). RBS 2 had a predicted strength of 4068 AU and RBS 3 had a predicted strength of 3618 AU. These two plasmids allowed for two eventualities; (A) the transcriptional attenuator reduces the level of expression so that even when fully induced a level of expression does not reach that of the un-induced original plasmid. In this case one of these two plasmids would be used to cover this part of the dynamic range. (B) A large decrease in the level of expression of dCas9 is not required because in this case one of these two plasmids would suffice to expand the dynamic range into a potential optimum.

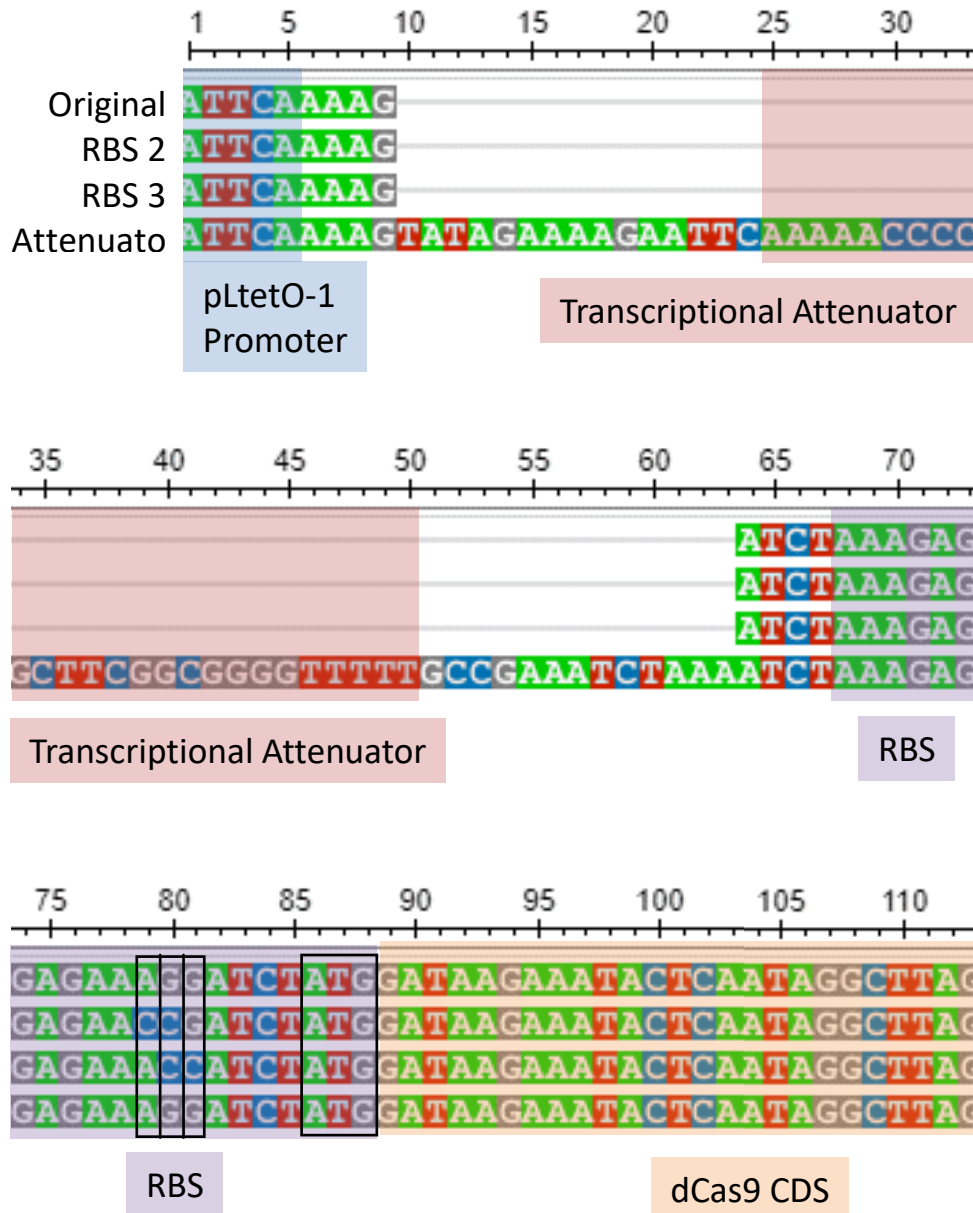


Figure 2.12 Alignment of the original dCas9 expression plasmid (AddGene plasmid #44249) with three variants engineered to reduce the level of expression of dCas9 either through the addition of a transcriptional attenuator (pdCas9-T) or through the mutation of the RBS sequence.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

The attenuator in pdCas9-T leads to very little repression of the reporter (though still statistically significant repression of $P < 0.05$). RBS2 does not result in a statistically significant change. RBS3 however does lead to a lower repression, statistically significant for the -asRNA condition.

In the same way the un-induced original plasmid had a higher level of induction than an apparent optimum level for fold change, conversely the un-induced pdCas9-T has a level of expression below the apparent optimum. Fortunately, the promoter can be induced to increase the level of expression, hence the need for a second induction level optimisation.

The level of induction of the crgRNA was varied as well as the level of induction for dCas9. In the previous optimisation, five different crgRNA induction levels (IPTG) and four different dCas9 (aTc) induction levels were tested. As it is dCas9 expression which has been reengineered five different induction levels of dCas9 (aTc) were tested, correspondingly the number of induction levels for the crgRNA (IPTG) was reduced to four. These results are shown in Figure 2.14.

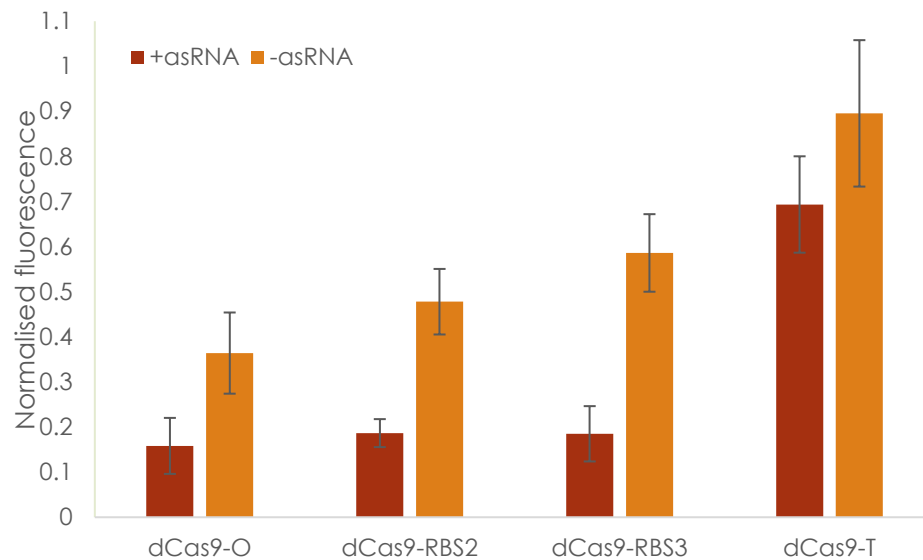


Figure 2.13 Varying dCas9 expression through two different approaches; the addition of a transcriptional attenuator (pdCas9-T) or a change in RBS strength. There is a substantial and statistically significant drop in the level of repression by the plasmid containing a transcriptional attenuator. The fluorescence measured is mCherry. The original AddGene plasmid #44249 is marked as dCas9-O. RBS 2 has a strength of 4068 AU, RBS 3 has a strength of 3618 AU. The plasmid with a transcriptional attenuator engineered into it is marked as pdCas9-T. The transcriptional attenuator has been experimentally characterised as having a termination efficiency of 83%, though this is partially context dependent. In all cases it is the K variant of the crgRNA/asRNA system which is used. As with previous experiments the levels of induction are 50 μ M IPTG (crgRNA) and 0 pg/mL aTc (dCas9). Error bars are standard deviation, n=24. Full data processing methods found in section 5.3.1.

At 400 pg/mL aTc induction with pdCas9-T, there is a comparable level of repression to that seen with the original dCas9 expressing plasmid (Figure 2.13, Figure 2.5). This is true in both the +asRNA and -asRNA states and at each of the IPTG induction levels observed. This is evidence that the level of dCas9 expression at 400 pg/mL aTc is comparable to the level of expression from the original plasmid when un-induced.

When performing the induction level optimisation with the original plasmid, the variation in the level of repression appeared to follow a log-exponential curve. With pdCas9-T, the dynamic range of the dCas9 promoter (aTc) covers a greater proportion of the dynamic range of repression (from 86% to 17% repression, -asRNA). With this greater dynamic range available the curve that can now be plotted appears to be log-sigmoidal rather than log-exponential.

The fold difference between the crgRNA +asRNA and -asRNA varies with the level of induction of both the dCas9 (aTc) and the crgRNA (IPTG). As discussed earlier in Section 2.2.11 the level of repression with pdCas9-T at 400 pg/mL aTc is comparable to that of the original plasmid when un-induced (5 fold). Correspondingly, the highest fold difference at this level of dCas9 induction is at 50 μ M IPTG (crgRNA) induction. The next lower dCas9 induction level (200 pg/mL aTc) shows the new peak fold difference. At this point the peak fold difference moves from being at 50 μ M IPTG (crgRNA) to 100 μ M IPTG with a very similar result at 250 μ M IPTG (10 fold). When either the crgRNA or dCas9 are un-induced there is a small fold difference between +asRNA and -asRNA. This is also true for high levels of induction of dCas9.

The fold difference between +asRNA and -asRNA correlates with the dCas9 repression/induction gradient (Δ Repression/ Δ induction). Where small changes in induction lead to large changes in the level of repression a greater fold difference is seen between +asRNA and -asRNA. For example a change from 100 μ M IPTG, 200

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

pg/mL aTc to 100 μ M IPTG, 400 pg/mL aTc leads to a fold change from 10 to 4 but the successive change 100 μ M IPTG, 800 pg/mL aTc results in a 4 to 3 fold change.

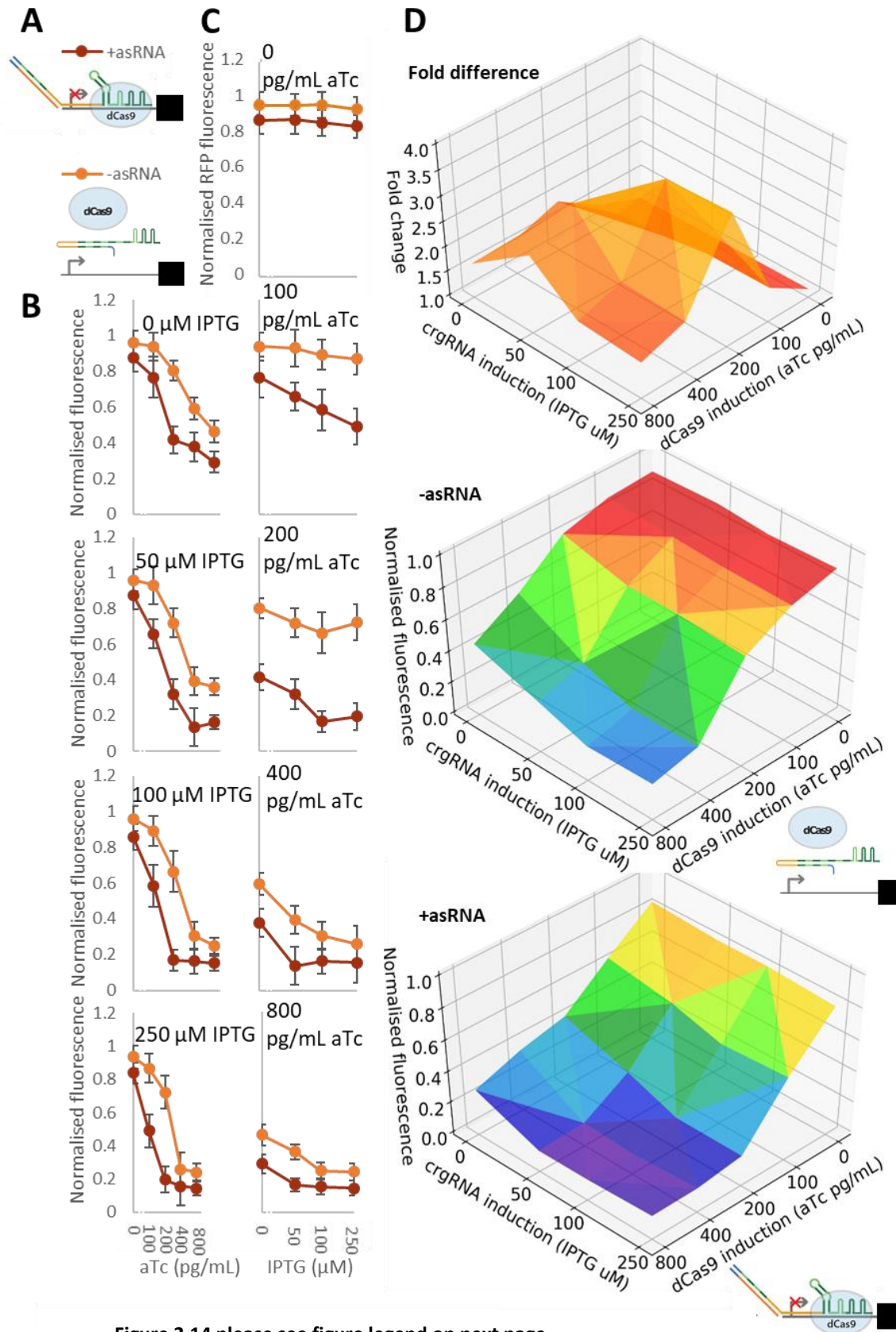


Figure 2.14 Testing for optimal induction levels of dCas9 and crgRNA while expressing dCas9 from pdCas9-T. All fluorescences measured are mCherry. Results are presented in a number of ways; in a series of slices with normalised fluorescence on the y axis and one inducer on the x-axis and increasing slice by slice to allow the presentation of error bars. Data are presented with each inducer on the x axis. The more intuitive way the level of repression is represented is in the three-dimensional, triangle plots (right). Both -asRNA and +asRNA show the greatest repression at the highest level of dCas9 and crgRNA induction, and the least repression in an un-induced state. The fold difference between the +asRNA and the - asRNA varies across the induction space and is portrayed in the uppermost triangle plot. Error bars are standard deviation of sample. n=24.

2.3 Discussion

The aim was to construct a RNA based logic system, in this chapter is laid out a proof of concept for the system and a detailed characterisation of the preliminary system, laying out strengths as well as points for improvement. Here, the main results that will be taken forward from this chapter into the next chapter will be explored, leaving an overarching compare and contrast discussion between this work and the broader field to the final Thesis Discussion chapter. The results seen here will be compared with those from other published systems.

2.3.1 Proof of Concept

Before producing the results laid out in this chapter the underlying mechanism explored here was entirely hypothetical. There was an original hypothesis: that it would be possible to design a gRNA with a *cis*-repressing element included in the transcript that would inactivate it, and that this element itself could then be sequestered through the expression of an asRNA to rescue the crgRNA activity and repress expression of an output gene. On the evidence laid out in this chapter the null hypothesis that this is not possible can be rejected. Nevertheless, there are a number of areas for improvement to make the best possible logic gate for use by practitioners of synthetic biology.

2.3.2 Methodological Insights

The work presented in this chapter has produced a number of insights into the best methodology to use in practice when further exploring this type of mechanism and these have been taken forward into the next chapter. This includes the three plasmid system used in this chapter. Expressing the system from a low copy plasmid such as pBR322 yields a functional system lacking in the toxicity observed when expressed from the original high copy synthesis plasmid. The most commonly used dCas9 expressing plasmid (AddGene plasmid #44249) expresses dCas9 at a level above the optimum for this system through leaky expression when un-induced. Consequently, the engineered version of this plasmid (pdCas9-T) was taken forward with a transcriptional attenuator that produces dCas9 over a dynamic

range including the optimal level of expression of dCas9 for this system. So far as synthesis of the constructs is concerned, large hair pins, repeats, inverted repeats and strong secondary structure are required by these designs, but producing to these requirements is beyond the limits of some synthesis companies, which needs to be taken into account for future work, when selecting which synthesis companies to use when exploring these mechanisms.

2.3.3 dCas9 expression

The use of CRISPRi in bacteria has been explored in a range of papers each of which requires the expression of dCas9. The majority of these papers either use the original AddGene plasmid pdCas9-bacteria #44249 or the operon from this plasmid in a different back bone. A number of papers include a dCas9 induction curve (Lee et al., 2016a; Vigouroux, Oldewurtel, Cui, Bikard, & van Teeffelen, 2018) covering only 40% - 60% of the dynamic range of dCas9 induction (depending on system used). The induction of the pdCas9-T plasmid covers >74% of the dynamic range of dCas9. The plasmid was not used at high levels of induction due to the requirements of this work but the remaining dynamic range within the promoter (expression levels reaching saturation at 10 ng/mL) implying the dynamic range covered may be up to 83%. Alternatively, pdCas9-T may be used alongside the original dCas9 expressing plasmid to reach such a combined coverage of the dCas9 induction range. This engineered pdCas9-T, when made available, offering increased dynamic range may make future insight into the performance of systems at low dCas9 expression more easily attainable.

2.3.4 *Cis*-repressing elements at the 5' and 3' terminus

The repression is lower for the W crgRNA than for the Y crgRNA when expressed in the absence of the asRNA. This variation is intuitive based on the differences between the two structures. The *cis*-repressing element in Y contains a direct complement of the spacer region of the gRNA leading to a more stable structure ($\Delta G = -38.05$ kcal/mol). The W variant on the other hand, has a complementary region of the *cis*-repressing element containing a number of mismatches reducing

the stability of the *cis*-repressed structure ($\Delta G = -24.37$ kcal/mol). Consequently, there is a lower predicted partial pairing probability between the *cis*-repressing element and the spacer region for W, leading to a higher proportion of the structural ensemble to be in a conformation suitable for repressing the output promoter. Therefore, the difference in repression from the two crgRNAs in their inactive state can be understood in terms of structural stability.

The repression is greater for the K crgRNA than for the Y or W crgRNA when expressed with the asRNA. As described in Section 2.2.4; when synthesizing the 3' *cis*-repressing elements, the possibility of toxicity led to the decision to change the asRNA promoter to a weaker one than that used in the 5' variants. This may account for the difference in percentage functional rescue between 5' and 3' variants.

The *cis*-repressed structure of K is more stable than the *cis*-repressed structure of Y or W due to the effects of entropy on the long sequence between the *cis*-repressing element and its complementary target (structural predictions performed with Vienna2.0). Armed with this information one might anticipate that the inactive K variant would be more stably inactivated than the Y variant. Our observations indicate the contrary however: the inactive K variant retains more activity than the inactive Y variant. There is clearly an extra factor to be considered here, and this can be seen in the difference between the two controls. The addition of the 3' element leads to a reduction in the level of repression seen even when the 3' element doesn't contain a *cis*-repressing element.

When cultured at 20 °C instead of 37°C, the two variants with 3' *cis*-repressing elements both resulted in complete repression of the fluorescent reporter protein at 20 °C with or without asRNA. There are a number of potential reasons for this change including (i) Lower temperature leading to a change in the RNA structural ensemble resulting in a change in crgRNA/asRNA functionality. (ii) Metabolic changes resulting from lower temperature. (iii) Change in expression of RNA

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

chaperone proteins or expression of heat shock proteins. The data available is not sufficient to interpret the mechanistic underpinning of these results.

2.3.5 Design insight

While issues with synthesis resulted in not all of the intended designs being tested, the data from those designs which were tested still offer insight into how future designs might be optimised.

Based on the evidence within this chapter, it was decided to continue to develop the 5' *cis*-repressing element over the 3' *cis*-repressing element due to the enhanced performance of the K variant when contrasted with the W and Y variants. The K variant, when expressed without asRNA (in its inactive state) still shows some dCas9 repression activity at 27% repression (pdCas9-T, 250 uM IPTG, 200 pg/mL aTc). When expressed with the asRNA, the crgRNA (in its active state) shows only 12% less repression than the gRNA control (pdCas9-T, 250 uM IPTG, 200 pg/mL aTc). This emphasises that the inactivation of the gRNA is more in need of optimisation than the reactivation by asRNA. There were mismatches were included in the complementation between the *cis*-repressing element and the spacer region of the gRNA in a number of the design variants which failed to be synthesised. These mismatches would have stabilised formation of the heteroduplex in the active state and conversely destabilised the *cis*-repressing element. The spacer hairpin in the inactive state changes in the relative ΔG of each of the competing structures ($\Delta\Delta G$). As it is the inactivation of the gRNA that needs to be optimised, the next library should seek to further stabilise the interaction between the *cis*-repressing element and the gRNA, as well as further sequestering sequences necessary for complex formation between dCas9 and the crgRNA. A promising avenue of approach for this optimisation would be to extend complementarity of the *cis*-repressing element from the spacer into the gRNA scaffold.

2.3.6 Characterisation

In this chapter, the most thoroughly characterised and best performing design variant was the K variant. In this chapter it is demonstrated that the addition of the *cis*-repressing element to the gRNA transcript reduced the degree of CRISPRi repression of a reporter gene, and that the addition of a corresponding asRNA to complement the *cis*-repressing element rescued CRISPRi repression activity. This repression, rather than being a direct RNA-RNA interaction, was dCas9 dependent and no repression was seen in the absence of dCas9. Similarly, repression seen was the result of a specific interaction rather than a large-scale metabolic change produced by the system as can be observed by the lack of change in control fluorescent protein expression.

The performance of the system varies with the level of expression of dCas9 and with the level of expression of the crgRNA. The level of repression seen by the system follows a log-sigmoidal curve when graphed against the level of expression induction. In both these curves, it is the point at which the induction/repression gradient is at its highest that the greatest difference was seen between the +asRNA and -asRNA states. This is true of both dCas9 induction level and crgRNA induction level.

The results show that the observations made in the stationary phase do not hold true for the exponential phase. In both exponential phase and stationary phase +asRNA strains show greater repression than -asRNA strains but the fold difference between the two states is greatest in stationary phase. It was hypothesized that the variation in system performance between the growth phases and specifically the lower level of repression in the exponential phase when contrasted with stationary phase observed relates to a dilution effect and the act of DNA replication. Once bound, dCas9 will remain bound to its DNA target sequence until replication (Jones et al., 2017). Therefore, the rapid replication seen during the exponential phase disrupts the repressive effect of dCas9 as the target promoter continues to express until a new crgRNA/dCas9 complex can bind. This effect may be accentuated by the effects of dilution. At each round of replication, the expressed dCas9 protein is

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

divided between the two daughter cells which continue to grow and increase in volume leading to a dilution effect. This means that rather than accumulating as in the stationary phase, the concentration of dCas9 is continually diluted, resulting in a lower concentration of dCas9 and therefore a lower level of repression. Hence the higher levels of fluorescence observed. The level of induction of dCas9 can be tuned for optimal performance in the exponential phase or the stationary phase.

The K variant shows a statistically significant difference in repression between +asRNA and -asRNA strains at a range of temperatures from 20 °C to 37 °C.

Chapter 3: Optimisation of crgRNA and orthogonality

3.1 Introduction

In the previous chapter (Chapter 2:), three design variants of the crgRNA-asRNA were designed, and characterised. While the system demonstrated functionality, there was also substantial room for improvement. Two crgRNA variants in which the *cis*-repressing element was at the 3' end, and one variant with the *cis* repressing element at the 5' were synthesised and characterised. Of these, the 5' variant exhibited the best properties in terms of achieving the greatest fold change in repression upon the addition of the asRNA and was consequently further characterised. In this chapter, the insights gained from characterisation of the system using the 5' variant were used as the basis for construction of a new library of variants to further improve our crgRNA-asRNA RNA based biological logic gate system.

In the previous chapter is described the design and testing of a crgRNA that has 27% basal repression of a target reporter gene, in the absence of asRNA and leads to a maximal repression of 81% when activated by asRNA, under optimal levels of induction of the plasmids expressing crgRNA, asRNA and the Cas9 enzyme. It was anticipated that there would still be room for improving the fold change observed by the current design, by both reducing basal repression of target promoter by crgRNA in the absence of asRNA and increasing maximal repression by crgRNA in the presence of asRNA. Improving the maximal repression by the crgRNA in the presence of asRNA can be done by increasing the level of expression of dCas9. However, this would also lead to an increase in basal repression from the crgRNA in the absence of asRNA, which is undesirable. Hence, reduction of basal repression level by improving the inactivation of the crgRNA by its *cis*-repressing element was considered the more reasonable design space to explore, as it would allow the system to be used even at higher dCas9 expression levels, improving both the basal repression and the active crgRNA repression, simultaneously.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Accordingly, throughout the design process the focus was on optimising the *cis*-repressing element to further inactivate the crgRNA, including extending the area of complementarity from the spacer region into the Cas9 binding scaffold. A library of variants was synthesised and tested.

Having identified a good design for the *cis*-repressing element, the next steps were to optimise the system for orthogonality so that multiple logic gates can be expressed in the same cell without cross-reactivity. This was done using an *in silico* machine learning approach to optimise the toeholds for orthogonality. The final step was to demonstrate that this system can be applied to multiple different promoters to produce predictable and consistent results.

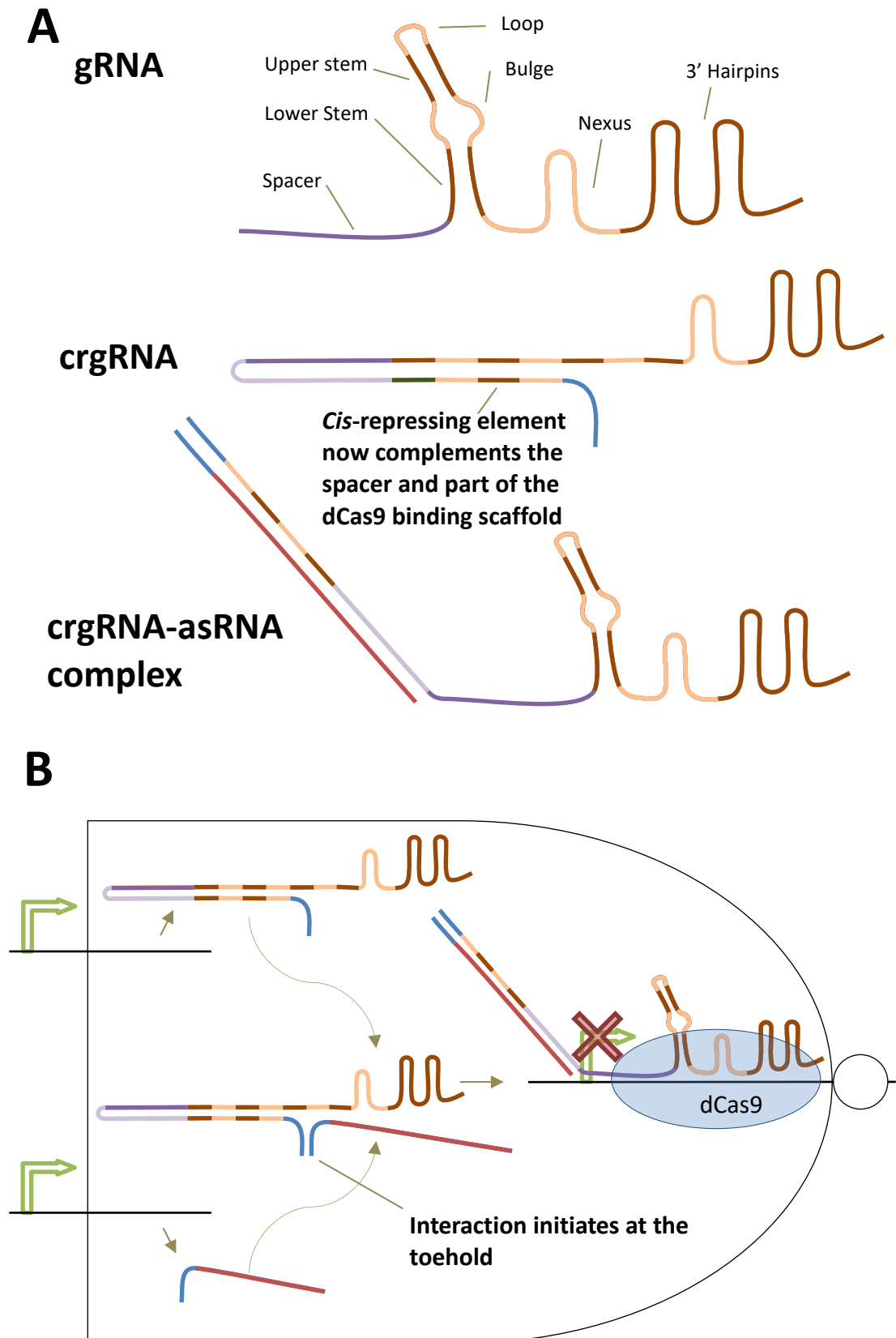


Figure 3.1 Please see figure legend on next page.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Figure 3.1 (A) Schematics of crgRNA with asRNA complexed and without. (B) System acting as NAND Logic gate. In both cases, colour changes through the Cas9 binding region of the transcript denotes different functional modules: Spacer in purple. *Cis*-repressing element colour matches that of the region of the transcript it compliments. Toehold in blue, asRNA in pink.

3.2 Results

3.2.1 Library Design

As discussed, the best way to improve the system is the further inactivation of the crgRNA through the *cis*-repressing element. There are two approaches to improving the *cis*-repression: the first is to further stabilise the inactive state through stabilising secondary structure formed between the *cis*-repressing element and the rest of the crgRNA. The second approach would be to increase inactivation of the crgRNA by further complementing and therefore sequestering other functional modules of the gRNA. As increasing the length of complementation of the *cis*-repressing element would both stabilise the *cis*-repressed structure and allow it to compliment other functional modules of the gRNA. The logical progression would be to extend complementarity from the spacer region into the 5' end of the Cas9 binding scaffold.

Mutation of the scaffold reveals that the 5' region contains a number of functional modules required for Cas9 binding and cleavage (Briner et al., 2014). The structure at the 5' end of the Cas9 binding scaffold is a hairpin formed by the combination of the crRNA and tracrRNA into a single transcript replacing the repeat-antirepeat duplex (Mali et al., 2013). It is formed of a number of modules, the lower stem, the bulge, the upper stem and loop sequence (Figure 3.1). Mutations on each of these modules have different consequences for Cas9 activity. This could be due to the prevention of complex formation between the Cas9 and its binding scaffold.

Alternatively, it is possible that as complex formation between dCas9 and the gRNA leads to substantial conformational rearrangements which are required for target DNA recognition, the mutated gRNA may lead to a gRNA-Cas9 complex which is inactive (Jiang et al., 2015; M. Jinek et al., 2014). When examining the change in efficacy resulting from mutation, module by module, it was discovered that when two of the 6 nucleotides of the lower stem are mutated to non-complementary bases, there is a drop of 45-100% in Cas9 cleavage activity. In the bulge module, there is a high tolerance for mutations including deletions which change the

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

sequence but retain the bulge within the secondary structure. However, if the bulge is removed through making the two sides complementary to one another, there is a 100% removal of cleavage activity. The upper stem and loop sequence can be deleted while retaining 52-81% activity.

To prevent repression of the target promoter by the crgRNA in the absence of the asRNA, a library of crgRNAs was created in which the region of complementarity within the *cis*-repressing element was extended. Each library member had the region of complementarity extended into the next successive module, extending from the spacer (Sp), first into the lower stem (LS), then into the bulge (Bg), the upper stem (US) and finally complementing the loop sequence as well (Lp) (Figure 3.2). It is the first two modules that have the greatest impact on complex activity in mutational studies using Cas9. Hence, the greatest stepwise improvement might be anticipated to be due to the complimentary of these two modules. However, the US and Lp variants may lead to a greater stabilisation of the crgRNA in its inactive state and it is hypothesised that the Lp variant will result in the greatest degree of *cis*-repression. In the previous chapter, the secondary structure of the crgRNAs were problematic for the synthesis company, consequently, a different synthesis company (GeneArt) was used to produce this library of variants.

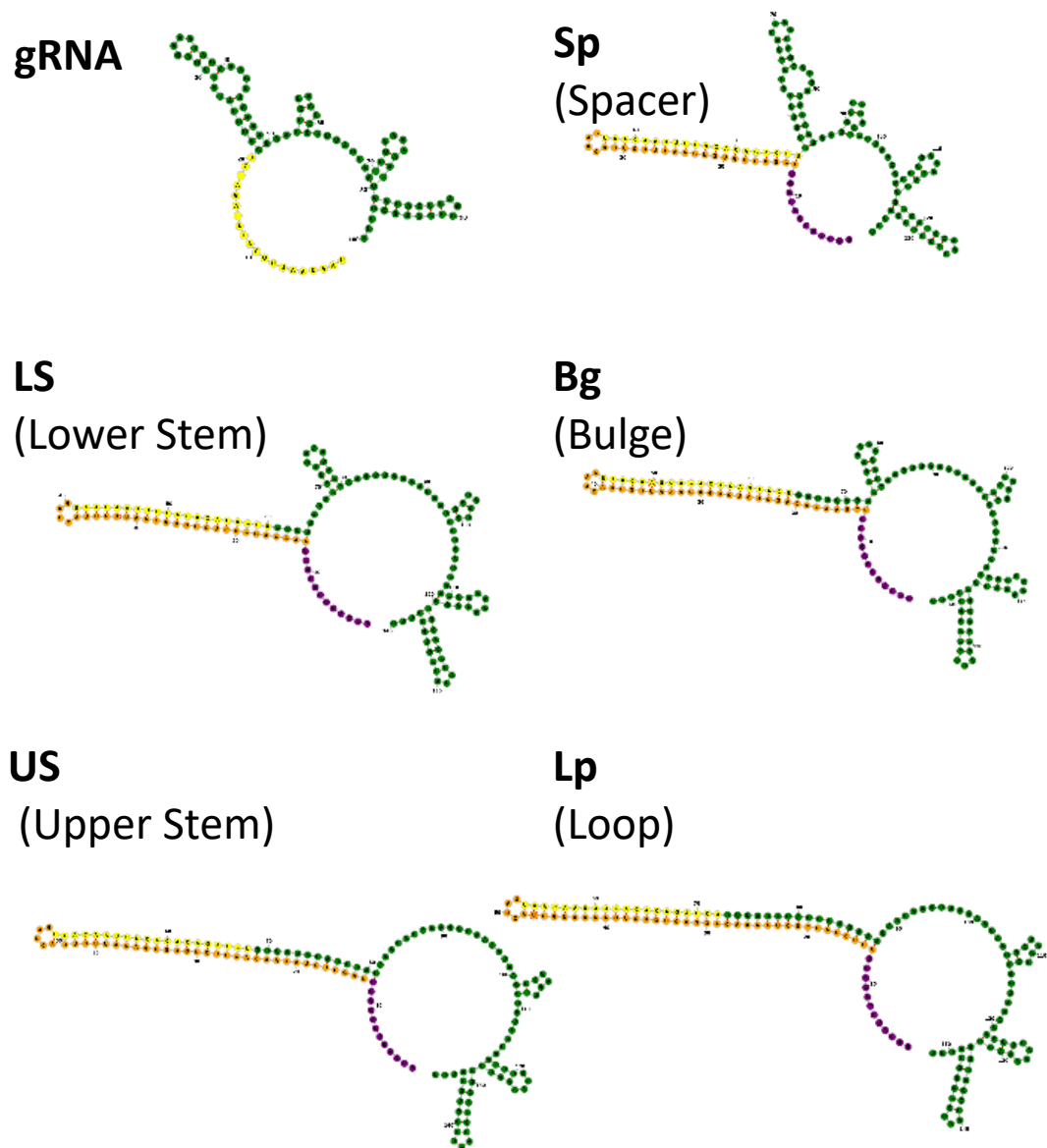


Figure 3.2 Structure schematic of library of crgRNAs. Purple represents the toe hold where interaction initiates with the asRNA, Orange is the region of the *cis*-repressing element which compliments and sequesters functional modules. Yellow is the spacer region, green is the Cas9 binding scaffold. A native state gRNA is included for reference, to demonstrate how modules of the Cas9 Binding scaffold are iteratively complemented by successive members of the library. Structures visualised using FORNA (Kerpedjiev, Hammer, & Hofacker, 2018).

3.2.2 *In vivo* Library testing

To test functionality of these designs, library variants were sub cloned into the pBR322 plasmid with the crgRNA expressed by the P_{LacO-1} promoter and the asRNA expressed by the J23119 promoter (in the same manner used to express the K variant). Similarly, the three plasmid system described in section 2.2.1 was used to test the library. The only change being that dCas9 was expressed from pdCas9-T to reduce the level of expression of dCas9 (2.2.11) (pdCas9-T). The crgRNAs targeted repression of mCherry expressed from the pZS2-123 plasmid, with two other fluorescent proteins are also expressed from the plasmid as controls (Cox et al., 2010). Based on the studies using the K variant of crgRNA, described in section 2.2.11, the optimal fold change was produced with an induction level of 0.2 ng/mL aTc (dCas9 induction, pdCas9-T plasmid) and 100 μ M IPTG (crgRNA induction). Therefore, these levels of induction were used initially for the new library.

In the Sp variant the *cis*-repressing element compliments the spacer region in the same manner as the K variant (the two are functionally equivalent with small differences in restriction sites and cloning). Therefore, as predicted, the levels of repression seen by Sp +asRNA and -asRNA are not significantly different from those of the K variant. The Sp and K variants had basal repression levels of 32.6% and 27.8% repression respectively. Additionally, Sp and K variants in the active +asRNA state resulted in 81.4% and 80.4% repression respectively (Figure 3.3).

Predictions in the library design section of this chapter are based on mutation data, in which changes are made to complementarity within the Cas9 binding scaffold and changes in their functionality are observed. This approach was used, as changes in sequence complementarity might be predicted to yield similar changes in scaffold structure by using a *cis*-repressing element to compliment that part of the structure. Predictions based on this approach do not however adequately predict the comparative level of performance seen by the library. Thus, the mutational data at the least shows that multiple nucleotides within this sequence can be exchanged for non-complementary ones, disrupting complementation, while the complex

maintains activity (Briner et al., 2014). However, extending complementation from the spacer region into the lower stem leads to a significantly increased level of crgRNA *cis*-repression ($P < 0.05$). Similarly, when the mutations removed the bulge by making the sequence entirely complementary, Cas9 was no longer able to cleave its target. When the *cis*-repressing element is extended from the lower stem into the bulge however, the level of *cis*-repression is reduced and the crgRNA activity increases. When the *cis*-repressing element is extended from the bulge into the upper stem no statistically significant change in the level of basal repression was seen. Similarly, Briner et al (2014) found that deleting the other stem led to little change in activity. The final member of the design set includes the extension of the *cis*-repressing element through the loop sequence (Lp). This variant expresses the highest degree of crgRNA inactivation in the absence of the asRNA (only exhibiting 15.3% basal repression of the target promoter).

There are also statistically significant differences in the level of repression of each of the crgRNAs in their active state (+asRNA), The LS variant exhibits significantly less repression (76.2% repression) than the Sp variant (81.4% repression) ($p < 0.05$). The Bg and US variants also show reduced repression, though not statistically significant (20.7% and 21.1% repression respectively). On the other hand, the Lp variant, showed the greatest degree of repression (85.0% repression). This is statistically significantly greater than the LS variant ($p < 0.05$) but not significantly greater than that of the Sp variant. To exclude any contribution of growth rate and other global effects in gene expression, on the repression of target promoter by various variants of crgRNA, the optical density and expression level of a control fluorescent protein, unaffected by the crgRNA variants were measured in all variants. Both the optical density and the level of fluorescence of the control fluorescent protein showed no statistically significant difference across variants (Figure 4B & 4C).

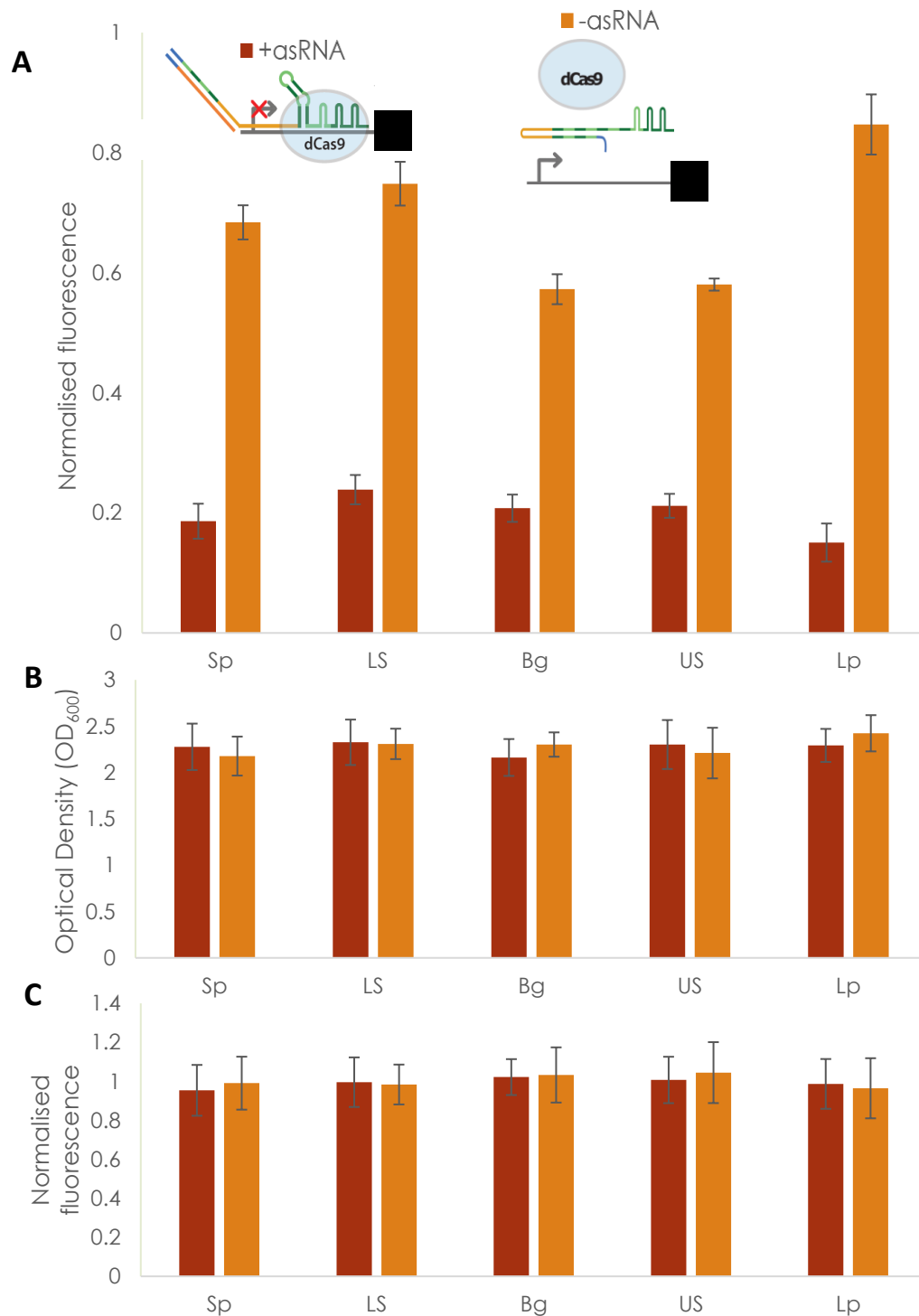


Figure 3.3 Functional test of the Library of crgRNA designs. Expressed in 0.2 ng/mL aTc (dCas9 induction) and 100 μ M IPTG (crgRNA induction). (A) Fluorescence of reporter gene (mCherry) normalised to positive control expressing mCherry with no crgRNA repression. (B) Optical densities of the cultures, no statistically significant differences (one way ANOVA, $p < 0.05$). (C) Level of fluorescence of control fluorescent protein normalised to positive control, no statistically significant differences (one way ANOVA, $p < 0.05$). Error bars represent standard deviation of sample $n=24$. sample. Full data processing methods found in section 5.3.1.

Based on the design variants expressed in this library it is evident that the Lp expression demonstrates the most optimal performance by having both the greatest *cis*-repressed inactivation (-asRNA) and the greatest repression in the active state (+asRNA) (Figure 3.3A). The inducer levels used in these experiments were the same levels found to be optimal for the K variant from section 2.2.11. Therefore, an induction level optimisation was required to get the best performance out of this variant. Whilst there is still approximately 15% of the dynamic range of both the upper and lower end of the spectrum (approximately 15% repression in inactive -asRNA state and 85% repression in active state +asRNA) to reach the maximum fold change, it is more important to maximise repression in the active state with +asRNA than basal repression in the inactive -asRNA state. This also corresponds to the requirements for biological logic gates where low levels of OFF state leakiness can cause greater problems than small changes in the level of ON state expression. Consequently, while optimising inducer levels for the induction of crgRNA and dCas9 higher dCas9 expression levels of both dCas9 and the crgRNA were explored, as both of these have the potential to lead to an increase in repression from the system, based on earlier observations.

As predicted, the increase in the level of expression of dCas9 increases the degree of repression of mCherry (Figure 3.4). The largest fold difference between +asRNA and -asRNA is seen at 100 μ M IPTG and 0.5 ng/mL aTc. This is the same concentration of IPTG as was used in the first test of this library, which was also optimal for the K variant. On the other hand, 0.5 ng/mL aTc represents an increase in the level of expression of dCas9. The increase in level of induction from 0.2 to 0.5 ng/mL aTc led to an increase in the fold change because the change in repression for the +asRNA state was greater in percentage terms than the change for the -asRNA.

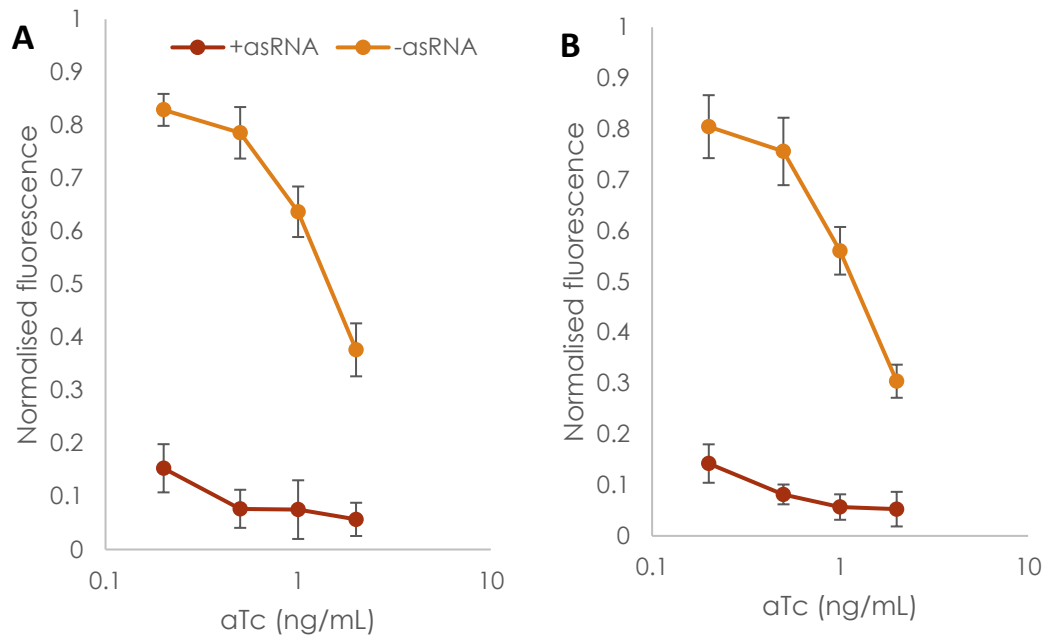


Figure 3.4 Induction curve of dCas9 with the Lp variant in both the active +asRNA and inactive -asRNA states. (A) 100 uM IPTG, (B) 250 uM IPTG (crgRNA induction). Concentrations of aTc; 0.2, 0.5, 1, 2 ng/mL aTc. Error bars represent standard deviation of sample (3 repeats with 8 replicates; n=24). Results are normalised to a positive control expressing mCherry without crgRNA to repress it. Full data processing methods found in section 5.3.1.

When the new optimal induction levels for the Lp variant (100 μ M IPTG and 0.5 ng/mL aTc) are applied to the entire library, the performance pattern of the library remains consistent but with an increase in the fold change of each library variant (Figure 3.5).

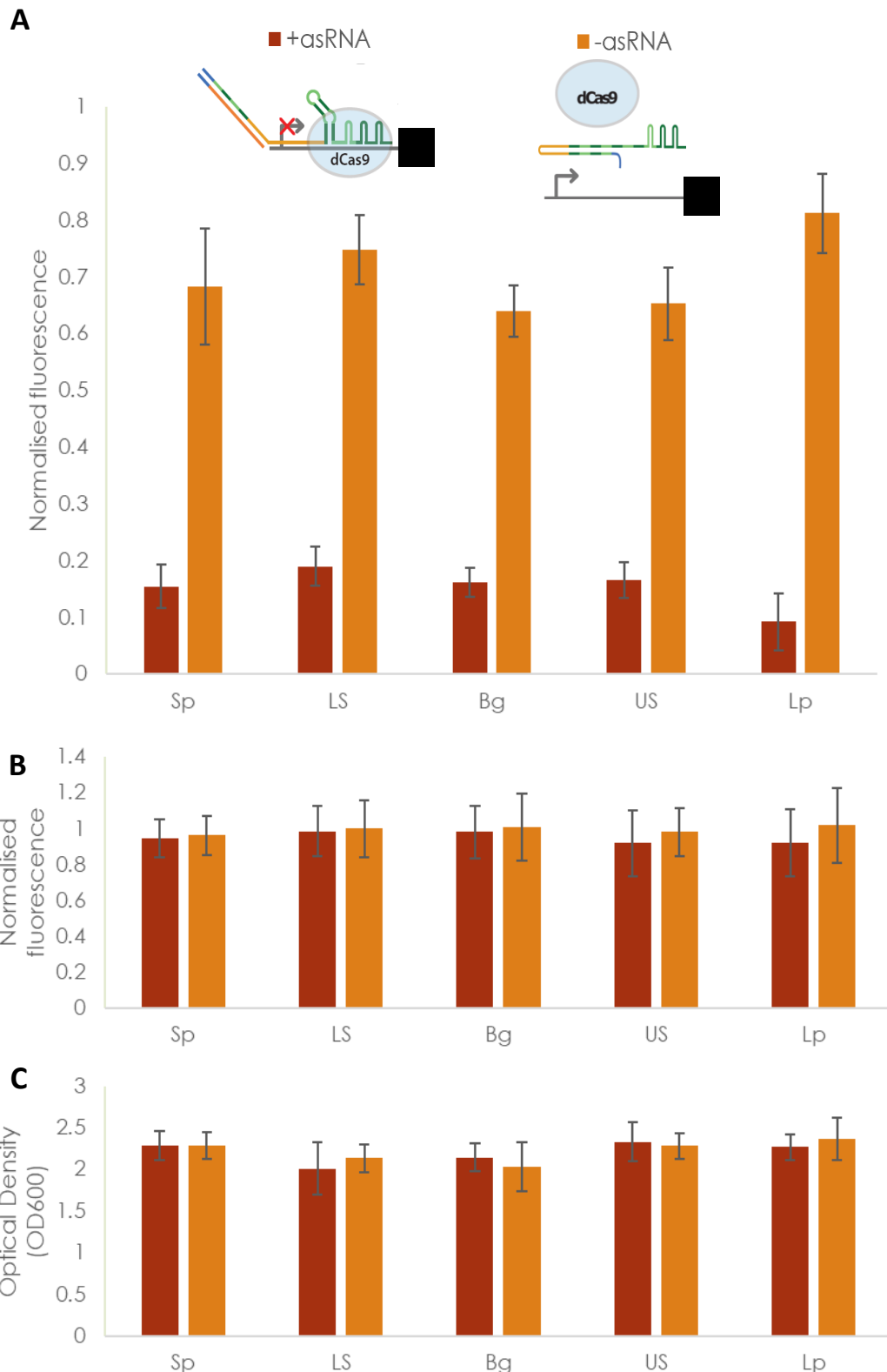


Figure 3.5 Library of crgRNA designs expressed in 0.5 ng/mL aTc (dCas9 induction) and 100 μ M IPTG (crgRNA induction). (A) Normalised fluorescence of target gene (mCherry). (B) Normalised level of fluorescence of control fluorescent protein, no statistically significant differences (one way ANOVA, $P < 0.05$). (C) Optical densities of the cultures, no statistically significant differences (one way ANOVA, $P < 0.05$). Error bars represent standard deviation of sample, $n = 24$. Full data processing methods found in section 5.3.1.

3.2.3 Toehold orthogonality

A potential advantage of the crgRNA system is the ability to have many crgRNAs gates functioning within the same cell. One way to achieve this is to use different crgRNA asRNA pairs which target different promoters or genes. Orthogonality between crgRNA asRNA pairs is critical for preventing off target effects and cross-talk between crgRNAs and non-cognate asRNAs. When the crgRNA is transcribed and the *cis*-repressing element folds, complementing the Cas9 binding scaffold, the toehold is left accessible for the asRNA to initiate complementation (Figure 3.1).

The entire *cis*-repressing element of the crgRNA is complemented by the asRNA. There are three main modules of the *cis*-repressing element (toehold, spacer complementing and scaffold complementing) each showing different degrees of sequence variability. The asRNA complements the entire *cis*-repressing element and so the same sequence constraints that apply to the *cis*-repressing element, also apply to each respective module of the asRNA. The first module of the crgRNA is the toehold which is unbound in the crgRNA's inactive conformation and initiates the interaction with the asRNA before the junction migrates up the hairpin. This module is entirely variable and can be optimised to avoid complementation with the rest of the crgRNA. Unintended toehold complementation within the structure could potentially hinder structure formation or lead to reduced probability of initiating an interaction with the asRNA. The toehold can also be optimised to minimise interactions with non-cognate asRNAs to ensure orthogonality.

As it was the Lp variant which showed the greatest fold change; it was then decided to make multiple functional homologues of the Lp variant that are orthogonal to each other. However, there are challenges in using the Lp variant to create orthogonal homologues, as it has a large region of the Cas9 binding scaffold complemented by the *cis*-repressing element imposing sequence constraints. Since the asRNA in turn, complements the *cis*-repressing element, the asRNA also bears these sequence constraints. This means when different versions of the Lp variant are made with different DNA targets the sequence constraints imposed result in a

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

segment of all asRNAs being identical and therefore matching a segment of all crgRNAs. Consequently, there is a high probability of non-cognate interactions and cross talk between different versions.

An element of the system that would mitigate this effect is the hairpin sequence (in the inactive state) that forms between the *cis*-repressing element and the rest of the crgRNA. The hairpin includes the sequence which is identical between versions reducing base accessibility and therefore reducing the chance of crosstalk. The toehold sequence is where interaction initiation occurs between the asRNA and the crgRNA. Therefore, maximising orthogonality at the toehold becomes the focus for maximising the orthogonality of the system.

A boost for orthogonality is the fact that the spacer module of the crgRNA would vary between versions depending on the target DNA sequence. Therefore, the corresponding regions of the *cis*-repressing element and the asRNA will also differ between variants. This has a positive impact for orthogonality, and it is possible to take this into consideration when optimising DNA targets for orthogonality. To change the spacer sequence while keeping the target gene the same to minimise variation within the experiment three different PAM sites were selected from within the coding region of the gene for use with the orthogonality experiment. After this when applying the system to multiple different promoters in each case the PAM closest to the transcriptional start site was used.

To generate multiple versions of the Lp system it was first necessary to ascertain whether sequences with different toeholds will retain the same functionality. To this end, three different versions of the Lp system were made, each with a different toe hold sequence (Lp- α , Lp- β , Lp- γ), sub-cloned into pBR322 and expressed in the same manner (0.5 ng/mL aTc, 100 μ M IPTG). There was no statistically significant difference between the three versions, indicating that indeed the system can function with different toehold sequences (Figure 3.6).

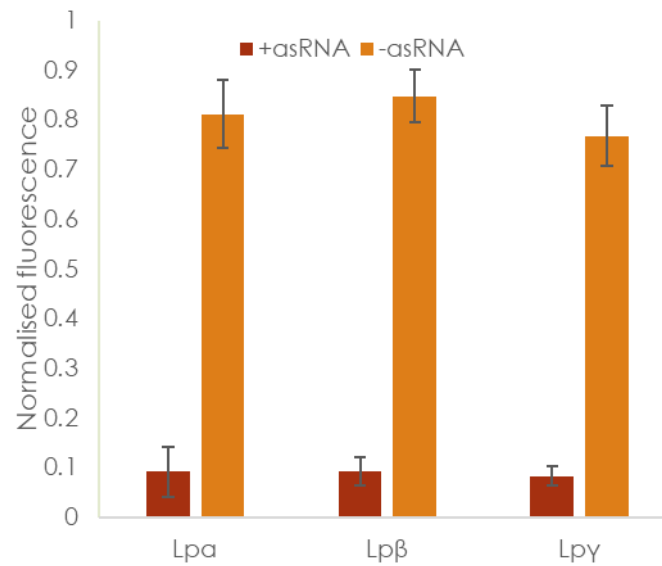


Figure 3.6 Variants on the Lp design with different toeholds. No statistically significant differences in repressive capability between the variants despite change of toehold sequence ($p=0.05$). Inducer concentrations were 0.5 ng/mL aTc, 100uM IPTG. Error bars are standard deviation, $n=24$. Full data processing methods found in section 5.3.1.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

To aid in the design process of crgRNA-asRNA pairs and maximise orthogonality, an *in silico* method was developed which generates toehold sequences that maximises the orthogonality of the different Lp versions. The method functions by optimisation in three stages (Figure 3.7). In the first stage a range of potential toeholds with maximum accessibility and minimal complementation with the rest of the crgRNA was generated. In the second stage, pairwise orthogonality tests were conducted between the asRNAs and crgRNAs with each of the toeholds from the earlier stage. This generated an interaction matrix, used in the third and final stage where the orthogonality of different combinations of toeholds, one for each cognate pair, were assessed to find the best combination of toeholds. The underlying script described here is included in Appendix 1: Toehold orthogonality script.

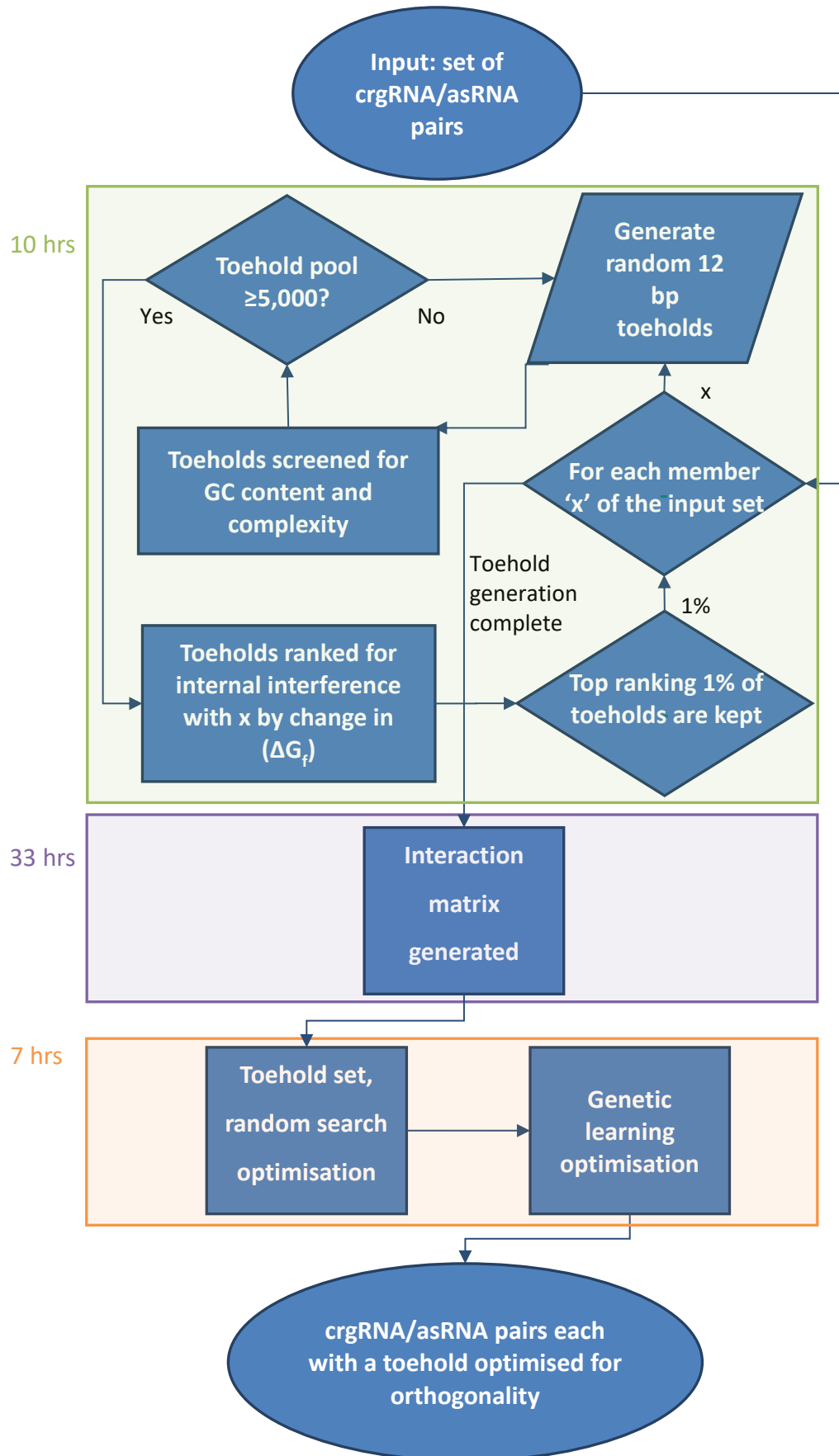


Figure 3.7 overview of the main processes by which toeholds are optimised for orthogonality. Computing time taken for each stage of the process is indicated next to that stage in hours (I7-3770 processor was used).

3.2.4 Stage 1: Toehold generation

A pool of randomly generated 12 nucleotide toehold sequences was produced. This pool was first screened for GC content, keeping only toeholds with a GC content of more than 45%. It is important for the toehold to have a high GC content to optimise the interaction initiation between the asRNA and the crgRNA. A higher GC content leads to a lower ΔG_f of the initial interaction between the toeholds of the crgRNA/asRNA pair, reducing transient interactions and consequently increasing the flux towards the pair forming a heteroduplex (Green et al., 2014a). A further screen to remove any sequence of low complexity like tandem repeats of 4 or more nucleotides such as 'AAAA', this left 56% of the initial toeholds pool. Random toeholds continued to be generated and screened until a pool of 5,000 toeholds which had passed the screen was achieved. This resulted in initially approximately 8,900 random toeholds being generated to reach the screened 5,000 toehold pool.

3.2.5 Stage 2: Toehold ranking and screening

In the next stage toehold sequences were scored for toehold base accessibility and unintended interactions with the rest of the transcript. The best toeholds will have the greatest proportion of bases within them available to complement with the antisense RNA within the thermodynamic ensemble. Undesirable secondary structure and complementation within the toehold or between the toehold and the rest of the crgRNA will reduce the base availability for the asRNA and so these sequences were removed. This was done using folding predictions from the Vienna 2.0 RNA folding package (Lorenz et al., 2011). The free energy of the thermodynamic ensemble of the crgRNA was calculated with the toehold ($\Delta_f G_{cT}$) and without ($\Delta_f G_c$). The same values were calculated for the cognate asRNA; with toehold ($\Delta_f G_{aT}$) and without ($\Delta_f G_a$). If the free energy dropped upon the addition of the toehold, this indicates complementation either within the toehold or between the toehold and the rest of the transcript. These free energy values are combined in Equation 3.1 to form a toehold availability score for each toehold.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Equation 3.1 *Toehold Availability Score* = $(-\Delta_f G_c + \Delta_f G_{cT}) + (-\Delta_f G_a + \Delta_f G_{aT})$

Out of the pool of 5,000 screened toehold sequences, the highest scoring 1% of toeholds for each cognate pair was selected to go on to the next stage. Therefore, each crgRNA/asRNA pair had a range of 50 candidate toeholds to be selected from.

3.2.6 Stage 3: Pairwise Interaction Matrix

To find a set of sequences which are mutually orthogonal it was necessary to score the pairwise interactions, both cognate and non-cognate, between the collective crgRNAs and asRNAs. As there are 50 candidate toeholds for each crgRNA/asRNA pair and 5 crgRNA/asRNA pairs a total of 250 toeholds will be included in the orthogonality optimisation. Scoring the pairwise interactions between all 250 asRNAs and all 250 crgRNAs yields 62,500 interactions. These interactions were scored before selecting combinations of toeholds with the maximum overall orthogonality.

The interaction score was based on interactions that might affect orthogonality. For example, orthogonality will be compromised if a non-cognate asRNA interacts with a crgRNA, displacing the *cis*-repressing element and allowing repression of the output gene. As the occurrence frequency of this non-cognate interaction was crucial, the score was based on pairing probability rather than free energy. Specifically, pairing probability between the asRNA and the *cis*-repressing element is used.

This scoring combines the pairing probabilities of each of the nucleotides in the *cis*-repressing element with the asRNA in an inverse geometric mean (Equation 3.2) where a_1, a_2, \dots are the partial pairing probabilities that each of the nucleotides of the *cis*-repressing element will complement with the asRNA. The score ranges from 0, indicating a negligible degree of interaction to 1, indicating a substantial level of interaction. The inverse geometric mean is used, as this leads to an even smaller

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

number of nucleotides having a high partial pairing probability with the asRNA resulting in large changes in the score.

$$\text{Equation 3.2} \quad \textit{interaction score} = 1 - \sqrt[n]{(1 - a_1)(1 - a_2) \dots (1 - a_n)}$$

The pairwise interactions formed a matrix, with each of the crgRNAs along the x-axis and each of the asRNAs along the y-axis. The matrix was grouped into crgRNA variants with the same DNA target and spacer region but different toeholds (Figure 3.8). The value in each cell of the matrix is an interaction score calculated by the method discussed here.

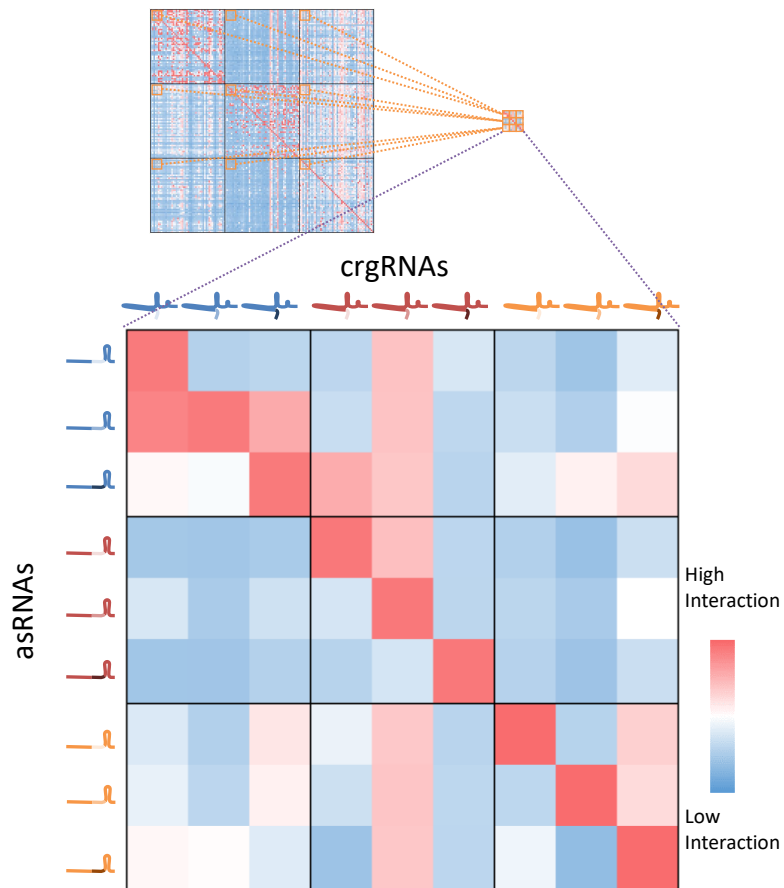


Figure 3.8 Full interaction matrix above, simplified interaction matrix below. crgRNAs are presented across the x-axis, asRNAs along the y-axis. Gridlines have within them crgRNA/asRNA pairs targeting a single genetic locus but with varying toehold sequences. Pairs of crgRNA/asRNAs are coloured to represent which genetic locus they target, toeholds are represented in different shades.

Several approaches to scoring the pairwise interaction within the interaction matrix were explored. An initial look at just the partial pairing probability between the spacer binding module of the *cis*-repressing element and the associated region of the asRNA created a score which was only dependent on stable interactions. These are stable interactions, as complementation of the *cis*-repressing element would have to be almost entirely unfolded for the spacer binding module to be complementing the asRNA (Figure 3.9A). While this measure was good for describing stable interactions with a very high probability of affecting orthogonality, it would hide low level interactions which may also affect orthogonality. This can be seen in the strongly bimodal matrix histogram. Using a similar approach, but instead targeting the scaffold binding module of the *cis*-repressing element (Figure 3.9B) reveals greater details of the interactions. As described in Section 3.2.3, due to the *cis*-repressing element complementing part of the Cas9 binding scaffold, there is a sequence common to all crgRNAs which is complementary to all asRNAs, cognate and non-cognate. This universality makes this section likely to be involved in any non-cognate interactions and using this region reveals where interactions between toeholds propagate along the *cis*-repressing element hairpin. As can be seen from the matrix (Figure 3.9B) this measure exposes more non-cognate interactions. Looking only at toehold interactions (Figure 3.9C) was too sensitive a measure as it captures many of the interactions between exposed toeholds, which would not lead to propagation of the interaction into the *cis*-repressing element with less detrimental effect on orthogonality. Taking into account these various issues, the measure that was finally chosen (Figure 3.9D) was a composite of these measures using the entire partial pairing probability between the *cis*-repressing element and the asRNA. This measure is dominated by interactions in the scaffold binding and spacer binding regions as they represent more of the sequence but also includes the interaction with the toehold. This interaction measure was used to generate the matrix used for optimisation.

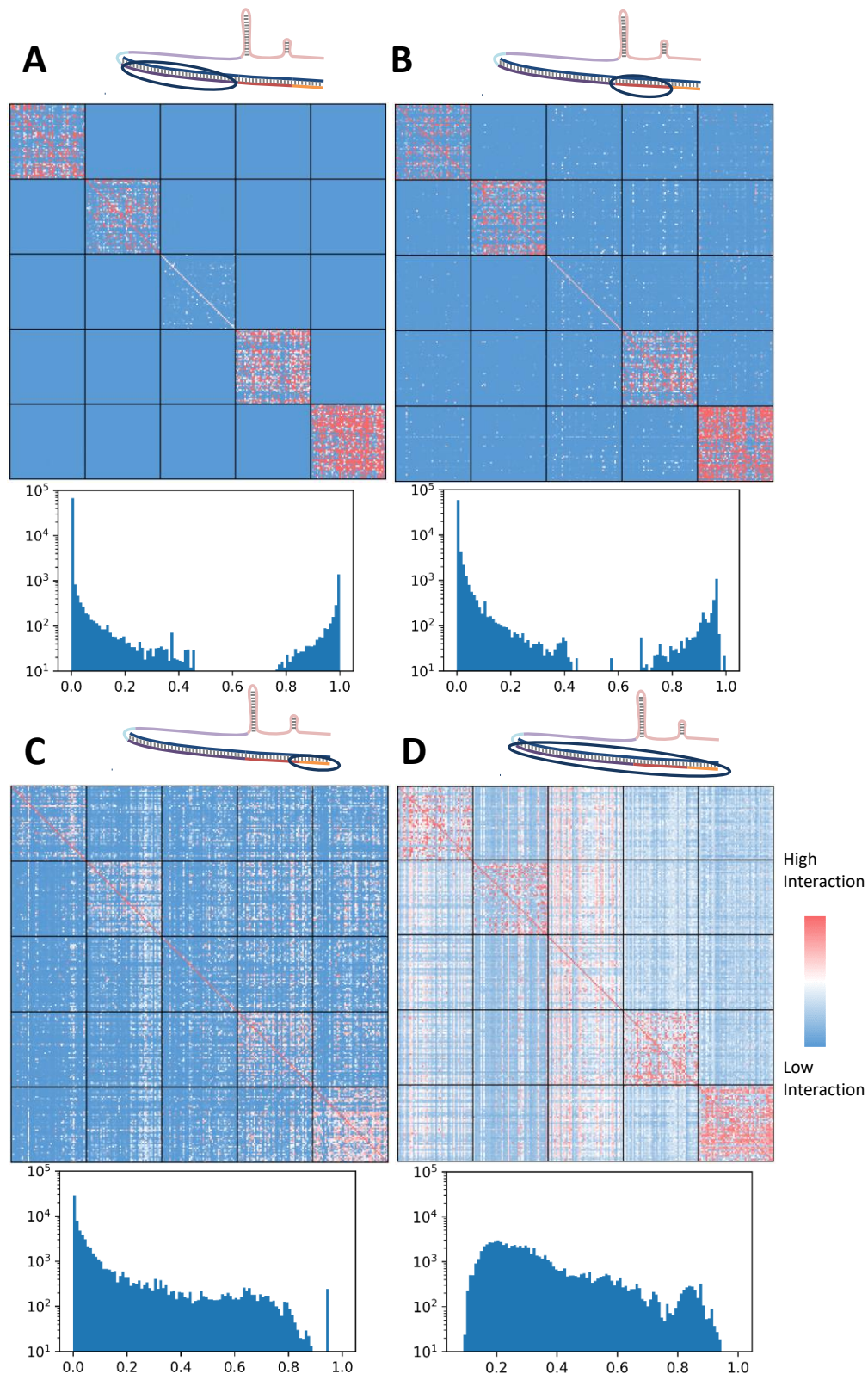


Figure 3.9 Pairwise interaction matrices for a library of asRNAs and crgRNAs calculated using folding predictions based either on (A) the spacer binding module of the *cis*-repressing element, (B) the scaffold binding module of the crgRNA's *cis*-repressing element, (C) the toehold or (D) all interactions between the *cis*-repressing element and the asRNA. Below each interaction matrix is a descriptive histogram with interaction score along the x-axis and score and frequency on the y-axis.

3.2.7 Toehold set optimisation

To obtain a full set of crgRNA/ asRNA pairs, one toehold has to be chosen for each of the five pairs (Figure 3.8). To calculate a combined orthogonality score for a set of crgRNA/asRNAs, the pairwise non-cognate interactions for a particular set of toeholds can be retrieved from the interaction matrix and summated. Trying all possible combinations of 5 toeholds, one from each set of 50 toeholds, for each of the cognate pairs, yields $3.1 \cdot 10^8$ different possible combinations. Calculating the orthogonality index for all of these combinations would not be feasible. Instead a two-stage approach was used: a random search (Figure 3.10), followed by a genetic machine learning optimisation.

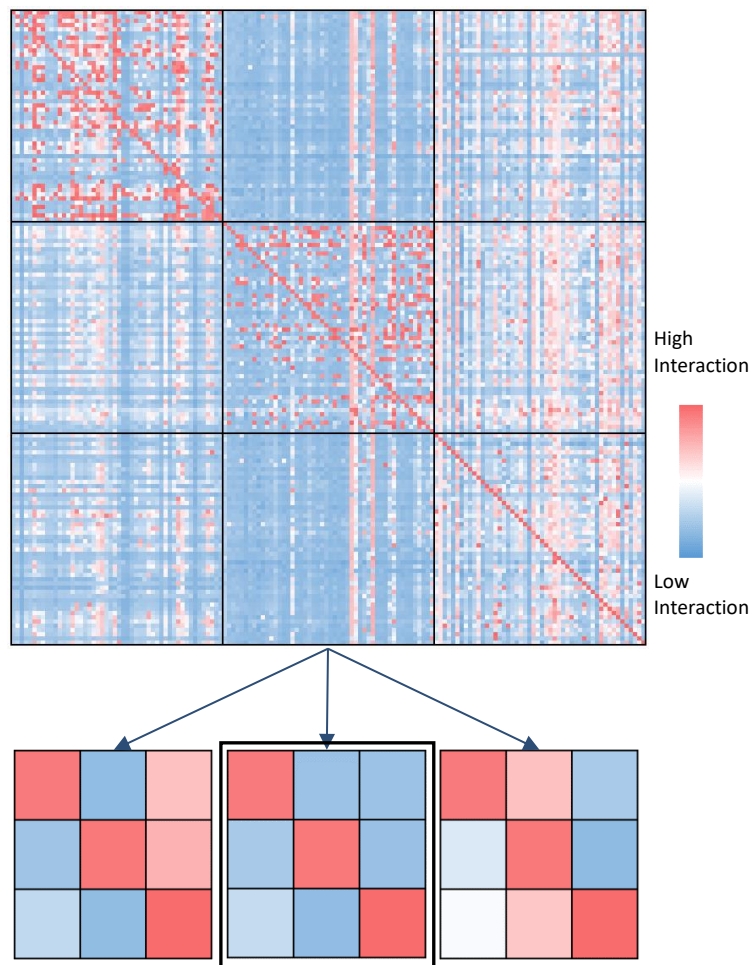


Figure 3.10 Simplified representation of random search optimisation. The interaction matrix above has three different crgRNA/asRNA pairs (within gridlines) each with 50 toeholds (between gridlines). The algorithm selects a random toehold from the matrix for each crgRNA/asRNA pair from the matrix. Creating a pool of different toehold selections below. The toehold selection with the lowest orthogonality index (the highlighted central one) represents the better toehold set for orthogonality.

In the random search stage (Figure 3.10), a million random combinations are assessed (Figure 3.11) and the top scoring 100 candidate combinations are passed into the second stage. These candidates go through 1 million iterations; each iteration includes selection of a random toehold set from the pool of 100 and 'mutated' to change some of the toeholds in the set, for alternatives from the interaction matrix. The new, altered set is given a combined orthogonality score. If the new score is better than the lowest scoring set in the pool, that set was replaced with the new set. In the mutation function, the number of toehold changes that are made to yield a set of toeholds and the selection of toeholds to be changed are random. The probability distribution for the number of toehold substitutions in a set is Equation 3.3 where n is the number of mutations and p is the probability. This probability distribution means that while most 'mutations' only change one or two toeholds in the set, a small number of mutations will change a higher number of the toeholds allowing the system to escape from local minima.

$$\text{Equation 3.3} \quad p = 0.5^n$$

The search resulted in crgRNA/asRNA pairs with toeholds optimised for orthogonality. The 5 selected optimal crgRNA/asRNA pairs were synthesized and cloned for functional testing *in vivo*. The script developed here is available (Appendix 1: Toehold orthogonality script) and can be used to design new sets of crgRNA asRNA pairs for use with different DNA targets in the same cell, combined by the user into the desired genetic circuitry.

3.2.8 Orthogonality index over time

As the program iterates through the toehold optimisation, the orthogonality index of the toehold set decreases (a lower index indicates higher orthogonality). A randomly selected toehold set from the matrix gives an average orthogonality index of 39 (arbitrary units). As the random search and learning optimisation progresses, the orthogonality index drops to 24 with a standard deviation of 0.4 (Figure 3.11).

The low variation seen at the end of the process suggests that the algorithm has, in each case, yielded a toehold set with an orthogonality index very close to the optimal minimum. It is also notable that, of the 140 runs of this optimisation process, no two runs yielded the same set of toeholds, implying this is an optimisation problem with a large number of potential answers with comparable levels of orthogonality. As the algorithm progresses, the reduction in orthogonality index appears to follow an asymptotic curve approaching a theoretical minimum. This reflects the underlying biology in which orthogonality is dependent on a number of factors and while optimising the toehold sequence appears to lead to substantial increase in orthogonality, there is only so far it can go. The curve formed by plotting orthogonality index appears to be asymptotically approaching a theoretical limit for the orthogonality with increasing the number of iterations (Figure 3.11). This is to be expected from a biological perspective as it is only the toehold which is being optimised; there are still regions of each asRNA which complement every crgRNA and so there is a maximum degree of orthogonality which can be achieved with this approach.

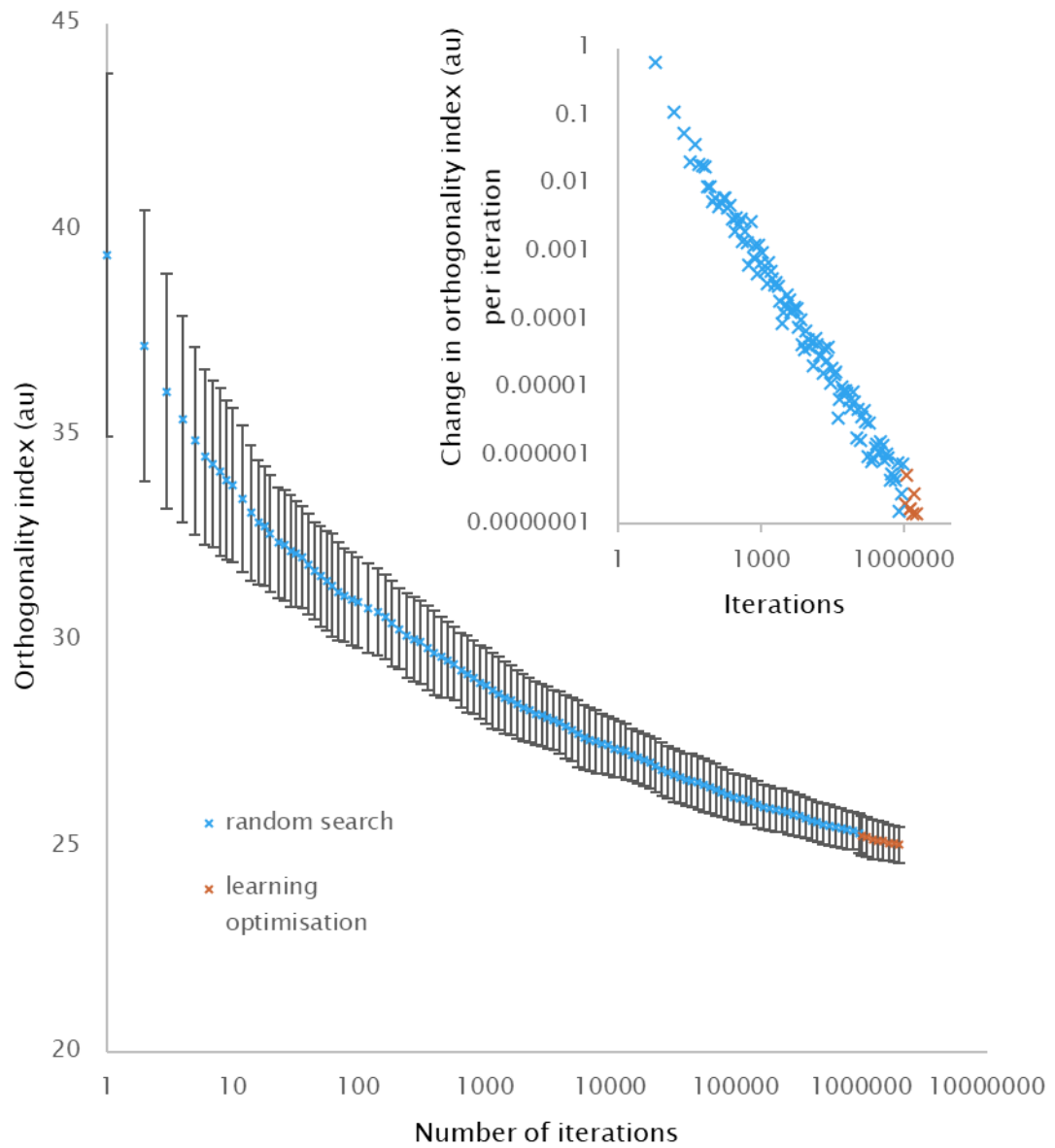


Figure 3.11 Orthogonality index (AU) over iterations of optimisation. The first million iterations are random search (in blue), the second million iterations are evolving toehold sets in orange. Error bars are standard deviation. n=140 different runs of the optimisation.

3.2.9 Result verification

Next validation of the results obtained using the Vienna 2.0 folding package (Lorenz et al., 2011) was sought to verification of the change in orthogonality using another package. The software package (NUPACK) was therefore used to measure the interaction score for a random set of crgRNA asRNA toehold sequences. The interaction score obtained with NUPACK software (Zadeh et al., 2011) was compared to that obtained from the Vienna 2.0 folding package (Figure 3.12). The interaction score derived from the NUPACK software represents the incidence of the heteroduplex within the structural ensemble, taken as a logarithm of base 10. This is analogous to the pairing probability used to generate the Vienna 2.0 based interaction score. As seen from Figure 3.12A, there is a positive correlation between the Vienna 2.0 based score and the NUPACK based score with cognate RNAs exhibiting strong interactions on both scales and a positive correlation amongst non-cognate interactions between the predicted interaction score of one system and the other providing a validation.

The same approach was repeated with a set of optimised toeholds (Figure 3.12B). Cognate RNA interactions still received high scores on both scales; however, within non-cognate interactions, lower levels of interaction are seen on both scales, denoting an increase in orthogonality within the set. As predicted, the reduction in non-cognate interactions was greater in terms of the scale that was used to optimise the toeholds (Vienna 2.0 suite), when compared to the NUPACK derived scale. It is notable that the positive correlation between scales seen in the random toehold set is mostly absent amongst non-cognate interactions within the optimised toehold set. This may be due to the fact that two million rounds of optimisation can lead to the selection of outliers, which are not necessarily the most accurate folding predictions. Therefore, the optimised toeholds no longer have a clear positive correlation between the indices derived from the two alternative folding prediction packages. This suggests that, even though it was possible to run more iterations of optimisation, any further decrease in the

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

orthogonality index would not necessarily be biologically relevant or correspond to a further increase in orthogonality.

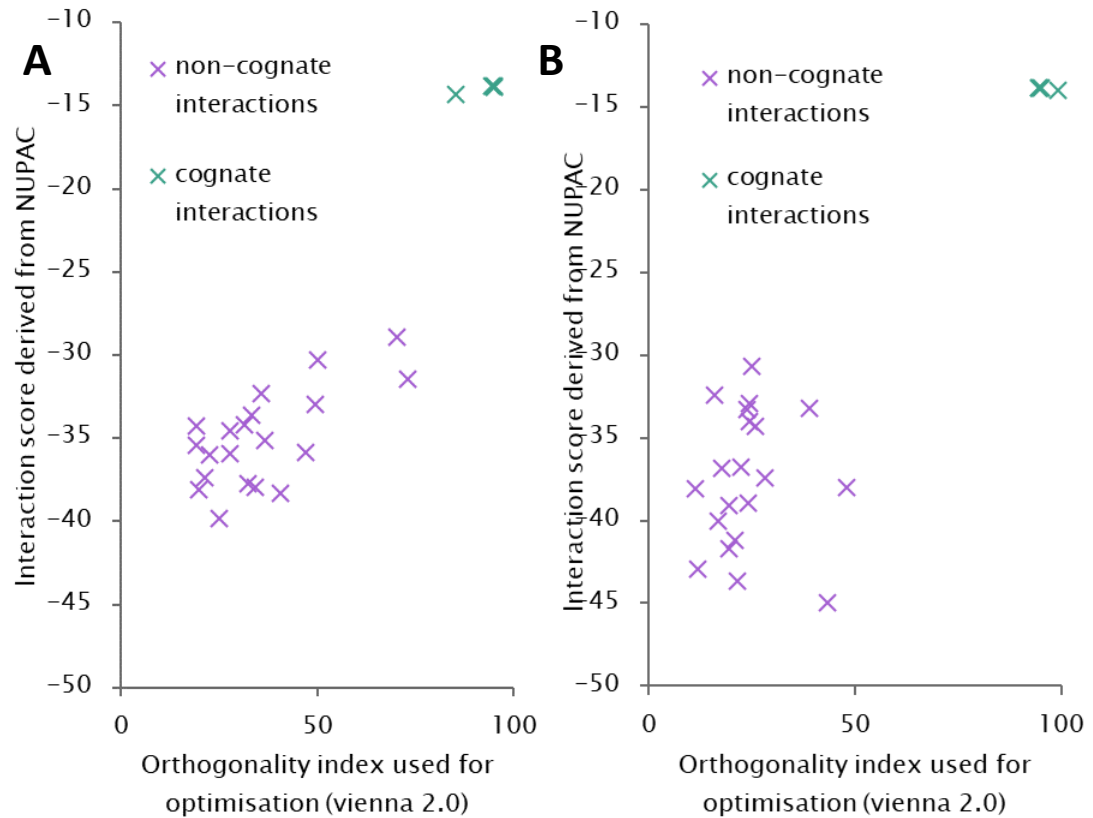


Figure 3.12 The orthogonality index used for optimisation of a toehold set (x axis) is plotted against orthogonality index calculated with an alternative folding package (NUPACK) plotted on the Y axis. Each graph shows both cognate and non-cognate interactions as different data series. **(A)** The orthogonality scores resulting from a random set of toeholds showing a positive correlation between the two measures for orthogonality. **(B)** The orthogonality scores resulting from a set of toeholds optimised using the Vienna 2.0 suite. The positive correlation between the two indices within non-cognate interactions has broken down in B. non-cognate interactions, n=20; cognate interactions n = 6.

3.2.10 Testing universalizability of Lp crgRNA

Having produced an *in silico* method for generating Lp versions which are functional homologues while exhibiting orthogonality from one another, the next step was to generate a set of homologues with different targets and test them. Of the five crgRNA asRNA pairs optimised for orthogonality, three were expressed with cognate and non-cognate interactions as a test for orthogonality. These repressed mCherry through targeting different sites in the coding region. The decision to target the same reporter gene in the orthogonality test was intended to avoid bias resulting from using multiple reporters. The remaining two optimised versions were used to target the Transcriptional Start Site (TSS) of two other fluorescent reporter genes (YFP and CFP) to be contrasted with the original Lp variant targeting the TSS of mCherry.

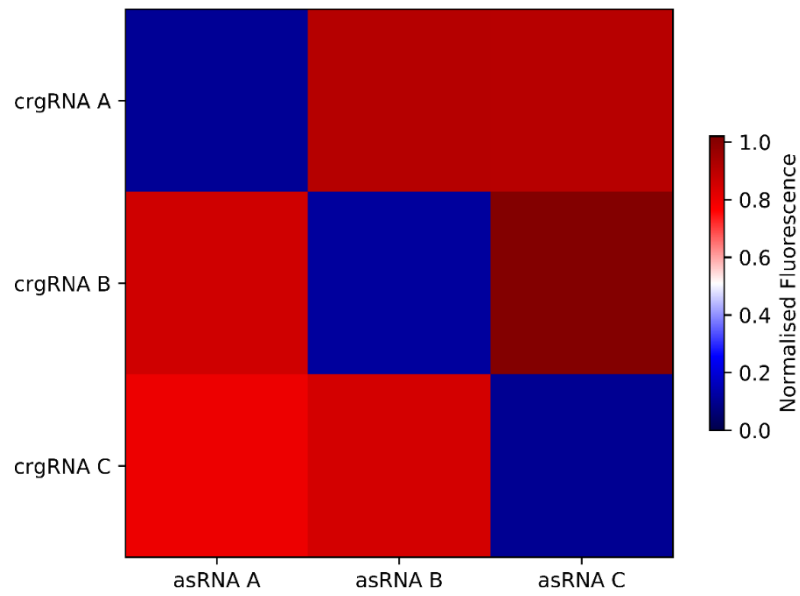
All five new crgRNA/asRNA pairs were used as the input for the program to produce toeholds to maximise orthogonality between all of them. It was only necessary to maximise orthogonality between the three, that were used to test orthogonality but applying the program to all 5 gave a more realistic understanding of the level of orthogonality that can be expected from 5 crgRNA, asRNA pairs operating in the same cell.

3.2.11 *In vivo* Orthogonality

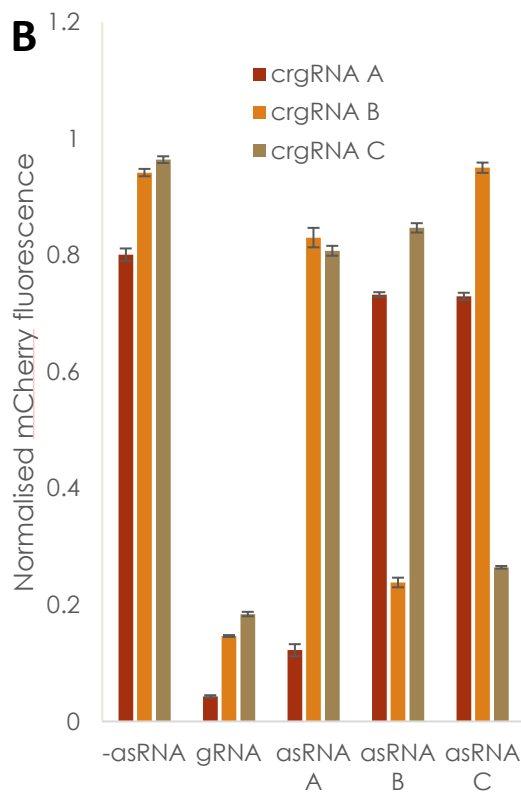
Here is described in greater depth how new functional homologs of Lp were generated and optimised for orthogonality testing. New DNA targets for dCas9 binding were selected and then the associated gRNA sequence is generated. The *cis*-repressing element derived from the resulting gRNA sequence complements the spacer region as well as the Cas9 binding scaffold up to and including the first loop sequence. After the selection of target sequences for various Lp versions, all Lp versions that are intended for use in the same cell are optimised for orthogonality using software described in Sections 3.2.3-3.2.8 and included in Appendix 1: Toehold orthogonality script.

Figure 3.13B shows results normalised to a positive control, Figure 3.13C shows results renormalized to reveal level of orthogonality and level of crgRNA functional rescue by cognate asRNA. The level of repression by each of the three crgRNAs used in the orthogonality test show variation in the level of repression when each of the crgRNAs is expressed with its cognate asRNA (88% to 74% repression) (Figure 3.13B). Control gRNAs with the same target but without the *cis*-repressing element for each crgRNA show variation in repression which correlates with the variation in crgRNA +asRNA repression (Figure 3.13B). To control for this variation resulting from underlying gRNA properties, the level of fluorescence was re-normalised so the gRNA repressed strains are 0 for their respective crgRNAs, to demonstrate the level of orthogonality, variation in the level of repression of each of the crgRNAs (Figure 3.13B, -asRNA) is masked by re-normalising, the crgRNA -asRNA level of fluorescence to 1 for each crgRNA (Figure 3.13C). After controlling for variation in the underlying gRNA no statistically significant differences were found between the repression levels of the crgRNAs expressed with cognate asRNA. When variation in crgRNA –asRNA is masked a degree of non-cognate cross reactivity is seen varies depending on the asRNA and on the crgRNA (0% to 20% with an average of 11% repression) this contrasts with the average 89% repression resulting from interaction with a cognate asRNA.

A



B



C

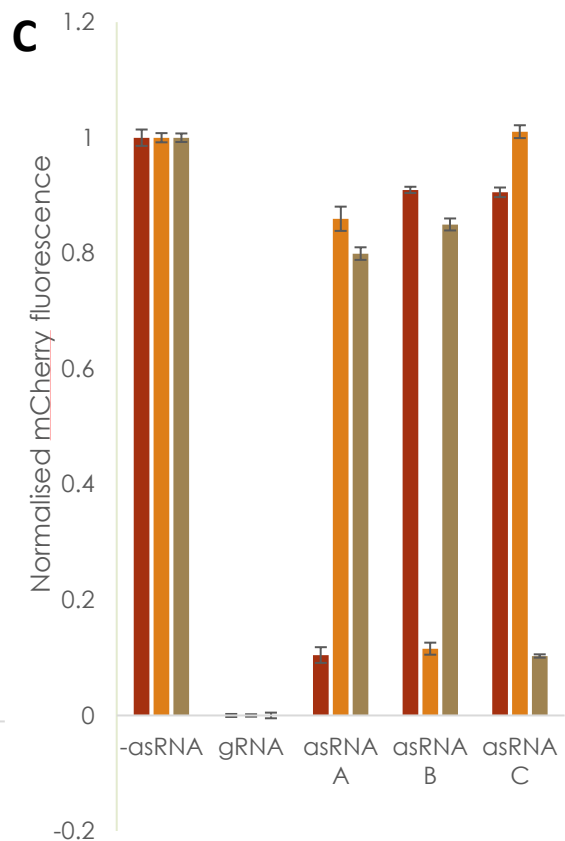


Figure 3.13 Please see figure legend on next page.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Figure 3.13 (B) data normalised to system expressing fluorescent proteins and dCas9 but no crgRNA (value of 1) and cultures expressing dCas9 but no fluorescent proteins (value of 0). (A,C) Data normalised to system expressing dCas9 and fluorescent proteins and either +crgRNA -asRNA for a value of 1 or the corresponding gRNA for a value of 0. (A) Assessing orthogonality of crgRNA system variants. The crgRNA-asRNA pairs have unique spacer sequences targeting three different sites in the mCherry gene. Each crgRNA is expressed with each asRNA. The gRNA only version of each crgRNA are included as controls. There is an average non-cognate repression of 11% (1.1 fold) contrasting with an average 89% (9.3 fold) repression for cognate interactions. Error bars are SEM, n = 24. Full data processing methods found in section 5.3.1.

3.2.12 Multiple crgRNA targets

The next step was to apply the system to three different genes; two additional Lp versions were generated targeting two additional fluorescent reporter genes (YFP, and CFP) to be combined with the original Lp targeting the mCherry gene. The sequence of the Lp crgRNA was changed to be target specific to each of these promoters. First the nearest PAM to the TSS of the target promoter was selected; the 20 bases of the spacer region in the crgRNA were modified to complement the sequence adjacent to the PAM to allow the crgRNA, asRNA, dCas9 complex to bind to the new target promoter. The *cis*-repressing element in the Lp variant complements the spacer region as well as part of the dCas9 binding scaffold, therefore the part of the *cis*-repressing element which complements the spacer also was changed to maintain complementarity with the spacer region. The dCas9 binding scaffold and corresponding segment of the *cis*-repressing element were kept constant. A sequence which did change was the toehold; this is where complex formation between the asRNA and the crgRNA initiates and therefore has consequences for orthogonality between non cognate asRNA, crgRNA pairs. The algorithm described in Sections 3.2.3-3.2.9 optimised each toehold sequence to prevent non-cognate interactions leading to an *in silico* increase in orthogonality. This orthogonality was demonstrated *in vivo* (Figure 3.13B).

Due to the promoter used, the Cyan Fluorescent Protein (CFP) is expressed at a low level; therefore, even though the normalisation allows it to be compared directly with the other two fluorescent proteins, the low level of expression means that intrinsic errors in measurement are larger (Figure 3.14). The three gRNAs, one targeted to each promoter TSS did not show statistically significant variation in level of repression (though crgRNA repressed CFP, has high error bars). Similarly, the level of repression in the three crgRNA +asRNA for each target fluorescent protein were not significantly different from one another (86% average repression). All three crgRNAs showed a small amount of repression for their target gene in the -asRNA state (19% average repression) (Figure 3.14).

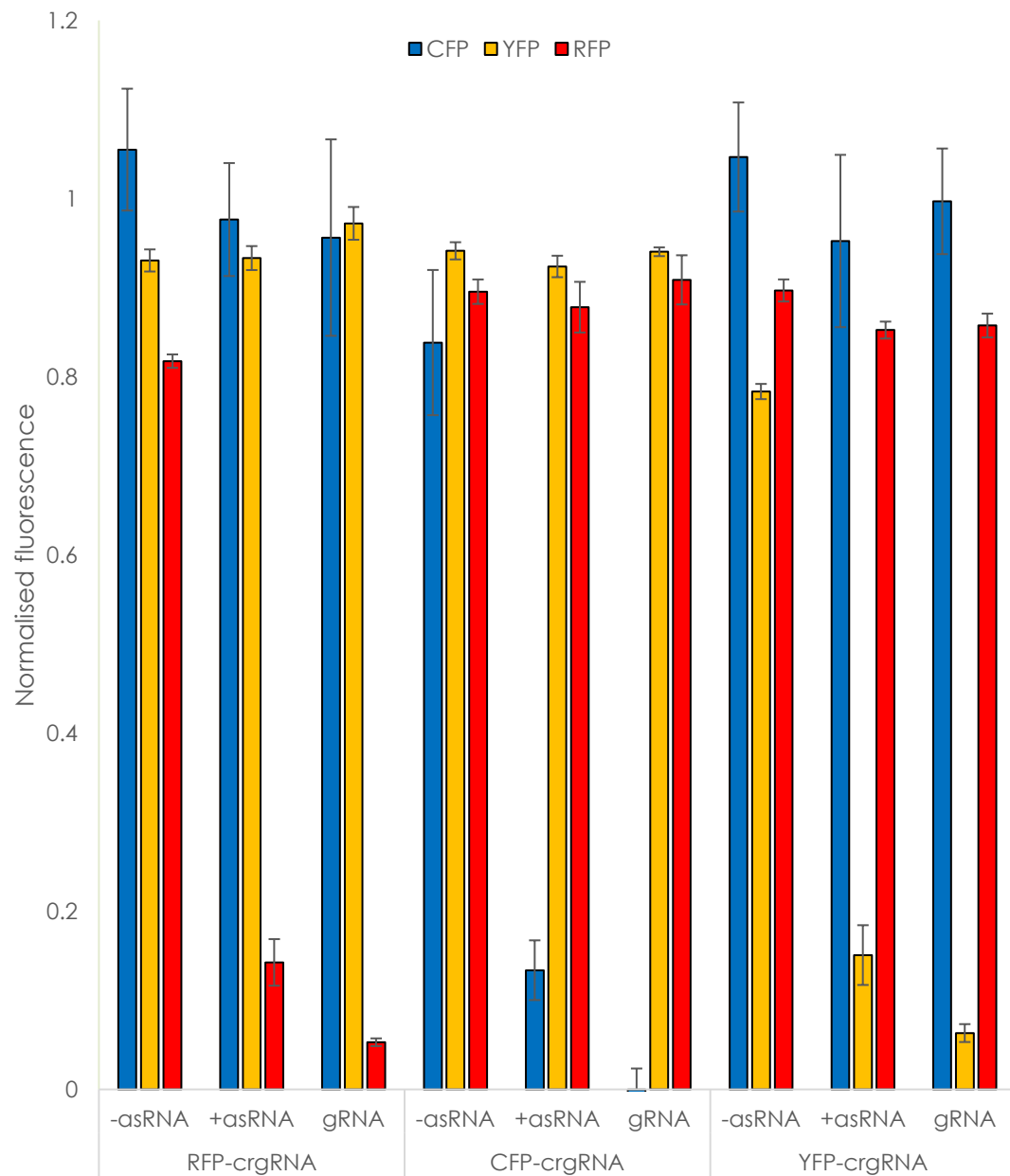


Figure 3.14 Three different crgRNAs are used to repress three different fluorescent reporter genes. gRNAs are crgRNA indicated by the colour but with the *cis*-repressing element removed. In each case '+asRNA' or '-asRNA' refers to the cognate asRNA to that crgRNA. Error bars are SEM, n = 24. Full data processing methods found in section 5.3.1.

3.2.13 Control Experiments

A number of control experiments were carried out to rule out non-specific effects on the cells from expression of components of the crgRNA asRNA system. Growth curves for strains for various conditions (with only crgRNA, with only asRNA alone, with both crgRNA and asRNA, and without crgRNA or asRNA) were measured (Figure 3.15A). There were no significant differences in the optical densities between different strains over 24-hs of growth. Therefore, there is no evidence that the expression of the crgRNA, the asRNA and the complex they form are toxic to the cell or impede growth when contrasted to an equivalent strain expressing neither the asRNA nor crgRNA.

The repression that results from expressing the system is dCas9 dependent, rather than being dependent on direct RNA-RNA interactions between the system and the mRNA of the target gene. This is demonstrated in Figure 3.15B where the system is expressed with or without dCas9. Repression is seen only when dCas9 is expressed.

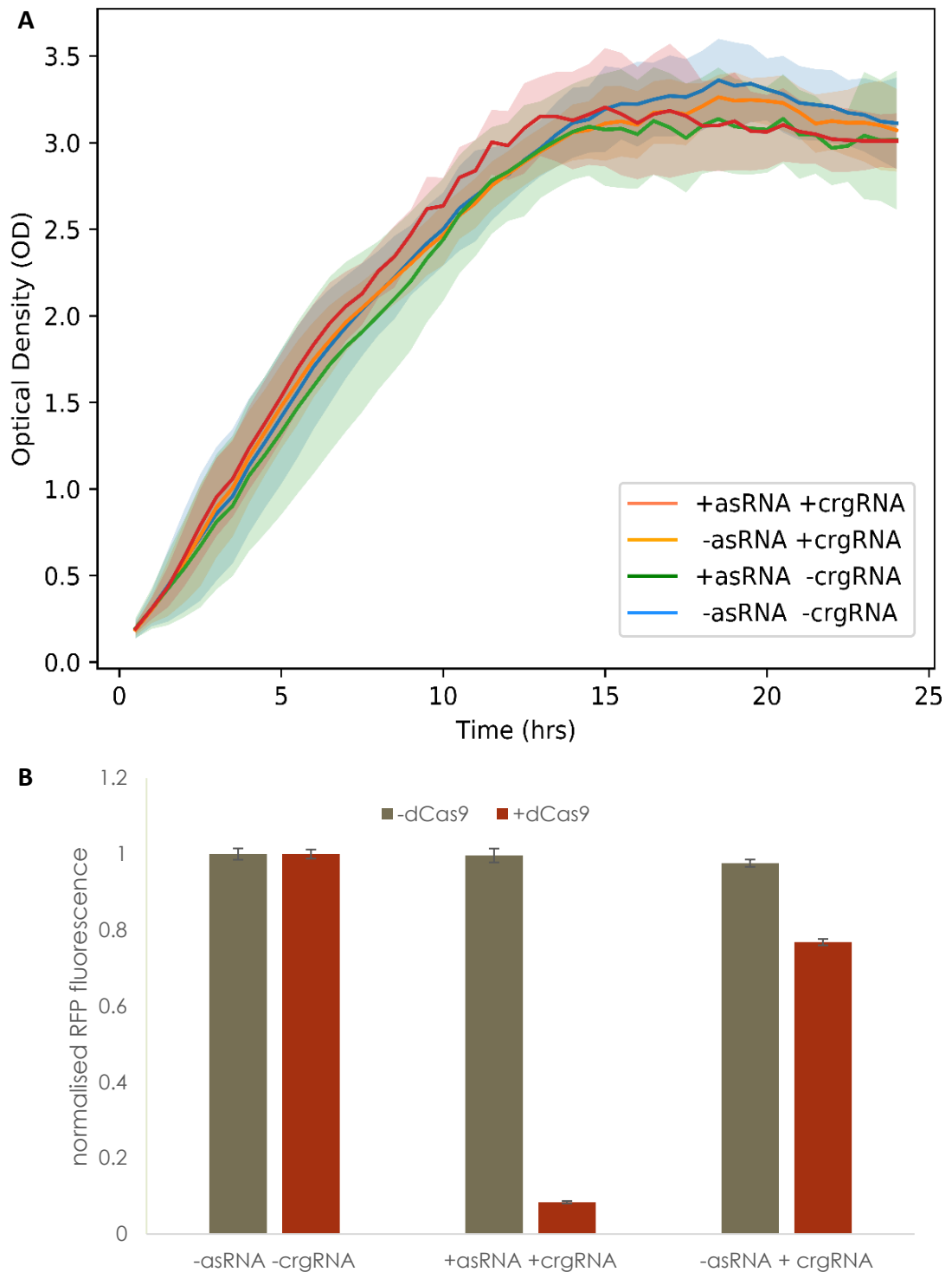


Figure 3.15 dCas9 dependency in the crgRNA asRNA system. (A) Growth curve of strains with different combinations of asRNA and crgRNA. There is no significant difference between the optical densities at the 24hr mark. Error bars rendered as translucent envelopes represent standard deviation from data set, n =24. (B) dCas9 dependency of crgRNA system. mCherry expression is only repressed by the crgRNA system when expressed with dCas9. Fluorescence is normalised to mCherry expressing control, error bars are SEM, n=24. Full data processing methods found in section 5.3.2.

3.3 Discussion

In this chapter is presented a framework for generating functionally complete NAND gates. The system designed here has a small genetic footprint and causes a low metabolic burden per additional gate while expressing a high degree of orthogonality. While there are a number of avenues that could be explored to further improve the system, this work forms a strong basis for further exploring and exploiting the potential crgRNA asRNA system for Boolean logic circuits in biological control systems.

3.3.1 No recruitment of dCas9 by asRNA

One of the concerns about extending the *cis*-repressing element into the Cas9 binding scaffold is that the corresponding asRNA which complements the entire *cis*-repressing element is therefore a complement for the sequence that complements the Cas9 binding scaffold. Thus, the extension of the *cis*-repressing element results in an asRNA with Cas9 binding scaffold sequence. However, complex formation between the gRNA and the dCas9 protein are dependent on structure more than sequence and in most cases mutations which change the sequence but not the structure of the gRNA do not affect its ability to recruit Cas9. Therefore, the fact that the longest asRNA only includes the 5' side of the first hairpin indicates that, while the asRNA contains Cas9 binding scaffold sequence, it does not exhibit the same folding or structure and, duly does not form a functional complex with dCas9. This fact is supported by the lack of transcriptional repression when the asRNA alone is expressed along with dCas9, confirming the finding by others that the length of Cas9 binding scaffold sequence in asRNA is less than the minimum sequence required for dCas9 functionality (Briner et al., 2014).

3.3.2 *Cis*-repressing element mechanism of inactivation.

The evidence presented in this chapter indicates that the addition of the Lp *cis*-repressing element to the gRNA transcript leads to a reduction in the level of dCas9 dependent repression. Similarly, the addition of the Lp asRNA to complement the *cis*-repressing element leads to a rescue in the level of repression of the system.

However, what is not clear from the evidence is how the *cis*-repressing element causes this change in repression. There are a number of steps in the formation of the final DNA-dCas9-RNA complex: at each of these steps the *cis*-repressing element may be interfering with the progression of complex formation. How complex formation between a gRNA and the dCas9 initiates, and which part of the gRNA forms the initial recognition interaction with the dCas9 is yet to be determined. Additionally, it is unclear how the gRNA transitions into complex formation with the Cas9, though there are detailed studies on the substantial conformational changes which occur within the protein during this process (M. Jinek et al., 2014). At either of these steps, the *cis*-repressing element may be interfering with complex formation. Similarly, the *cis*-repressing element may be interfering with complex formation between the ribo-protein complex and the target DNA.

The Sp variant exhibits *cis*-repression despite not complementing the Cas9 binding scaffold. It is possible that in the Sp variant the crgRNA complexes with the dCas9 and the hairpin sits in the same major cleft as the spacer region lies and prevents complex formation between the ribo-protein complex and the DNA. How this mechanism might change, when the *cis*-repressing element is extended in to the Lower Stem, Bulge and Upper Stem is unclear. Particularly when the extension into the bulge and upper stem sequence stabilizes spacer complementation and potentially interferes with ribo-protein complex formation, the level of *cis*-repression is reduced. This contrasts with the Lp variant where the level of *cis*-repression is significantly increased. The exact mechanism of the system functionality is not clear and requires further investigation.

3.3.3 Orthogonality Algorithm

In this chapter is presented a new algorithm for optimization of the system in terms of orthogonality. This algorithm nevertheless has many ways in which it may be improved. The interaction matrix, which was generated exhaustively, looks at the pairwise interactions of 50 toeholds from each Lp functional homologue (with different DNA targets) to be included in the same system. This means the number of

pairwise interactions which need to be scored, where n is the number of asRNA crgRNAs pairs in a system, is given by Equation 3.4. Since the amount of computational time required for this calculation rises as the square of the number of crgRNA asRNA pairs being assessed, having a large number of crgRNA asRNA pairs optimized for orthogonality in this manner becomes unfeasible.

$$\text{Equation 3.4: } \textit{interactions} = 50n^2$$

The levels of cross reactivity that a particular toehold has with one alternative toehold correlates with the score it will receive from others. It would be possible to increase the efficiency of the algorithm by removing toeholds with above-average cross-reactivity before completing the full pairwise interaction matrix.

One potential alternative approach would be to generate many more candidate toeholds for each crgRNA-asRNA pair and use a neural learning approach in which the neural network could pick a set of toeholds, one for each crgRNA pair. The toehold set would be scored and iteratively the neural network would learn which toeholds are more likely to produce lower scores. Building such a system would reduce the computational barriers for generating large numbers of orthogonal crgRNA-asRNA pairs to run in the same cell.

3.3.4 Observed orthogonality *in vivo*

When characterising different versions of this system there are multiple attributes a system has at which variation may occur. One variable which changes the level of repression is the DNA target which is chosen: a target closer to the transcription start site may have a higher level of repression than a DNA target which is in the coding region. Additionally, there is variation in the underlying gRNA which may have a high or low affinity with its DNA target depending on amongst other things the GC content of the seed region (Vigouroux et al., 2018). Finally, there are the two relevant factors in identifying the properties of the Lp version. Firstly, the *cis*-repressing element's ability to inactivate the rest of the transcript may vary due to

changes in the spacer and the toe-hold sequence. Secondly, can changes in the toehold and spacer sequences lead to changes in the ability of the asRNA to complex with the crgRNA rescuing repression activity. Identifying these two factors is important in understanding how the system is working and controlling for variation in these two factors is necessary for identifying orthogonality levels or the change in repression by a crgRNA resulting from the co-expression of a non-cognate asRNA.

An examination of the level of repression by the three crgRNAs and control gRNAs used in the orthogonality test show variation in the underlying level of repression from the gRNA; the same pattern is seen when each of the crgRNAs is expressed with its cognate asRNA. There was also variation in the level of repression of each of the crgRNAs (Figure 3.13B). When variation in these two factors was controlled for, the level of orthogonality was found to average 11% reporter repression from a crgRNA with non-cognate asRNA. This is a low level of cross-reactivity considering a large part of the asRNA transcript is uniform between sequences, it is also sufficient orthogonality for most biological applications.

3.3.5 Application of system to multiple genes

There are several issues raised in the Section 3.3.4 with variation in the level of repression by crgRNA systems due variation in repression by the underlying gRNA either due to DNA target or gRNA affinity. This is not found when the system is targeted to the TSS as is done as in section 3.3.5 (Figure 3.14) where no statistically significant difference in level of crgRNA repression was observed. When the system is used by a practitioner targeting repression of genes, the system should be targeted to the TSS.

Chapter 4: Discussion

4.1 Summary

The aim of this work was to generate a biological NAND gate for use in the cellular environment with a large potential library size, near-binary output, low metabolic burden and orthogonality. The approach taken has been to engineer gRNAs (crgRNAs) to require an asRNA for dCas9-mediated repression of an output promoter. The greatest challenges faced in this work have been (1) reducing the level of repression of the output when it is in an ON state, (2) increasing the level of repression of the output when it is in an OFF state and (3) maximising orthogonality between non-cognate RNAs.

The first positive results, with the K variant, produced an asRNA-dependent 1.86-fold change but highlighted the need to reduce the level of repression when the output is in the ON state, as expression of the crgRNA without the asRNA (which should give a high output) still resulted in 74% repression. Both of the main components in CRISPRi repression, the crgRNA and dCas9, were transcribed by inducible promoters, allowing their expression to be tuned to the optimal level by varying the level of inducer. For the K variant, the optimal level of induction was 0 pg/mL aTc (for dCas9 expression) and 50 μ M IPTG (for crgRNA expression). Leaky expression of dCas9 from the uninduced promoter therefore provided sufficient repression. By extrapolating the induction curve for dCas9 to lower expression levels than were possible with this experimental setup, greater fold changes were predicted. Hence, lower expression dCas9 plasmids were investigated (see Section 4.2.2).

Meanwhile, alternative designs were tested that used 3' *cis*-repressing elements (the W and Y variants), rather than the 5' *cis*-repressing element used for the K variant. Although induction levels were optimised for the W and Y variants, the K variant still proved to have the highest fold change of the synthesized variants.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Subsequently, a time series experiment demonstrated that the system's performance was dependent on growth phase, with lower levels of repression observed during exponential phase when contrasted to stationary phase. This might be explained by DNA replication-induced dissociation of dCas9 from the target DNA, as well as dilution of dCas9 during cell division, which occurs at a higher rate during exponential phase. As such, the optimum level of induction of dCas9 was higher in exponential phase than when measuring at stationary phase.

The addition of a transcriptional attenuator to the dCas9 expression plasmid (pdCas9-T) allowed the characterisation of the K variant at lower levels of dCas9 expression, increasing the fold change to 4-fold, surpassing the previous maximum of 3-fold.

To further reduce the level of repression resulting from the crgRNA, a new generation of variants were created in which the complementarity of the *cis*-repressing element was extended into the dCas9 binding scaffold of the transcript to varying degrees. Of this new generation, the Sp variant displayed the least repression when expressed without the asRNA (ON state) and the greatest repression when expressed with it (OFF state). The *cis*-repressing element of the Sp variant complements the spacer, lower stem, bulge, upper stem and loop sequences as contrasts with the K variant, where the *cis*-repressing element only complements the spacer.

Next, orthogonal versions of the Sp variant were created. However, the sequence constraints imposed by the Sp variant complicates the design of orthogonal versions as the sequence of sections of the *cis*-repressing element (and therefore asRNA) is defined by the sequence of the dCas9 binding scaffold which is complemented. As such, an *in silico* approach was taken to maximise the orthogonality of the toeholds used. This was done with a pair-wise interaction prediction and the optimisation was conducted using a two-step approach involving a Monte Carlo optimisation followed by a genetic machine learning optimisation. This approach was used to

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

optimise five crgRNA/asRNA pairs. Three were tested for orthogonality through exhaustive pairwise *in vivo* interaction testing. Despite the high sequence identity between the different crgRNA/asRNA pairs, only a 17% average crosstalk was recorded. In addition, versions of the Lp variant were used to regulate three different fluorescent reporter proteins. The levels of repression exhibited by each of the versions showed no statistically significant differences, demonstrating the ability to generate a library of functionally uniform versions.

In this thesis a new biological NAND gate is presented. This NAND gate compares favourably with other attempts to make guide RNA functionality dependent on two inputs when evaluated in terms of how close to binary the output is (discussed in Section 4.2.1). The gate requires low level background expression of dCas9, but since the gate is dependent on the interaction of two small RNA transcripts each additional gate only requires a genetic footprint of 304 bp per gate of which only 196 are the actual RNA sequences. The metabolic load produced by the gate was negligible. The protocol presented here for generating this NAND gate exploits the predictable nature of RNA folding and interaction allowing the creation of an almost arbitrarily large library size.

4.2 Future work

4.2.1 Logic circuit layering

The ultimate ambition of this line of work would be a biological computer based on Logic implemented at the molecular level. Though there are many applications which may successfully exploit progressions in biological logic and computing before this lofty goal is achieved. One of the first steps from logic gate to biological computer is to produce a system which is able to conduct complex logical operations by layering gates into more sophisticated logic circuits as has been attempted with a number of other logic gates (Rosado, Cordero, & Rodrigo, 2018; Wong, Wang, Poh, & Kitney, 2015). The logic gate described here exhibits some attributes beneficial to layering as well as some which may be detrimental. First, the input of this gate, polymerase flux or transcription of the two input RNAs is the

same as the output from the gate, polymerase flux or transcription from the output promoter which makes it well suited for layering. The ultimate ambition of this line of work is to generate a biological computer, based on logic implemented at the molecular level.

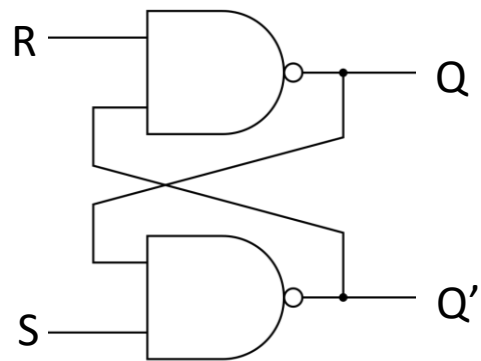
A potential difficulty with applying layering to this transcription-based system however, is the propensity for signal drift; as described in section 1.2.2, the voltage transfer characteristic is used to characterise electronic logic gates, and is a useful concept when considering biological logic gates (in this system voltage is analogous to rate of transcription or expression level). In the gate presented here, as in the case of other biological and electronic logic gates, there are indications that logic gate layering may lead to a degree of 'signal drift'. Signal drift occurs when the binary nature of a signal degrades as it passes through layers of the system. For example in a set of inverters connected in series with a binary 1 or 0 input, if the input to the first inverter is 1, but the output is repressed to 0.1 rather than 0. Then this low-level input to the next inverter may result in an output of 0.8 (rather than 1.0), which may then produce an output of 0.3 rather than 0, after the next inverter, and so on. The signal quality is therefore degraded at each successive logical layer. This behaviour can limit the number of logical layers that may be used.

The NAND gate described here has high and low outputs which neither reach 1 nor 0 (19% and 91% repression respectively). The optimisation of component expression levels detailed in (sections 2.2.8, 2.2.11, 3.2.2) have shown that repression efficiency is dose-dependent on the level of crgRNA expression. In a layered system crgRNA would be both the output of one gate and the input to the next gate, and so it can be anticipated that signal drift would occur.

The issue of signal drift has already been faced and addressed in computer architecture at the electronics level; in order to deal with signal drift and asynchrony (signal timing), logic circuits in modern computers output to a 'flip-flop' (Figure 4.1). The flip-flop is based on a positive feedback loop, that pushes non-

binary values to the closest binary value, e.g. 0.3 becomes 0, and 0.7 becomes 1 (Tanenbaum & Goodman, 2005). The simplest flip-flop, the 'SR NAND flip-flop', can be constructed using two connected NAND gates (Figure 4.1A), while the flip-flop which is most commonly used to store values in a computer is the 'gated D flip-flop' (Figure 4.1 B). Notably both the implementations displayed in Figure 4.1 are entirely composed of NAND gates (the type of logic gate presented in this thesis). Whether the crgRNA asRNA-mediated logic gate could be used to form a flip-flop or not would depend on whether the logic gate exhibited cooperativity. Even if it were found not to be possible to utilise asRNA and crgRNA to form flip-flops, there are other candidate flip-flop and toggle switch mechanisms which may be used to combat signal drift in a biological system, and extend the degree of layering possible (Atkinson, Savageau, Myers, & Ninfa, 2003; Bothfeld, Kapov, & Tyo, 2017; Gardner, Cantor, & Collins, 2000).

A



B

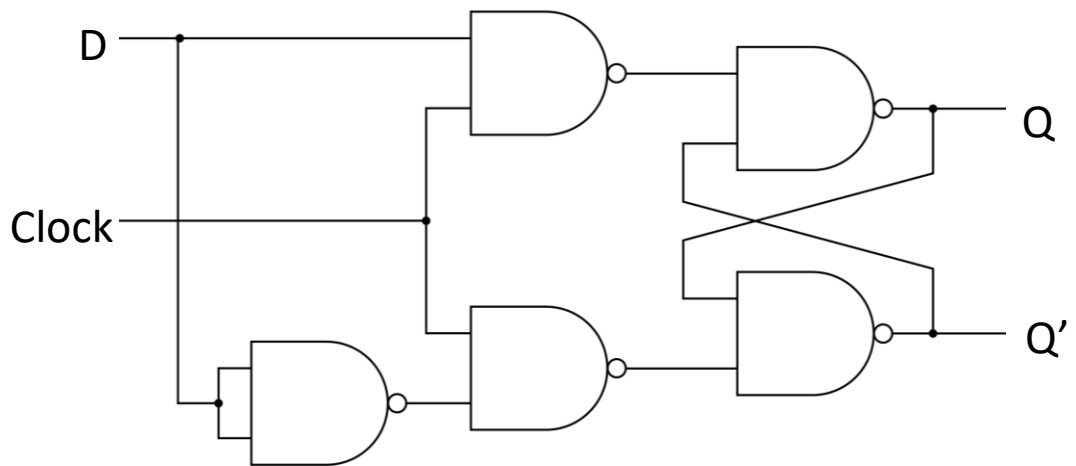


Figure 4.1 Flip-flops can be used to store data and rescue signal degradation. (A) The SR flip-flop is the simplest flip-flop, the S input is used to set the value stored, and the R input resets the flip-flop. The output Q maintains the signal input from S until reset. Q' will always have the opposite value to Q. (B) The gated D flip-flop is commonly used to store values in a computer (SRAM). Input D is the 'data' input and sets the value that the flip-flop holds. The clock input is used to keep the system in synchrony.

4.2.2 Reduced background repression

When the inactive crgRNA is expressed without the asRNA there is still a low level (19%) dCas9 dependent repression of the output gene (basal repression). RNA exists within a structural ensemble in which the RNA may pass briefly through energetically unstable structures. Complex formation between the gRNA and dCas9 is attended by large conformational changes in the dCas9 protein, and represents a large $\Delta\Delta G$ making the reaction principally one directional as the dissociation constant is very low. The levels of repression observed indicate the crgRNA exists in a structural ensemble in which some of the more rare structures are able to complex with dCas9. This complex formation has a low dissociation constant and so leads to a significant level of repression. One way to reduce this effect would be to express a nonsense gRNA in stoichiometric excess of the dCas9, thus reducing the probability of dCas9 complexing with transient crgRNA structures. However, with this approach the nonsense gRNA could also compete with the crgRNA asRNA complex for the unbound dCas9 pool, requiring a slightly higher level of expression of dCas9 to reach the same level of active state crgRNA repression. Also the reduced background repression may allow the use of a higher level of dCas9 expression to be used resulting in greater crgRNA +asRNA repression. Hence, the constitutive expression of a nonsense gRNA may improve fold change in crgRNA-mediated repression of the target gene by simultaneously reducing basal repression in the absence of asRNA and increasing maximal repression in the presence of asRNA.

4.2.3 Potential network size

When considering how many gates may be included in the single cell there was negligible metabolic load from a single gate and there are 275 predicted sRNA systems alone in *E. coli* (Rivas, Klein, Jones, & Eddy, 2001). If an investigator is prepared to impose a metabolic burden on cells it may be possible to add more than 275 gates to a cell. For more context, when understanding the value of a

number of logic gates, the guidance computer of the spacecraft Apollo was composed of only 4,100 logic gates (O'Brien, 2010).

How many gates may be expressed while maintaining orthogonality is a different question. In this thesis, the level of orthogonality expected for a library of 5 logic gates is ascertained using measurements from a subset of three of the five gates. The level of orthogonality seen is achieved by optimising from a base of 50 screened candidate toeholds per gate. Using a different optimisation approach, possibly heuristics or machine learning, it may be possible to optimise from a larger pool of screened candidate toeholds per gate and generate sets of gates with many more members and a lower orthogonality index (more orthogonal).

4.2.4 Future work: autoregulation

The evidence presented here demonstrates the variation in dCas9 level induction that is necessary for optimal performance in different growth phases; a low level of expression in the stationary phase and a high level of expression in the exponential phase, with an apparent gradient required as the cultures transition from the exponential phase into the stationary phase. This provides an interesting problem; how to produce a system which performs at close-to-optimal levels, irrespective of the growth phase. To address this challenge there are several designs that could be explored in future work. One way of introducing a negative feedback loop to smooth the variation in levels of repression seen between growth phases, would be to make dCas9 expression auto-regulated. If dCas9 is used to repress the expression of dCas9 then this creates a negative feedback loop for the repression level created by dCas9. The level of repression could be tuned through the introduction of mismatches into the seed and non-seed regions of the spacer (Vigouroux et al., 2018). Such a mechanism, if placed on the genome, may produce a noisy output due to a single copy having a binary level of expression; dCas9 is either bound or unbound. But placed on a plasmid there are many copies and output becomes dependent on the percentage of plasmids which have the dCas9 bound. Combined with the dynamical error introduced by the slow degradation rate of previously

expressed dCas9, this would have a smoothing effect. Such a mechanism to produce a cellular circuit with performance closer to the optimum, irrespective of growth phase, holds the promise of increasing the utility of such a system.

4.3 Contrast with other systems

4.3.1 Contrast with a gRNA antisense system

In the mechanism put forth in this thesis, an element designed to repress the gRNA is included in the gRNA transcript and an asRNA sequesters this *cis*-repressing element to rescue gRNA function. The closest comparable work explores direct sequestration and degradation of a gRNA by an asRNA (Lee et al., 2016). Lee et al., (2016) used three rounds of improvement, initially the asRNA complemented the spacer region of the gRNA (Figure 4.2). As well as inhibiting transcriptional repression by the dCas9 and the gRNA, an Hfq binding scaffold was included in the asRNA transcript. Hfq stabilises the asRNA and promotes interactions with other RNAs as well as recruiting RNase-E to degrade the two RNAs (Morita & Aiba, 2011). This initial method of sequestration and degradation led to a derepression of 15% which rose to 43%, 55% and finally 95% through the extension of the complementary spacer, exchange of the MicF Hfq scaffold for Spot42 and finally swapping the location of the system from the 5' end of the gRNA transcript into an extended linker in the 3' end of the gRNA transcript. This allowed the investigators to further decrease the $\Delta\Delta G$ of asRNA binding.

Whereas the mechanism presented in this work is dependent on two interactions and therefore has two, sometimes competing, objectives; the Lee *et al.* (2016) mechanism is based on one interaction and has one objective. In the mechanism put forth by Lee *et al.* (2016), the asRNA sequesters and degrades the gRNA producing as high a derepression percentage as possible. The two interactions in the work presented here are firstly the interaction between the *cis*-repressing element and the crgRNA, and secondly the interaction between the asRNA and the *cis*-repressing element. The first interaction, equivalent to derepression, ensures an output value as close as possible to 1 when the output is ON. The second

interaction, equivalent to de-derepression (rescuing repression activity), ensures an output value as close as possible to 0 when the output is OFF. This second interaction means that avenues for optimising the first one, do not include degradation of the crgRNA, an avenue open to and exploited by Lee *et al.* (2016) through the use of Hfq. However, the mechanism adopted in this thesis does allow the repressing element to be included in the gRNA transcript (making a crgRNA), an avenue not open to Lee *et al.* (2016). Inclusion in the same transcript means the *cis*-repressing element is in proximity with the target region within the gRNA during folding of the nascent RNA transcript, thus increasing the rate of initial interactions and stabilising the resulting structure.

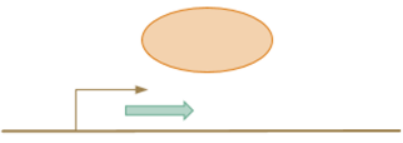
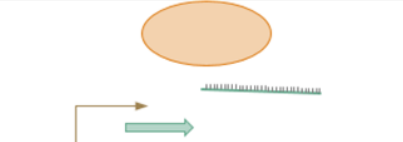
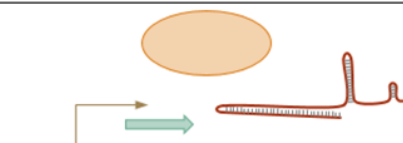
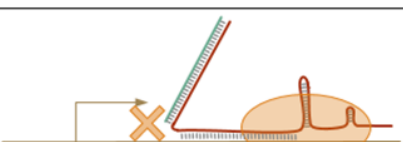
To achieve an optimal performance for this system, the level of expression of dCas9 had to be reduced to below the saturation level. Similarly, to achieve the greatest percentage de-repression, Lee *et al.* (2016) reduced the level of dCas9 induction to 200 pg/mL aTc (from an initial 10 ng/mL aTc), although the rationale given was to reduce toxicity. While the Lee *et al.* (2016) mechanism expresses dCas9 from the same plasmid as used here, both the gRNA and antisense RNAs are expressed from a plasmid with the high copy ColE1 origin by Lee *et al.* (2016) rather than the medium copy pMB1 used in this work (the same gRNA promoter was used). This is one possible reason for the difference in optimal dCas9 expression levels between the two systems, together with the different requirements of the two mechanisms and the use of different strains.

Both mechanisms have two input RNAs which combine in a logical operation to yield a repressed or non-repressed output (analogous to OFF/ON or 0/1) and as such, both can be represented as logic gates. The crgRNA system equates to a NAND gate, as the output is only repressed when both input RNAs are expressed (or ON) (Figure 4.2). The Lee *et al.* (2016) antisense/gRNA system on the other hand equates to an IMPLY gate as the output is only repressed when the gRNA input is ON and the asRNA input is OFF. The NAND gate is a functionally complete logic gate, meaning any logical operation can be generated through combination of NAND

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

gates. The NAND gate is a symmetric gate, commonly used in computer processor design (Tanenbaum & Goodman, 2005). The IMPLY gate however is asymmetric and appears rarely in computational design. Lee *et al.* have therefore presented their mechanism in terms of gene regulation more generally, rather than emphasising use as a logic gate.

NAND Gate

gRNA	asRNA	Mechanism	output
0	0		1
0	1		1
1	0		1
1	1		0

IMPLY Gate

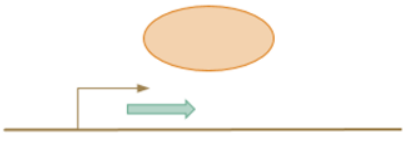
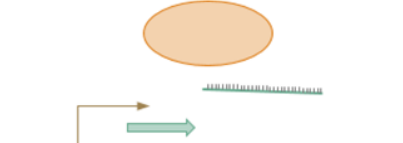
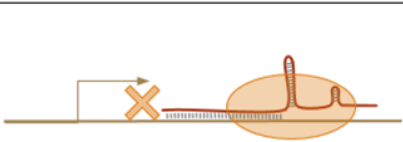
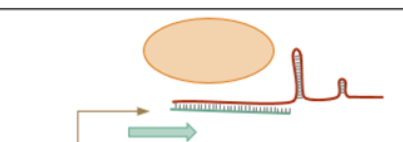
gRNA	asRNA	Mechanism	output
0	0		1
0	1		1
1	0		0
1	1		1

Figure 4.2 Truth tables of the crgRNA system NAND Gate and the antisense gRNA IMPLY gate with mechanistic schematics.

4.3.2 Contrast with riboswitched gRNAs: study I

Many approaches to generating biological logic gates have been published, and similarly a great number of engineering approaches explored for transcriptional regulation using the dCas9 system (Lee & Moon, 2018). Here, the advantages of dCas9 transcriptional regulation are applied to the requirements of a logic platform. Basic dCas9 repression requires two main components; the dCas9 protein and the gRNA. Both of these components have the potential to be engineered. The protein element can be engineered in a similar manner to other protein transcription factors such as TetR through fusing of additional functional domains to the protein (Peres-Pinera *et al.*, 2013; Roybal *et al.*, 2016). The gRNA offers the opportunity to use the predictable nature of RNA folding to engineer transcriptional repression in a way that is novel amongst other transcriptional regulation systems. One of the main ways this opportunity has been exploited has been through the inclusion of aptamers in the gRNA transcript (Tang, Hu, & Liu, 2017). These have been used to recruit proteins in a gRNA specific manner to the gRNA target binding site. Aptamers have also been used to make gRNA functionality ligand-dependent (Liu *et al.*, 2016).

The main challenge encountered in this work has been to use the predictable nature of RNA folding, to develop a sequence which can be included in a gRNA transcript which will effectively inactivate/abrogate gRNA activity in a reversible manner. This is a similar challenge to that faced by those who wish to make gRNA functionality dependent on a ligand, through the inclusion of riboswitches and aptazymes. It is therefore instructive to compare and contrast the approaches taken in papers producing ligand-dependent gRNAs to the approaches taken in this chapter.

Tang, *et al* (2017) included, in two different instances, either the theophylline aptazyme or the guanine aptazyme in the gRNA transcript in combination with a “blocking sequence”. This blocking sequence complemented part of the gRNA sequence to abrogate activity. Tang, *et al* (2017) explored a number of approaches

including complementation of the spacer region, separating the gRNA into the crRNA and tracrRNA, and using the blocking sequence to aggregate complementation between the crRNA and tracrRNA to create a functional unit. Tang, *et al* (2017) found that the most effective of the three approaches was using the blocking sequence to complement the spacer region, the same approach pursued in the mechanism presented in this thesis. Tang, *et al* (2017) also found that the greatest repression of gRNA activity was achieved with the longest complementation within the blocking sequence that they tried (a similar correlation to that found in the library tested here; section 3.2.2). The 17nt maximum blocking sequence used by Tang, *et al* (2017) contrasts with the 20nt *cis*-repressing element used in the K design variant. The mechanism presented here displaces the *cis*-repressing element from the gRNA, rather than relying on post cleavage duplex melting and disassociation. As a result, the inclusion of bulges in the blocking sequence lead to an increase in ON state activity.

Tang, *et al* (2017) explored the use of a theophylline aptazyme gRNA for theophylline-dependent cleavage of DNA using Cas9, and also explored the use of a guanine-dependent aptazyme for guanine-dependent activation of a promoter through the use of dCas9 fused to a transcriptional activator. Direct comparisons between the ON and OFF states achieved by Liu *et al* and those presented in this thesis are challenging. Tang, *et al* (2017) used a mammalian system, whereas the system presented in this chapter was characterised in a bacterial system (*E. coli*). The most immediately comparable of the two approaches taken by Tang, *et al* (2017) would appear to be the guanine-based system as it is used to regulate transcription, although the lack of a positive control using a wild type gRNA means the induction can only be presented in terms of fold increase from the negative control. This makes it impossible to assess what percentage activity of the guanine ribosyme-gRNA is rescued by the addition of guanine. Theophylline, on the other hand, has both controls, although it is measuring DNA cleavage catalyzed by Cas9 in a mammalian system, rather than transcriptional repression. The theophylline-

ribozyme-gRNA exhibits a background cleavage rate of 23% in the absence of theophylline, and 58% cleavage rate at 2 mM theophylline. This compares favourably with the +asRNA 91% repression and –asRNA 19% repression presented in Chapter 3:

4.3.3 Contrast with riboswitched gRNAs: study II

Liu *et al.*, 2016 generated a series of gRNA containing aptamers to make the gRNA functionality dependent on interactions between the aptamer and the aptamer ligand. In this paper, they demonstrated the applicability of their method to both small molecule-dependent riboswitches and protein-dependent riboswitches. This additional functional module was inserted in the 3' end of the gRNA transcript. An antisense sequence with engineered, ligand-dependent nucleotide availability from the 3' module was complemented within the 5' spacer region. In seeking to achieve bistability so as to maximise the dynamic range achievable through the addition of the riboswitch ligand, a number of different antisense lengths were explored. Irrespective of the $\Delta\Delta G$ of ligand binding, they found an antisense of 15 bases maximised ligand-dependent change in repression or activation. If the antisense region had a length of 18 bases, the stable interactions lead the gRNA to have reduced activity with low dependence on ligand concentration. Conversely, when the antisense region was 11 bases the gRNA retained its activity also with a low dependence on ligand concentration. The percentage repression efficiencies of gRNAs into which Liu *et al* had engineered with 15 nt antisense showed relatively little variation in repression efficiency when the cognate aptamer ligand was absent (1~2%). When the ligand was added to the system the level of repression ranged between variants within the library from 34% to 77%. This compares with the +asRNA 91% repression and –asRNA 19% repression presented in this thesis. The authors comment on this variability, including in the method that several variants must be generated and the optimal variant selected.

While the authors do discuss the substitution of the two 3' hairpins in the gRNA for other aptamers, they don't discuss the design decision to include the riboswitch

module in the 3' terminus of the transcript or whether any designs including the module in the 5' of the gRNA were explored in preliminary work. The decision to place a riboswitch functional module in the 3' terminus of the gRNA transcript leads to a structure in which complementation between the antisense and spacer regions has a substantial cost to entropy. To maintain a ΔG which would render this a stable structure in the absence of the ligand, requires a higher enthalpy of the interaction between the antisense and spacer regions of the gRNA. Other than changing the GC percentage, this can be achieved by increasing the number of complementary nucleotides. Hence, by locating the riboswitch module in the 3' terminus rather than 5' terminus of the gRNA transcript, a larger number of complementary nucleotides are required to maintain the same ΔG . This is important both for maintaining stability and for switching upon the addition of the ligand. The increased number of complementary nucleotides for this design schema potentially offers greater interference with complex formation between the gRNA, dCas9, and the target DNA in the absence of the riboswitch ligand. In contrast, this approach is not advantageous for the mechanism presented in this thesis as the area that requires greatest optimisation is the full inactivation of the gRNA. Reduced entropy effects for the inactivated state of the gRNA of having the *cis*-repressing element in the 5' terminus rather than the 3' terminus provide an advantage for inactivating the gRNA. Consequently, whilst fusing the *cis*-repressing module to the 3' terminus is advantageous for Liu *et al.* conversely, fusing the *cis*-repressing element to the 5' terminus of the gRNA is advantageous for the crgRNA presented here.

Liu *et al* also take this one step further and arrange their circuitry in the form of logic gates. They attempt all of the Boolean logic gates. Whilst these certainly have logical value they are also flawed. The two inputs of the logic gates are in the form of ligands to modulate the two riboswitched gRNAs, and in all cases (except the XOR logic gate) the combinatorial effect of the two inputs is evaluated through two complexes binding to/near the promoter, as opposed to one. Consequently, the effect of having one input in the ON state instead of two, is an intermediate result,

approximate to being halfway between having no inputs (OFF) and both (ON) (an output value which might approximate to 0.5). Additionally, there are barriers to the layering of these gates as making the output of one gate the input of the other would require that the expression of the output promoter of one gate would result in the synthesis of the input ligand of the next gate. Similarly, developing a library of gates is a laborious process requiring the creation of new ligand responsive riboswitches, and their inclusion into the gRNA framework. To both these points, the system described here is far easier to expand, with simple design rules allowing reliable synthesis of RNA-dependent logic gates. This is because the design is dependent on predictable RNA-RNA interactions, without the complexities of aptamer-ligand interactions. In a similar vein, the logic output of the system presented here, is dependent on the direct interaction of two RNAs, creating an output closer to the binary 1 or 0 than the additive effects of two effectors. It is worth noting that the critique presented here is entirely in terms of the objectives of this project, which are quite different to the objectives of the authors in terms of the application of the circuits to cancer, which should not be understated.

Chapter 5: Methods and Materials

5.1 Materials

Glycerol stock

20% glycerol (v/v) dissolved in dH₂O and autoclaved at 115 °C for 15 minutes.

Stored at room temperature

Ampicillin

Stock concentration was 100 mg/ml in dH₂O, filter sterilised (0.22 µm). Working concentration was 100 µg/ml. 1 mL aliquots stored at -20 °C.

Kanamycin

Stock concentration was 50 mg/ml in dH₂O, filter sterilised (0.22 µm). Working concentration was 100 µg/ml. 1 mL aliquots stored at -20 °C.

Chloramphenicol

Stock concentration was 40 mg/ml in 100 % ethanol; working concentration was 40 µg/ml. 1mL aliquots stored at -20 °C.

Anhydrotetracycline (aTc) stock

200 µg/mL aTc dissolved in 50% ethanol, 50% dH₂O, filter sterilised (0.22 µm), stored at -20 °C.

Isopropyl-beta-D-thiogalactoside (IPTG) stock

500 mM IPTG prepared in 1 mL aliquots stored at -20 °C.

Lysogeny broth (LB)

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

To make 1 Litre, 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl. pH adjusted to 7.5 with NaOH, autoclaved at 115 °C for 15 minutes. Requisite inducers and antibiotics added to LB.

S.O.C broth

To make 1 Litre, 20 g Bacto Tryptone, 5 g Bacto Yeast Extract, 2 ml of 5 M NaCl, 2.5 ml of 1 M KCl, 10 ml of 1 M MgCl₂, 10 ml of 1 M MgSO₄, 20 ml of 1 M glucose. pH adjusted to 7.0 with NaOH, autoclaved at 115 °C for 15 minutes.

LB Agar

15 g Agar was added to 1 L of LB broth prior to autoclaving. Agar was cooled to *ca* 55°C before adding requisite antibiotics.

Tris-acetate-EDTA buffer (TAE buffer)

To make 1 litre: 4.844 g Tris Base, 1.21 g Acetic Acid, 0.372 g EDTA.

Tris EDTA buffer (TE buffer)

To make 1 litre: 10 mL of 1 M Tris base (pH to 8.0 using HCl), EDTA; 2 mL 0.5 M, dH₂O 988 mL.

Transformation Buffer 1 (TFB1)

30 mM CH₃COOK, 100 mM RbCl, 10 mM CaCl₂, 50 mM MnCl₂, 15% (v/v) glycerol; pH adjusted to 5.8 with acetate; sterilised by filtration (0.22 µm)

Transformation Buffer 2 (TFB2)

10 mM MOPS (pH 6.5), 75 mM CaCl₂, 10 mM RbCl, 15% (v/v) glycerol; pH adjusted to 6.5 with KOH; sterilised by filtration (0.22 µm)

Restriction enzymes:

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

All restriction enzymes described were purchased from NEB

QIAprep Spin Miniprep Kit

Kits were purchased from QIAGEN and used according to the manufacturer's instructions. Plasmid mini-prep kits were prepared from 5 mL overnight cultures in LB broth, with supplements/antibiotics as required. DNA was stored at -20°C until required.

Q5

5X Q5 Reaction Buffer, 10 mM dNTPs and 2 U/μl Q5 High-Fidelity DNA Polymerase were purchased from NEB

Agarose

UltraPure™ Agarose was purchased from Invitrogen.

QIAquick Gel Extraction Kit

Kits were purchased from QIAGEN and used according to the manufacturer's instructions. DNA was stored at -20°C until required.

10X T4 DNA ligase and ligase buffer

Purchased from NEB (B0202S)

Gel Loading Dye, Purple

Purchased from NEB (B7024S)

5.1.1 List of plasmids

Sequences are to be found in

Appendix 2: Nucleotide sequences

pZS2-123	pBR322-LS-c
pdCas9 addgene #44249	pBR322-Bg
pdCas9-T	pBR322-Bg-c
pdCas9-RBS2	pBR322-US
pdCas9-RBS3	pBR322-US-c
pBR322	pBR322-Lp (- α)
pBR322-K	pBR322-Lp (- α)-c
pBR322-Kc	pBR322-Lp- β
pBR322-W	pBR322-Lp- β
pBR322-Wc	pBR322-Lp- γ
pBR322-Y	pBR322-Lp- γ -c
pBR322-Yc	pBR322-Lp-CFP
pBR322-Z	pBR322-Lp-CFP-c
pBR322-gRNA	pBR322-Lp-YFP
pBR322-Sp	pBR322-Lp-YFP-c
pBR322-Sp-c	pBR322-LP-crgRNA-A/asRNA-A
pBR322-LS	pBR322-LP-crgRNA-A/asRNA-B
pBR322-LP-crgRNA-A/asRNA-C	

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

pBR322-LP-crgRNA-A/asRNAc

pBR322-LP-crgRNA-B/asRNA-A

pBR322-LP-crgRNA-B/asRNA-B

pBR322-LP-crgRNA-B/asRNA-C

pBR322-LP-crgRNA-B/asRNAc

pBR322-LP-crgRNA-C/asRNA-A

pBR322-LP-crgRNA-C/asRNA-B

pBR322-LP-crgRNA-C/asRNA-C

pBR322-LP-crgRNA-C/asRNAc

5.1.2 *E. coli* strains

All cloning and experimentation was performed in Top10 from lab stock generated from NEB commercial competent cells. Genotype: F- mcrA Δ (mrr-hsdRMS-mcrBC) ϕ 80lacZ Δ M15 Δ lacX74 nupG recA1 araD139 Δ (ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ -

5.2 Molecular Biology Methods

5.2.1 Storage of *E. coli* strains

For short term storage, strains were kept on LB agar plates with appropriate antibiotics in a fridge at 4°C. For long term storage, *E. coli* strains were stored at -80 °C in 15 % (v/v) glycerol, either in a volume of 500 μ L in a cryotube or in 350 μ L volume in the well of a round bottomed 96 well plate. To prepare *E. coli* glycerol stocks, 5 mL LB containing the appropriate antibiotic was inoculated with the desired *E. coli* strain and grown for 16-24 hr (shaking at 200 rpm) at 37 °C before being mixed with 20 % (v/v) glycerol.

5.2.2 Preparation of chemically competent cells

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Chemically competent cells were prepared using the RbCl method. 5 mL of LB was inoculated with *E. coli* TOP10 and grown at 37 °C for ~ 16 hr (shaking at 200 rpm). The culture was diluted 1:400 in 500 mL LB containing 20 mM MgSO₄ and grown at 37 °C (shaking at 200 rpm) until the OD_f reached ~ 0.5. The cells were transferred to ten 50 mL Falcon tubes (50 mL per tube) and incubated on ice for 10 min. The cells were kept cold from this point on. The cells were centrifuged (3750 RCF for 5 min at 4°C), re-suspended in 20 mL cold TFB1 buffer, and incubated on ice for 5 min. The cells were centrifuged as before, re-suspended in 2 mL cold TFB2 buffer, and incubated on ice for 15 min. Cells were aliquoted, snap-frozen on dry ice and 100 % ethanol, and stored at -80°C.

5.2.3 Transformation into *E. coli*

Competent cells are thawed on ice, the aliquot was divided into 50 µL per reaction for cloning reactions or 20 µL for purified plasmid DNA. For single transformations and cloning reactions, the transformation was performed in a 1.5 mL Eppendorf micro-centrifuge tube. For the large number of three plasmid system transformations, the transformation was conducted in a 96 well plate (Manufacturer: Ritter, model: riplate PP - 1 mL, Part number: #43001-0116).

Between 1 µL and 10 µL of cloning reaction or purified plasmid was added to the competent cells and allowed to incubate for 20 min on ice. The cultures were then heat shocked at 42°C for 45 seconds before being returned to the ice bath. Due to the large thermal mass of the 96 well plate, both stages of the heat shock were performed with agitation to the 42 °C water bath or ice bath to ensure the cells reached the correct temperature rapidly. For cultures in 1.5 mL Eppendorf tubes, 1 mL of SOC media is added. For transformations in 96 well plates 200 µL of SOC media is added to each reaction. Transformations were allowed to recover for 1 hour with shaking 200rpm in a 37 °C incubator. 1.5 mL Eppendorf tubes are centrifuged for 1 min at 18,626 G, the excess supernatant poured off and the pellet re-suspended in the remaining ~50 µL supernatant before being spread on an LB agar plate containing the relevant antibiotics. Transformations in 96 well plates are

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

transferred directly to LB agar plates containing the relevant antibiotics. LB agar plates are dried before being inverted and incubated at 37 °C for ~ 16 hr.

5.2.4 Plasmid preparation

5 mL of LB containing the appropriate antibiotic was inoculated with a colony of *E. coli* containing the desired plasmid and grown at 37 °C for ~ 16 hr (shaking at 200 rpm). Plasmid extraction was performed using the QIAprep Spin Miniprep Kit according to manufacturer's instructions, eluting in 20 µL elution buffer.

5.2.5 Polymerase Chain Reaction (PCR)

PCR reactions were performed using Q5 DNA Polymerase.

For A standard 25 µL Q5 PCR reaction:

Component	Final concentration	Volume (µL)
Q5 reaction buffer - 5x	1X Q5 reaction buffer	5 µL
dNTPs - 10 mM	200 µM	0.5 µL
Primers - 0.5 µM	0.5 µM	1.25 (each)
template	n/a	0.2
Q5 High-Fidelity DNA Polymerase.	0.02 U/µl	0.25 µL
H ₂ O	n/a	16 µL

The thermocycling program was: 98°C x 30 s; (98°C x 10 s, what temperature was used for annealing 20 s, 72°C x 30s per kb of PCR product) x30 cycles, 72°C x 15 min. Hold at 16°C. The PCR products were purified by gel extraction using QIAquick Gel Extraction Kit.

5.2.6 Restriction digest

Restriction digests of plasmid minipreps were performed as follows. Mixed up to 1 µg plasmid, 1 µL each restriction enzyme (NEB) and 1X CutSmart buffer (NEB) in a final volume of 50 µL. Incubated at 37 °C for at least 1 hr and analysed by agarose gel electrophoresis.

5.2.7 Agarose gel electrophoresis

Restriction digests or PCR reactions were mixed with 1X Gel Loading Dye, Purple and loaded onto an agarose-TAE gel. The percentage of agarose was typically 1 % but was adjusted based on the expected size of the bands to between 0.7% and 2%. The gel was run at 100 V in 1X TAE buffer and imaged using a Gel Doc™ XR+ Gel Documentation System (BioRad).

5.2.8 Gel extraction

PCR reactions or restriction digests were run on agarose-TAE gels and purified using the QIAquick Gel Extraction Kit, according to the manufacturer's instructions. Elution was in 10 µl elution buffer Sanger sequencing. Band specific gel extractions were performed with a transilluminator (Safe Imager™ 2.0 Blue Light Transilluminator)

Plasmids were sequenced by MRC PPU DNA Sequencing and Services (University of Dundee) using the Sanger sequencing method.

5.2.9 Plasmid construction

The pdCas9-T plasmid (addgene plasmid #46569 with transcriptional attenuator engineered in) was constructed by performing a PCR on the original plasmid in which the two primers bind, back to back in the 5' UTR of the dCas9 gene, facing outwards to PCR round the whole plasmid. One of the primers was an ultramer oligo including the transcriptional attenuator and a restriction site (EcoRI) the other primer also introduces the same restriction site.

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

A DpnI digestion was used to remove the plasmid template. An EcoRI digest was used to generate 'sticky ends' before ligation and transformation. After digestion confirmation, NsiI and SmaI were used to digest and ligate the engineered region of the plasmid into the original addgene plasmid #46569 to remove any PCR mutations in the backbone. The pdCas9-T was verified by sequencing. The same procedure was carried out to mutate the RBS site.

Library of crgRNA asRNA parts were synthesised by IDT (K,W,Y variants) or GeneArt (all variants used in Chapter 3:) and were cut out from the supplied donor plasmid using EcoRI and HindIII restriction enzymes and directly ligated into pBR322 acceptor plasmid which carries a different antibiotic resistance. All library members had the inserts sequence verified.

5.3 Data Gathering Methods

5.3.1 End point reading

Each experiment had three repeats conducted on different days and each repeat including multiple replicates, usually 8. When characterising each member of the design libraries, three plasmids were used, a reporter plasmid expressing three different fluorescent proteins pZS2-123, pBR322 expressing the asRNA/crgRNA of the library member and pdCas9 addgene #44249 or the pdCas9-T with an attenuator (Cox et al., 2010; Qi et al., 2013). All three plasmids were co-transformed at the same time. Eight colonies were picked from each transformation and inoculated into 200 μ L of LB broth in each well of a 96 well plate and cultured overnight.

A fluorescence plate was prepared (Greiner-Bio CELLSTAR 96 Well Black μ Clear® # 655090) each well was filled with 340 μ L of LB broth containing the requisite concentrations of antibiotics and inducers. 10 μ L of overnight culture was used to inoculate each culture. Every plate included sterility blanks of non-inoculated media, every plate also included 8 cultures expressing dCas9 and reporter fluorescent protein without the asRNA or crgRNA (positive control). There is also a

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

negative control expressing the dCas9 and the crgRNA/asRNA but without any of the fluorescent reporter genes. These two controls are used to normalise the results. A third control is also included expressing fluorescent reporter proteins, dCas9 and a gRNA to allow a comparison between the level of repression from the 'wt' gRNA and the crgRNA/asRNA system. The 96-well fluorescence plate is cultured at 37°C (Unless otherwise specified) while shaking at 200 rpm. After 24 hours the plate is read with a BMG FLUOstar Omega which reads the optical density (OD₆₀₀) of each culture as well as the level of fluorescence from three channels excitation/emission (355/490, 485/520, 584/620-10 nm). The gain is calculated by adjusting until the positive control is 80% of the saturation level.

5.3.2 Time course

For the time course, the same protocol is used as the endpoint scheme detailed in Section 5.3.1, but rather than shaking in an incubator overnight, the fluorescence plate is placed directly in the plate reader (a BMG FLUOstar Omega) which incubates the plate at 37 °C while double orbital shaking (500 rpm). Reading the same optical density and channels as detailed in Section 5.3.1 every 30 min. This provides data both for observing the system at different growth phases but also for comparing the rate of growth between different cultures. In experiments exploring the effect of temperature or dCas9 dependency, the same protocol as in Section 5.3.1 was carried out at either a different temperature or without the dCas9 expressing plasmid.

5.4 Data Processing

5.4.1 Controlling for row bias

In both the fluorescence plates used for Endpoint analysis and time courses, there is variation in the level of evaporation from wells in the plate depending on their position. Wells at the edge of the plate exhibited an elevated level of evaporation. To control for this, cultures are organised so the 8 cultures from the same transformation fill an entire column and each column represents a single set of plasmids. This leads to each row having the same value, on average, due to

including one of each of the different plasmid sets hence any variation in level of fluorescence between rows can be controlled for by dividing the average level of fluorescence for a row by the average level of fluorescence for the entire plate and using the resulting value as a multiplier for each of the values within a row to control for any row bias.

5.4.2 Propagation of noise

In each of the experiments, there has been no statistically significant difference between the optical densities of each of the cultures there is however, well to well variation which exceeds the level of variation seen in the level of fluorescence. It is standard procedure when working with this kind of fluorescence to control for optical density of a culture. Yet in this case, as there is no statistically significant variation in the level of optical density and the degree of variation in optical density between wells is greater than that of the fluorescence measurements it appears that the plate reader measures fluorescence to a greater accuracy than it does optical density. Therefore, dividing the level of fluorescence by the optical density increases the level of noise rather than reducing error thus unless otherwise specified this step has not been included.

5.4.3 Normalisation

To normalise the level of fluorescence for each culture the first step is to remove auto fluorescence by subtracting the level of fluorescence seen in the negative control from each cell. The next step is to divide the level of fluorescence by the positive control. For both these steps, the average of the 8 cultures for each control is used.

$$\text{Equation 5.1} \quad X_N = \frac{X_O - \frac{\sum_{n=0}^{n=8} e_n}{n}}{\frac{\sum_{n=0}^{n=8} p_n}{n} - \frac{\sum_{n=0}^{n=8} e_n}{n}}$$

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

Equation 5.1 is used to calculate the normalised level of fluorescence where X_N is the normalised level of fluorescence, X_0 is the pre normalisation level of fluorescence, p_n are the positive controls, and e_n are the negative controls. After normalisation, values can be compared between plates from different repeats, Or in a time-series, from an earlier or later time point.

Appendix 1: Toehold orthogonality script

```
>>>Language: Python 3.6

import random

import time

import os

import subprocess

import math

import copy

begin = time.clock()

asRNA = ""

crgRNA = ""

def randomnt():

    for i in range(3):

        f = 1

        s = 1

        while f == s :

            s = f

            f = random.randrange(0,100)

        if f < 25 :
```

```
o = 'A'

elif f < 50 :

    o = 'T'

elif f < 75 :

    o = 'G'

else :

    o = 'C'

return(o)

def sequencegenerator(length= 12, GC = 45 ):

    # GC is the minimum GC content

    #there is a basic complexity screen , no more than three of a neucleotide in a row

    gc = False

    while not gc :

        seq = ""

        for i in range(length):

            seq = seq + randomnt()

        if float(seq.count('G') +seq.count('C'))/float(length) > (float(GC)/100.0):

            gc = True

        if 'AAAA' in seq or 'TTTT' in seq or 'CCCC' in seq or 'GGGG' in seq:

            gc = False
```

```
    return (seq)

def RNAcofold (sequence1 , sequence2, concentration1 = "", concentration2 = ""):

    # note free energy read outs may be inaccurate below -99.99

    keep_on = True

    cycles = 0

    while keep_on :

        try :

            if os.path.isfile('concfile.txt'):

                os.remove('concfile.txt')

            if os.path.isfile('temp_RNAcofold_inputs_24544.txt'):

                os.remove('temp_RNAcofold_inputs_24544.txt')

            if os.path.isfile('temp_RNAcofold_outputs_24544.txt'):

                os.remove('temp_RNAcofold_outputs_24544.txt')

            if "!=" concentration1 and concentration2 != "":

                conc_w = open('concfile.txt','w')

                conc_w.write(str(concentration1)+' ' + str(concentration2) + '\n')

                conc_w.close()

                concentrations = '--concfile=concfile.txt'

            else :

                concentrations = ""
```

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

```
in_w = open('temp_RNAfold_inputs_24544.txt', 'w')

in_w.write(sequence1.replace(' ','') + '&' + sequence2.replace(' ',''))

in_w.close()

in_r=open('temp_RNAfold_inputs_24544.txt', 'r')

out_w = open('temp_RNAfold_ouputs_24544.txt', 'w' )

p = subprocess.Popen(['C:/Program Files (x86)/ViennaRNA
Package/RNAfold.exe', concentrations, '-p', '-a'], stdin = in_r , stdout = out_w
,shell =True )

p.wait()

out_w.close()

in_r.close()

out_r = open('temp_RNAfold_ouputs_24544.txt','r')

output = out_r.read().splitlines()

#print (output)

out_r.close()

#print ('b')

keep_on = False

except PermissionError :

keep_on = True

time.sleep(1)
```

```
cycles = cycles + 1

print ('RNAfold run permission error :' + str(cycles)+ ' attempts.' )

onwards = True

cycles = 0

while onwards :

    try :

        if os.path.isfile('concfile.txt'):

            os.remove('concfile.txt')

        if os.path.isfile('temp_RNAfold_inputs_24544.txt'):

            os.remove('temp_RNAfold_inputs_24544.txt')

        if os.path.isfile('temp_RNAfold_outputs_24544.txt'):

            os.remove('temp_RNAfold_outputs_24544.txt')

        onwards = False

    except PermissionError :

        onwards = True

        time .sleep (1)

        cycles= cycles +1

        print ('RNAfold file wipe permission error ' + str(cycles) + 'cycles')

return (output)

def RNAfold (seq):
```

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

```
a = RNACofold(seq,'ATGC')

return (float(a[6].split('\t')[3]))

def RNACofoldcentroid (seq1,seq2):

    #gives the free energy of the centroid structure formed by two RNA sequences.

    a = RNACofold (seq1,seq2)[1]

    b= float(a[ a.find (' '):].replace('(','').replace(')','').replace(' ',''))

    return (b)

def generatetoehold2(crgRNA,asRNA, length,no_toeholds, no_runs ):

    # establishes the delta gibbs free energy of the crgRNA then adds a

    #random (with screens) toehold and tests the delta gibbs free energy

    #again, most toe holds will complement within the crgRNA but if there

    #is negligible complementation (-0.5) the toehold is returned as one

    #that doesn't do internal folding

    term = 'UUAA'+'AAAAAACCCCGCTTCGGCGGGGTTTT'# yunr + terminator with the

    last two 'T's removed to reflect the probable transcription stop point.

    e = RNAfold(crgRNA)

    asf = RNAfold(asRNA+term)

    i = 0

    accumulator = []

    start = time . clock()
```

```
l = start

for o in range (no_runs):

    i = i +1

    a = sequencegenerator()

    b = a + crgRNA

    r = RNAfold(b)

    ast = asRNA + complement(a) +term

    asr = RNAfold(ast)

    accumulator = accumulator + [[a,r-e + asr-asf, r,e,asr,asf]]

    if time.clock() > l +5 :

        l = time . clock()

        print(str((o/no_runs)*100)[:4] + '% done producing toeholds for system
variant' )

    accumulator = order(accumulator)[:no_toeholds]

    return(accumulator)

def complement(seq):

    seq = seq.upper ()

    out = ""

    for i in seq :

        o = ""
```



```
if i == 'A':  
    o = 'T'  
  
elif i == 'T':  
    o = 'A'  
  
elif i == 'G' :  
    o = 'C'  
  
elif i == 'C':  
    o = 'G'  
  
else:  
    o = '*'  
  
    print ('ERROR : complementing non nucleotide sequence' + i)  
  
out = o + out  
  
return (out)  
  
def toeholdorthogonalitytest (toe1,toe2, crgRNA= crgRNA, asRNA=asRNA,without  
='') :  
  
    a = without  
  
    b = 0  
  
    c = RNAcofoldcentroid (toe2+crgRNA,asRNA +complement(toe1))  
  
    d = RNAcofoldcentroid (toe1+crgRNA,asRNA +complement(toe2))  
  
    e = 0
```

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

```
x = [a,b,c,d,e,]
```

```
y = []
```

```
for i in x :
```

```
    y = y + [i - x[0]]
```

```
z = y[2]+y[3]
```

```
return([z,y])
```

```
def generatingtoeholds(crgRNAs_asRNAs, no_of_toeholds,no_runs = 100):
```

```
    #input is the asRNAs and gRNAs in the format [[name, asrna,crgrna],[name, asrna,  
grna]] etc.
```

```
    term = 'UUAA'+ 'AAAAAACCCCGCTTCGGCGGGGTTTT'
```

```
    a = crgRNAs_asRNAs
```

```
    no = no_of_toeholds
```

```
    crgRNA_out = []
```

```
    asRNA_out = []
```

```
    z = 0
```

```
    start = time.clock()
```

```
    l = start
```

```
    complete = len(a)
```

```
for i in a :

    z = z +1

    toeholds = generatetoehold2(i[2],i[1], 12, no_of_toeholds, no_runs)

    for o in range(len(toeholds)):

        toehold = toeholds[o][0]

        crgRNA_out =crgRNA_out + [[i[0]+' crgRNA '+str(o),toehold+i[2]]]

        asRNA_out =asRNA_out + [[i[0]+' asRNA '+str(o),i[1]+
complement(toehold)+term ]]#here

        if time.clock()>l + 1 :

            l = time.clock()

            print (str((z/complete)*100)[:5]+'% complete. '+str((((1/(z/complete))*(l-
start))-(l-start))/60)[:5] + ' minutes predicted remaining of toehold generation')

        return([asRNA_out,crgRNA_out])

def pairing_probability (asRNA,starta,finisha,crgRNA,startc,finishc, mean = 'arith',
allow_internal = False):

    if not allow_internal:

        startc = startc + len (asRNA)

        finishc = finishc + len (asRNA)

    a = RNAcofold(asRNA,crgRNA)

    a = open ('ABdot5.ps','r')

    b = a .read()
```

```
a.close()

probabilities = b[b.find('%start of base pair probability data')+37:b . find
('showpage\nend\n%%EOF')]

probabilities = probabilities .split('\n')

new = []

for i in probabilities :

    new = new + [i.split(' ')]

probabilities = new

new = []

for i in probabilities :

    z = []

    for o in i[:-1]:

        z = z + [float(o)]

    new = new + [z]

probabilities = new[:-1]

selected = []

for i in probabilities:

    if starta<=i[0]<= finisha and startc<=i[1]<= finishc:

        selected = selected + [i]

if mean == 'geo':
```

```
score = 1

for i in selected:

    score = score * i[2]

score = math.pow(score,1/len(selected))

return (score)

else :

    score = 0

    for i in selected :

        score = score +i[2]

    return (score)

def genmatrix(no):

    #generates square matrix where there isn't assignment cross over

    x = ['']

    y = ['']

    while len (y)<no:

        y = copy.deepcopy(y) + copy.deepcopy(x)

    z = [y]

    while len(z)<no:

        z = copy.deepcopy(z)+ [copy.deepcopy(y)]

    return (z)
```

```
def evaluatecombo1(variants, matrix):  
  
    score = 0  
  
    for i in variants :  
  
        for o in variants :  
  
            if not i==o:  
  
                score = score + float(matrix[i][o])  
  
    if score != 0:  
  
        score = score / ((len(variants)*len(variants))-len(variants))  
  
    return(score)  
  
def cleanrandom():  
  
    s = 0  
  
    t = 0  
  
    while s == t :  
  
        s = random.randrange(0,2)  
  
        t = random.randrange(0,2)  
  
def genrandomcombo(together,no_toeholds,matrix ):  
  
    bottom = 0  
  
    combo = []  
  
    for i in range(len(together)):  
  
        cleanrandom()
```

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

```
    combo = combo + [random.randrange(bottom , bottom+no_toeholds)]

    bottom = bottom + no_toeholds

    return ([combo,evaluatecombo1(combo,matrix)])

def random_optimisation (no_combos, together, no_toeholds, matrix ):

    z = 0

    complete = no_combos

    start = time.clock()

    l = start

    output = []

    outputs = []

    best = 1

    for o in range(no_combos):

        k = genrandomcombo(together,no_toeholds, matrix)

        output = output + [k]

        if k[1]<best:

            best = k[1]

            top = k

            a = open ('best combintation .txt','w')

            a.write(str(time.clock()+ '\n\n'+str(k) )

            a.close()
```

```
z = z + 1

if time.clock()>l + 2 :

    l = time .clock()

    print (str((z/complete)*100)[:5]+'% complete. ' + str((((1/(z/complete))*(l-
start))-(l-start))/60)[:5] + ' minutes predicted remaining of optimisation')

    outputs = outputs + output

    output = []

return(outputs)

def no_changes():

    go = True

    changes = 1

    while go :

        cleanrandom()

        if random .randrange(2)==1:

            go = False

        else:

            changes = changes + 1

    return (changes)

def learning_optimisation (no_combos, together, no_toeholds, matrix,
no_random_combos):
```



```
z = 0

complete = no_random_combos

start = time.clock()

l = start

output = []

outputs = []

best = 1

for o in range(no_random_combos):

    k = genrandomcombo(together,no_toeholds, matrix)

    output = output + [k]

    if k[1]<best:

        best = k[1]

        top = k

        a = open ('best combintation .txt','w')

        a.write(str(time.clock())+ '\n\n'+str(top) )

        a.close()

        a = ""

    z = z + 1

    if time.clock()-l > 2 :

        l = time .clock()
```

```
print (str((z/complete)*100)[:5]+'% complete. ' + str((((1/(z/complete))*(l-  
start))-(l-start))/60)[:5] + ' minutes predicted remaining of random search')
```

```
outputs = outputs + output
```

```
outputs = order (outputs, fast = True)
```

```
output = []
```

```
#
```

```
outputs = outputs + output
```

```
outputs = order (outputs, fast = True)
```

```
for_file = ""
```

```
for i in outputs :for_file = for_file + str(i) +'\n'
```

```
a = open ('combinations from random search .txt','w')
```

```
a.write(for_file)
```

```
a.close()
```

```
print ('\n\ncombination optimisation\n')
```

```
z = 0
```

```
complete = no_combos
```

```
start = time.clock()
```

```
l = start
```

```
parameters = []
```

```
lowerbound = 0
```

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

#parameters are the upper and lower bounds of the regions of the x and y axis
that have each of the crgRNAs on

for i in together :

parameters = parameters+ [[lowerbound,lowerbound + no_toeholds]]

lowerbound = lowerbound + no_toeholds

#gets rid of all but the top 10 outputs so far. one of these will be mutated then
reinserted. the list will then be reordered and the last one removed

outputs = outputs[:10]

for i in range(no_combos):

subject = outputs[random.randrange(0,len(outputs))]

changes = no_changes()

for i in range(changes) :

change = random.randrange(len(parameters))

subject[0][change] =

random.randrange(parameters[change][0],parameters[change][1])

subject = [subject[0] ,evaluatecombo1(subject[0], matrix)]

outputs = [subject] + outputs

outputs = order (outputs)

outputs = outputs [:100]

z = z + 1

if time.clock()>l + 60 :

```
l = time .clock()

print (str((z/complete)*100)[:5]+'% complete. ' + str((((1/(z/complete))*(l-
start))-(l-start))/60)[:5] + ' minutes predicted remaining of combination
optimisation' + str(z) + ' iterations ')

a = open('outputs learning artih mean '+str(z)+' iterations .txt','w')

a.write(str(outputs))

a.close()

return(outputs)

def order(combinations, fast = False ):

    out = [[''],-1000],[[''],1000]]

    start = time.clock()

    l = start

    z = 0

    complete = len (combinations)

    for i in combinations:

        done = False

        for o in range(len(out)):

            if i[1]>=out[o][1] and out[o+1][1]>=i[1] and not done:

                out = out[:o+1]+ [i]+out[o+1:]

                done = True
```

```
if not done :

    print ('failed to order "' + str(i) + "'")

z = z + 1

if l+5 < time.clock():

    l = time.clock()

    print ('ordering '+str((z/complete)*100)[:5]+'% complete')

    if fast :

        out = out[:200]+ [[''],1000]]

out = out[1:201][:-1]

return (out)

def gen3uniform (s, gen3crgRNA, gen3asRNA ,no_toeholds):

    # both gen3crgRNA and gen3asRNA need to be in the form that already has the
    toehold attached

    crgRNAs = s[1][:0-no_toeholds]

    asRNAs = s[0][:0-no_toeholds]

    crgRNAs = crgRNAs + no_toeholds*[['Gen3 crgRNA 0',gen3crgRNA]]

    asRNAs = asRNAs + no_toeholds*[['Gen3 asRNA 0',gen3asRNA]]

    return([asRNAs,crgRNAs])

def readmatrix():

    a = open ('generated interference matrix .txt','r')
```

```
b = a .read ()

a .close()

b = b .replace('[[','').replace(']]','')

lines = b.split ('\n', [])

matrix = []

for i in lines :

    matrix = matrix + [i .split(',')]

return (matrix)

def interactionmatrix(x,y):

    matrix = genmatrix(len(x))

    #for targeting the spacer region

    asRNAstart = 0

    asRNAfinish= 19

    crgRNAstart = 12 + 16

    crgRNAfinish = 12 + 16 + 20

    #for targeting the scaffold region

    asRNAstart = 19

    asRNAfinish= 19 +16

    crgRNAstart = 12

    crgRNAfinish = 12 + 16
```

```
#for capturing the entire interaction (this set used)

asRNAstart = 0

asRNAfinish= 77

crgRNAstart = 0

crgRNAfinish = 155

#

z = 0

start = time.clock()

l = start

complete = len(x)* len(y)

for i in range(len (x)) :

    for o in range(len (y)):

        matrix [i][o]= pairing_probability
(y[o][1],asRNAstart,asRNAfinish,x[i][1],crgRNAstart,crgRNAfinish)

        z = z + 1

        if time.clock()>l + 1 :

            l = time.clock()

            print (str((z/complete)*100)[:5]+'% complete. '+str((((1/(z/complete))*(l-
start))-(l-start))/60)[:5] + ' minutes predicted remaining of pairwise interaction
calculations')

        print ('writing interference matrix to file ')
```

```
a = open('generated interference matrix .csv','w')

a . write (str(matrix).replace('[[','').replace(']]','').replace(', ', ['\n']))#.replace(',','\t')

a.close()

return (matrix)

def read_toeholds():

a = open ('generated toeholds .txt','r')

b = a.read()

a.close()

print ('-----')

print (b.find('['), [''])

print (b [ b.find('['), [']-20: b.find('['), [']+20])

b = b .split('['), [''])

print (len (b))

asRNA = b[0].split("", [''])

crgRNA = b[1].split("", [''])

asRNAs = []

crgRNAs = []

for i in asRNA :

    asRNAs = asRNAs + [i . split("", "")]

for i in crgRNA :
```


Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

```
crgRNAs = crgRNAs + [i . split("", "")]

print (crgRNAs[:3])

return([asRNAs,crgRNAs])

def return_optimal_crgRNA_asRNA_set(combination,crgRNAs, asRNAs): # x =
crgRNAs , Y = asRNAs

    print (combination)

    out = []

    for i in combination[0]:

        out = out + [asRNAs[i],crgRNAs[i],]

    out = [out,combination]

    return (out)

#reading inputs

a = open ('input crgRNA set .txt','r')

together = ast.literal_eval(a.read())

a.close()

#important : number of runs and iterations at each stage

no_toeholds = 50

no_runs = 5000

no_combinations = 1000000

no_random_combos = 1000000
```

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

```
#these are the quick run values, to test changes
```

```
##no_random_combos = 1000
```

```
##no_combinations = 1000
```

```
##no_toeholds = 4
```

```
##no_runs = 6
```

```
##together = together[:4]
```

```
#generating toeholds
```

```
print ('\n\ngenerating toeholds\n')
```

```
s = generatingtoeholds(together, no_toeholds, no_runs = no_runs)
```

```
x = s [1]# crgRNA
```

```
y = s[0] # asRNA
```

```
print ('writing toeholds to file')
```

```
a = open ('generated toeholds .txt', 'w')
```

```
a .write (str(s))
```

```
a.close()
```

```
print ('\n\ncreating interaction matrix\n')
```

```
matrix = interactionmatrix(x,y)
```

```
print ('\n\nrandom optimisation\n')
```

```
combinations = learning_optimisation (no_combinations, together, no_toeholds,  
matrix, no_random_combos)
```

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

```
print ('writing combinations to file')

a = open ('combinations of toeholds .txt', 'w')

a.write(str(combinations))

a.close()

RNASET = return_optimal_crgRNA_asRNA_set(combinations[0],x, y)

RNASETS = str(RNASET[1])+'\n'

for i in RNASET[0]:

    RNASETS =RNASETS + str(i) + '\n'

RNASETS =RNASETS + '\n total run time ' + str((time.clock()-begin)/60) + ' minutes'

a = open ('script result .txt','w')

a .write(RNASETS)

a.close()

print (' total run time ' + str((time.clock()-begin)/60) + ' minutes')
```

Appendix 2: Nucleotide sequences

Sequences from results chapter 1

5' variants

asRNA-J

AGTAACAATTTACACAGACCTGTTTTAGGGAGTTGTGAGG

asRNA-K

AGTAACAATTTACACAGACCTGGAGTTGTGAGG

asRNA-L

AGTAACAAATTCACAGAGACCTGGAGTTGTGAGG

asRNA-M

AGTAACAGTTTCTCACACACCTGGAGTTGTGAGG

asRNA-N

AGTTACAGTTTCTCACACACCTGGAGTTGTGAGG

asRNA-nonsense RNA

CGTTAACATATTCTTACGTATGACGTAGCTATGT

crgRNA-J

cctcacaactccctaaaacAGGTCTGTGTGAAATTGTTActaatcggAACAAATTTACACAGACCTGT
TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGC
ACCGAGTCGGTGCTTTT

crgRNA-K

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

cctcacaactccAGGTCTGTGTGAAATTGTTActaatcggAACAATTTACACAGACCTGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG
AGTCGGTGCTTTT

crgRNA-L

cctcacaactccAGGTCTCTGTGAATTTGTTActaatcggAACAATTTACACAGACCTGTTTTAGA
GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA
GTCGGTGCTTTT

crgRNA-M

cctcacaactccAGGTGTGTGAGAACTGTTActaatcggAACAATTTACACAGACCTGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG
AGTCGGTGCTTTT

crgRNA-N

cctcacaactccAGGTGTGTGAGAACTGTAAActaatcggAACAATTTACACAGACCTGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG
AGTCGGTGCTTTT

gRNA only control

AACAATTTACACAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG
TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT

j23119 (asRNA promoter)

ttgacagctagctcagtcctaggtataatgctagc

pLlacO-1 (crgRNA promoter)

ataaatgtgagcggataacattgacattgtgagcggataacaagatactgagcacg

3' variants

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

asRNA-W

CTCTCACCTTCCTCGTTAACAATTCACACAGACCT

asRNA-X

TAGGAAAGGCGAGCGATAACAGTTTCTCACACACCT

asRNA-Y

GGCGCTGTGCAATTGATTACAGTTTCTCACACACCT

asRNA-Nonsense control

AGGGAACGGGAAGACAGTCAGCGCTGGGACGATCCC

crgRNA-W

AACAATTCACACAGACCTgttttagagctagaaatagcaagttaaataaggctagtcgGAGGGGAGG
AAGGTCTGTGTGAAATTGTTAACGAGGAAGGTGAGAGAGGGGAGGcaaaGCCCCGCCgaaa
GGCGGGCttttttt

crgRNA-X

AACAATTCACACAGACCTgttttagagctagaaatagcaagttaaataaggctagtcgATTGCTTGCT
AGGTGTGTGAGAACTGTTATCGCTGCCTTTCCTACGACATATcaaaGCCCCGCCgaaaGGC
GGGcttttttt

crgRNA-Y

AACAATTCACACAGACCTgttttagagctagaaatagcaagttaaataaggctagtcgAGGGCACGT
TAGGTGTGTGAGAACTGTAATCAATTGCACAGCGCCTCCATCTGcaaaGCCCCGCCgaaaGG
CGGGCttttttt

crgRNA-nonsense 3' element

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

AACAATTTACACAGACCTgttttagagctagaaatagcaagttaaataaggctagtcgTCTGTCTTAT
TGGGATCGTCCCAGCGCTGACTGTCTTCCCGTTCCTCTTCCCAcaaaGCCCGCCgaaaGGCG
GGCttttttt

j23107 (asRNA promoter)

tttacggctagctcagccctaggtattatgctagc

PLlacO-1 (crgRNA promoter)

ataaatgtgagcggataacattgacattgtgagcggataacaagatactgagcacg

Sequences from results chapter 2

Variant library and toehold change

control nonsense asRNA

GCGTTAACATATTCTTACGTATGACGTAGCTATGT

asRNA-Sp

AACAATTTACACAGACCTACTGTGAATGCC

asRNA-LS

AACAATTTACACAGACCTGTTTACTGTGAATGCC

asRNA-Bg

AACAATTTACACAGACCTGTTTTAGAACTGTGAATGCC

asRNA-US

AACAATTTACACAGACCTGTTTTAGAGCTAACTGTGAATGCC

asRNA-Lp(α)

AACAATTTACACAGACCTGTTTTAGAGCTAGAAAAGTGTGAATGCC

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

asRNA-Lp β

AACAATTTACACAGACCTGTTTTAGAGCTAGAAAGTCTGGGTATC

asRNA-Lpy

AACAATTTACACAGACCTGTTTTAGAGCTAGAAAATTCGACGAGGA

asRNA only

AACAATTTACACAGACCTGTTTTAGAGCTAGAAAAGTGAATGCC

crgRNA-Sp

GGCATTACAGTAGGTCTGTGTGAAATTGTTACTAGTAACAATTTACACAGACCTGTTTTA
GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
GAGTCGGTGCTTTTTTTT

crgRNA-LS

GGCATTACAGTAAACAGGTCTGTGTGAAATTGTTACTAGTAACAATTTACACAGACCTG
TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG
CACCGAGTCGGTGCTTTTTTTT

crgRNA-Bg

GGCATTACAGTTCTAAACAGGTCTGTGTGAAATTGTTACTAGTAACAATTTACACAGA
CCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
GTGGCACCGAGTCGGTGCTTTTTTTT

crgRNA-US

GGCATTACAGTTAGCTCTAAACAGGTCTGTGTGAAATTGTTACTAGTAACAATTTACAC
AGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA
AAAGTGGCACCGAGTCGGTGCTTTTTTTT

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

crgRNA-Lp(α)

GGCATTACAGTTTTCTAGCTCTAAAACAGGTCTGTGTGAAATTGTTACTAGTAACAATTC
ACACAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT
TGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

crgRNA-Lp β

GATACCCAAGACTTTCTAGCTCTAAAACAGGTCTGTGTGAAATTGTTACTAGTAACAATTC
ACACAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT
TGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

crgRNA-Lp γ

TCCTCGTCGAATTTCTAGCTCTAAAACAGGTCTGTGTGAAATTGTTACTAGTAACAATTC
ACACAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT
TGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

gRNA only control

ACTAGTAACAATTTACACAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT
AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

j23119 (asRNA promoter)

ttgacagctagctcagtcctaggtataatgctagc

PLlacO-1 (crgRNA promoter)

ataaatgtgagcggataacattgacattgtgagcggataacaagatactgagcacg

Orthogonality

asRNA-A

TGATAGATTCAATTGTGAGGTTTTAGAGCTAGAAATAGCGATGGACC

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

crgRNA-A

GGTCCATCGCTATTTCTAGCTCTAAAACCTCACAATTGAATCTATCATCTAGATGATAGATT
CAATTGTGAGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT
TGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT

asRNA-B

CACCTGCCATGGTTTCCAAGTTTTAGAGCTAGAAACCATGGGCCTCC

crgRNA-B

GGAGGCCCATGGTTTCTAGCTCTAAAACCTGGAAACCATGGCAGGTGCCTAGGCACCTGCC
ATGTTTTCCAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
TTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT']

asRNA-C

GTCACGAGTTCGAGATCGAGTTTTAGAGCTAGAAAGCAGACTAGCTC

crgRNA-C

GAGCTAGTCTGCTTTCTAGCTCTAAAACCTCGATCTCGAACTCGTGACCCTAGGGTCACGAG
TTCGAGATCGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
TTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT

Multiple targets

asRNA-CFP

AGATACTGAGCACATCAGCGTTTTAGAGCTAGAAACAGCCAGTGCCA

crgRNA-CFP

TGGCACTGGCTGTTTCTAGCTCTAAAACGCTGATGTGCTCAGTATCTCCTAGGAGATACTG
AGCACATCAGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
TTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

asRNA-YFP

AGATACTGAGCACATCAGCGTTTTAGAGCTAGAAAGCCAGATCCGTC

crgRNA-YFP

GACGGATCTGGCTTTCTAGCTCTAAAACGCTGATGTGCTCAGTATCTTCTAGAAGATACTG
AGCACATCAGCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
TTGAAAAGTGGCACCGAGTCGGTGCTTTTTTT

asRNA-mCherry

AACAATTCACACAGACCTGTTTTAGAGCTAGAAAAGTGAATGCC

crgRNA-mCherry

GGCATTACAGTTTTCTAGCTCTAAAACAGGTCTGTGTGAAATTGTTACTAGTAACAATTC
ACACAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT
TGAAAAGTGGCACCGAGTCGGTGCTTTTTTT

j23119 (asRNA promoter)

ttgacagctagctcagtcctaggtataatgctagc

PLlacO-1 (crgRNA promoter)

ataaatgtgagcggataacattgacattgtgagcggataacaagatactgagcacg

dCas9 expression

dCas9 coding sequence used:

atggataagaaataactcaataggcttagctatcggcaciaaatagcgtcggatgggaggatgactgatgaatataag
gttccgtctaaaaagttcaaggttctgggaaatacagaccgccacagatcaaaaaaatcttataggggctcttttatt
tgacagtggagagacagcgggaagcgactcgtctcaaacggacagctcgtagaaggatatacacgtcggagaatcgta
ttgttatctacaggagatTTTTTcaaatgagatggcgaaagtagatgatagttctttcatcgacttgaagagtctttttg
gtggaagaagacaagaagcatgaacgtcatcctatttttgaaatatagtagatgaagttgcttatcatgagaaatc

caactatctatcatctgcgaaaaaattggttagattctactgataaagcggatttgcgcttaatctatttggccttagcgc
atatgattaagtttcgtggtcatttttgattgaggagatttaaatcctgataatagtgatgtggacaaactatttatcca
gttggtacaaacctacaatcaattttgaagaaaacctattaacgcaagtggagtagatgctaaagcgattctttctg
cacgattgagtaaatcaagacgattagaaaatctcattgctcagctccccggtgagaagaaaaatggcttatttgggaa
tctcattgctttgtcattgggtttgaccctaattttaaatcaaattttgatttggcagaagatgctaaattacagctttcaa
aagatacttacgatgatgatttagataatttattggcgcaaattggagatcaatatgctgatttgttttggcagctaaga
atttatcagatgctattttactttcagatatcctaagagtaataactgaaataactaaggctcccctatcagcttcaatga
ttaaacgctacgatgaacatcatcaagacttgactcttttaaagcttttagttcgacaacaactccagaaaagtataaa
gaaatctttttgatcaatcaaaaaacggatatgcaggttatattgatgggggagctagccaagaagaattttataaatt
tatcaaaccaattttagaaaaaatggatggtagggaattattggtgaaactaaatcgtaagatttgcctgcgcaag
caacggacctttgacaacggctctattccccatcaaattcacttgggtgagctgcatgctattttgagaagacaagaaga
ctttatccatttttaaagacaatcgtagagaagattgaaaaatcttgactttcgaattccttattatgttggctcattgg
cgctggcaatagtcgtttgcatggatgactcggaagtctgaagaacaattacccatggaattttgaagaagttgtc
gataaaggcttcagctcaatcatttattgaacgcatgacaaactttgataaaaatcttccaaatgaaaaagtactacc
aaaacatagtttgctttatgagtattttacggtttataacgaattgacaaggctcaaatatgttactgaaggaatgcgaa
aaccagcatttcttcaggtgaacagaagaagccattgttgattactcttcaaaacaatcgaaaagtaaccgttaa
gcaatataaagaagattttcaaaaaatagaatgttttgatagttgaaatttcaggagttgaagatagatttaatg
cttcattaggtacctaccatgatttgctaaaaattattaagataaagatttttggataatgaagaaaatgaagatatct
tagaggatattgttttaacattgaccttatttgaagataggagatgattgaggaaagacttaaacatatgctcacctc
ttgatgataaggatgaaacagcttaaactgcgcttatactgggtggggacgtttgtctgaaaattgattaatggt
attagggataagcaatctggcaaaacaatattagatttttgaaatcagatggttttccaatcgcaattttatgcagctg
atccatgatgatgttgacatttaagaagacattcaaaaagcacaagtgctggacaaggcgatagtttatcatgaac
atattgcaatttagctggtagcctgctattaaaaaggattttacagactgtaaaagttgttgatgaattggtcaaa
gtaatggggcggcataagccagaaaatatcgttattgaaatggcacgtgaaaatcagacaactcaaaaggccagaa
aaattcgcgagagcgtatgaaacgaatcgaagaaggtatcaagaattaggaagtcagattcttaagagcatcctgt
tgaataactcaattgcaaaatgaaaagctctatcttattatctcaaaatggaagagacatgtatgtggaccaagaa
ttagatattaatcgtttaagtgattatgatgctgatgccattgtccacaaagtttccttaagacgattcaatagacaat
aaggtcttaacgcttctgataaaaatcgtgtaaatcggataacgttccaagtgaagaagtagtcaaaaagatgaaa
aactattggagacaacttctaaacccaagttaatcactcaacgtaagtttgataatttaacgaaagctgaacgtggag
gtttgagtgaacttgataaagctggttttatcaaacgccaattggttgaaactcgccaaatcactaagcatgtggcacia

Synthetic Logic Circuits encoded on Toehold Strand-Displacement Switchable CRISPR guide RNAs.

atTTGGatagTcgcatgaatacTaaatacGatgaaaatgataaacttattcgagaggTtaaagtgattacctTaaaatc
taaattagtttctgacttccgaaaagatttccaattctataaagtacgtgagattaacaattaccatcatgcccatgatgc
gtatcTaaatgccgtcgttggaaactgctttgattaagaaatccaaaactgaaatcggagtttgtctatggtgattataa
agtttatgatgttcgTaaatgattgctaagtctgagcaagaaataggcaaagcaaccgcaaaatatttctttactcta
atatcatgaaacttctcaaaacagaaattacacttgcaaatggagagattcgcaaacgccctctaatacgaaactaatgg
ggaaactggagaaattgtctgggataaagggcgagatttggcacagtgcgcaaagtattgtccatgccccagtcaat
attgtcaagaaaacagaagtacagacaggcggattctccaaggagtcaattttacaaaaagaaattcggacaagctt
attgctcgtaaaaagactgggatccaaaaaatatggtggttttagatgccaacggtagcttattcagtcctagtggT
tgctaaggtggaaaaagggaaatcgaagaagtTaaatccgtTaaagagttactagggatcacaattatggaaagaa
gttcctttgaaaaaatccgattgactttttagaagctaaaggatataaggaagtTaaaaagacttaatacTaaact
acTaaatatagtctttttagttagaaaacggctgTaaacggatgctggctagtgcggagaattacaaaaaggaaat
gagctggctctgccaagcaaatatgtgaatttttatatttagctagtcattatgaaaagttgaagggtagtcagaaga
taacgaacaaaaacaattgtttgtggagcagcataagcattatttagatgagattattgagcaaatcagTgaattttcta
agcgtgttatttttagcagatgccaatttagataaagtcttagtgcatataacaacatagagacaaaccaatacgtga
acaagcagaaaaatattattcatttattacgttgacgaatctggagctcccgctgctttTaaatattttgatacaacaatt
gatcgtaaacgatatacgtctacaaaagaagttttagatgccactcttatccatcaatccatcactggctttatgaaaca
cgattgatttgagtcagctaggaggtgactaa

pLtetO-1 (promoter used to express dCas9)

gttgacactctatcgttgatagagttattttaccactccctatcagtgatagagaa

References

- Atkinson, M. R., Savageau, M. A., Myers, J. T., & Ninfa, A. J. (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, *113*(5), 597–607.
- Bernhart, S. H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2006). Partition function and base pairing probabilities of RNA heterodimers. *Algorithms for Molecular Biology*, *1*(1), 3. <https://doi.org/10.1186/1748-7188-1-3>
- Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., & Marraffini, L. a. (2013a). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*, *41*(15), 7429–7437. <https://doi.org/10.1093/nar/gkt520>
- Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., & Marraffini, L. a. (2013b). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*, *41*(15), 7429–7437. <https://doi.org/10.1093/nar/gkt520>
- Boole, G. (1854). *The Laws of Thought*. *Buffalo: Prometheus Books*.
- Bothfeld, W., Kapov, G., & Tyo, K. E. J. (2017). A Glucose-Sensing Toggle Switch for Autonomous, High Productivity Genetic Control. *ACS Synthetic Biology*, *6*(7), 1296–1304. <https://doi.org/10.1021/acssynbio.6b00257>
- Bowsher, C. G., Voliotis, M., & Swain, P. S. (2013). The fidelity of dynamic signaling by noisy biomolecular networks. *PLoS Computational Biology*, *9*(3), e1002965. <https://doi.org/10.1371/journal.pcbi.1002965>
- Bradley, R. W., Buck, M., & Wang, B. (2016). Recognizing and engineering digital-like logic gates and switches in gene regulatory networks. *Current Opinion in Microbiology*, *33*, 74–82. <https://doi.org/10.1016/J.MIB.2016.07.004>
- Briner, A. E., Donohue, P. D., Goma, A. A., Selle, K., Slorach, E. M., Nye, C. H.,

... Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. *Molecular Cell*, 56(2), 333–339.
<https://doi.org/10.1016/j.molcel.2014.09.019>

Chappell, J., Takahashi, M. K., & Lucks, J. B. (2015). Creating small transcription activating RNAs. *Nature Chemical Biology*, 1–9.
<https://doi.org/10.1038/nchembio.1737>

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), 819–823. <https://doi.org/10.1126/science.1231143>

Cox, R. S., Dunlop, M. J., & Elowitz, M. B. (2010). A synthetic three-color scaffold for monitoring genetic regulation and noise. *Journal of Biological Engineering*, 4(1), 10. <https://doi.org/10.1186/1754-1611-4-10>

Cress, B. F., Jones, J. A., Kim, D. C., Leitz, Q. D., Englaender, J. A., Collins, S. M., ... Koffas, M. A. G. (2016). Rapid generation of CRISPR/dCas9-regulated, orthogonally repressible hybrid T7-lac promoters for modular, tuneable control of metabolic pathway fluxes in *Escherichia coli*. *Nucleic Acids Research*, 44(9), 4472–4485. <https://doi.org/10.1093/nar/gkw231>

Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767), 335–338.
<https://doi.org/10.1038/35002125>

Espah Borujeni, A., Channarasappa, A. S., & Salis, H. M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, 42(4), 2646–2659. <https://doi.org/10.1093/nar/gkt1139>

Gaber, R., Lebar, T., Majerle, A., Šter, B., Dobnikar, A., Benčina, M., ... Šter, B. (2014). Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nature Chemical Biology*, 10(january), 203–208.
<https://doi.org/10.1038/nchembio.1433>

- Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, *403*, 339–342.
- Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., ... Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, *159*(3), 647–661.
<https://doi.org/10.1016/j.cell.2014.09.029>
- Green, A. A., Kim, J., Ma, D., Silver, P. A., Collins, J. J., & Yin, P. (2017). Complex cellular logic computation using ribocomputing devices. *Nature Publishing Group*, *548*(7665), 117–121. <https://doi.org/10.1038/nature23271>
- Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014a). Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell*, *159*(4), 925–939.
<https://doi.org/10.1016/j.cell.2014.10.002>
- Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014b). Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell*, *159*(4), 925–939.
<https://doi.org/10.1016/j.cell.2014.10.002>
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R., & Hofacker, I. L. (2008). The Vienna RNA Websuite. *Nucleic Acids Research*, *36*(Web Server), W70–W74. <https://doi.org/10.1093/nar/gkn188>
- Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research*, *31*(13), 3429–3431. <https://doi.org/10.1093/nar/gkg599>
- Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte Für Chemie Chemical Monthly*, *125*(2), 167–188.
<https://doi.org/10.1007/BF00818163>
- Huang, H. (2007). Design and characterization of artificial transcriptional terminators. Retrieved from <https://dspace.mit.edu/handle/1721.1/45981>
- Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., & Collins, J.

J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7), 841–847.
<https://doi.org/10.1038/nbt986>

Jansen, R., Embden, J. D. A. van, Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), 1565–1575.

Jiang, F., Zhou, K., Ma, L., Gressel, S., & Doudna, J. A. (2015). STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science (New York, N.Y.)*, 348(6242), 1477–1481.
<https://doi.org/10.1126/science.aab1452>

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., ... Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (New York, N.Y.)*, 343(6176), 1247997.
<https://doi.org/10.1126/science.1247997>

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., ... Doudna, J. A. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science*, 343(6176), 1247997–1247997.
<https://doi.org/10.1126/science.1247997>

Kerpedjiev, P., Hammer, S., & Hofacker, I. L. (2018). forna (force-directed RNA): Simple and Effective Online RNA Secondary Structure Diagrams. *BIOINFORMATICS*, 00(00), 1–2. Retrieved from <http://>

Knight, S. C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L. B., Bosanac, L., ... Tjian, R. (2015). Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*, 350(6262), 823–826. <https://doi.org/10.1126/science.aac6572>

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., ... Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517(7536), 583–588.
<https://doi.org/10.1038/nature14136>

- Kortmann, J., & Narberhaus, F. (2012). Bacterial RNA thermometers: molecular zippers and switches. *Nature Reviews Microbiology*, *10*(4), 255–265. <https://doi.org/10.1038/nrmicro2730>
- Lee, Y. J., Hoynes-O'Connor, A., Leong, M. C., & Moon, T. S. (2016a). Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Research*, *44*(5), 2462–2473. <https://doi.org/10.1093/nar/gkw056>
- Lee, Y. J., Hoynes-O'Connor, A., Leong, M. C., & Moon, T. S. (2016b). Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Research*, gkw056. <https://doi.org/10.1093/nar/gkw056>
- Lee, Y. J., & Moon, T. S. (2018). Design rules of synthetic non-coding RNAs in bacteria. *Methods*, *143*, 58–69. <https://doi.org/10.1016/J.YMETH.2018.01.001>
- Lehman, E., Leighton, F. T., & Meyer, A. R. (2017). *Mathematics for computer science*. Retrieved from https://books.google.co.uk/books/about/Mathematics_for_Computer_Science.html?id=vpbGAQAACAAJ&redir_esc=y
- Liang, C., Li, F., Wang, L., Zhang, Z.-K., Wang, C., He, B., ... Zhang, G. (2017). Tumor cell-targeted delivery of CRISPR/Cas9 by aptamer-functionalized lipopolymer for therapeutic genome editing of VEGFA in osteosarcoma. *Biomaterials*, *147*, 68–85. <https://doi.org/10.1016/J.BIOMATERIALS.2017.09.015>
- Liu, Y., Zhan, Y., Chen, Z., He, A., Li, J., Wu, H., ... Cai, Z. (2016). Directing cellular information flow via CRISPR signal conductors. *Nature Methods*, *13*(11), 938–944. <https://doi.org/10.1038/nmeth.3994>
- Lorenz, R., Bernhart, S. H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology : AMB*, *6*, 26. <https://doi.org/10.1186/1748-7188-6-26>

Lovelace, Ada; Babbage, C. (1842). Sketch of The Analytical Engine Invented by Charles Babbage. *The Military Engineers*.

Ma, H., Tu, L.-C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., & Pederson, T. (2016). Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature Biotechnology*, *34*(5), 528–530. <https://doi.org/10.1038/nbt.3526>

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science*, *339*(6121), 823 LP-826. Retrieved from <http://science.sciencemag.org/content/339/6121/823.abstract>

Mathews, D. H. (2006). Revolutions in RNA Secondary Structure Prediction. *Journal of Molecular Biology*, *359*(3), 526–532. <https://doi.org/10.1016/J.JMB.2006.01.067>

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., & Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7287–7292. <https://doi.org/10.1073/pnas.0401799101>

McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, *29*(6–7), 1105–1119. <https://doi.org/10.1002/bip.360290621>

Meyer, S., Chappell, J., Sankar, S., Chew, R., & Lucks, J. B. (2015). Improving fold activation of small transcription activating RNAs (STARs) with rational RNA engineering strategies. *Biotechnology and Bioengineering*, n/a-n/a. <https://doi.org/10.1002/bit.25693>

Mojica, F. J., Díez-Villaseñor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*, *36*(1), 244–246.

- Mojica, F. J., Ferrer, C., Juez, G., & Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Molecular Microbiology*, *17*(1), 85–93.
- Morita, T., & Aiba, H. (2011). RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes & Development*, *25*(4), 294–298. <https://doi.org/10.1101/gad.2030311>
- Nielsen, A. A., & Voigt, C. A. (2014). Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular Systems Biology*, *10*(11), 763–763. <https://doi.org/10.15252/msb.20145735>
- Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., ... Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), 935–949. <https://doi.org/10.1016/j.cell.2014.02.001>
- O'Brien, F. (2010). *The Apollo Guidance Computer : architecture and operation*. Praxis. Retrieved from <https://books.google.co.uk/books?id=3fKzL0HfJp4C&printsec=frontcover&dq=156347185X&hl=en&sa=X&ved=0ahUKEwj6koqCuvfeAhVJWsAKHQnLDsQ6AEILzAB#v=onepage&q&f=false>
- Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, *31*(9), 839–843. <https://doi.org/10.1038/nbt.2673>
- Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kadi, A. M., Polstein, L. R., ... Gersbach, C. A. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods*, *10*(10), 973–976. <https://doi.org/10.1038/nmeth.2600>
- Prindle, A., Selimkhanov, J., Li, H., Razinkov, I., Tsimring, L. S., & Hasty, J. (2014). Rapid and tunable post-translational coupling of genetic circuits.

Nature, 508(7496), 387–391. <https://doi.org/10.1038/nature13238>

Qi, L. S., Larson, M. H., Gilbert, L. a, Doudna, J. a, Weissman, J. S., Arkin, A. P., & Lim, W. a. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), 1173–1183.

<https://doi.org/10.1016/j.cell.2013.02.022>

R. Stanković, J. A. (2008). Reprints from the Early Days of Information Sciences: TICSP Series On the Contributions of Akira Nakashima to Switching Theory. *Tampere International Center for Signal Processing*.

Reeder, J., & Giegerich, R. (2004). Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics*, 5(1), 104. <https://doi.org/10.1186/1471-2105-5-104>

Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L., & Corn, J. E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature Biotechnology*, 34(3), 339–344. <https://doi.org/10.1038/nbt.3481>

Rivas, E., & Eddy, S. R. (1999). A dynamic programming algorithm for RNA structure prediction including pseudoknots 1 Edited by I. Tinoco. *Journal of Molecular Biology*, 285(5), 2053–2068. <https://doi.org/10.1006/jmbi.1998.2436>

Rivas, E., Klein, R. J., Jones, T. A., & Eddy, S. R. (2001). Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Current Biology*, 11(17), 1369–1373. [https://doi.org/10.1016/S0960-9822\(01\)00401-8](https://doi.org/10.1016/S0960-9822(01)00401-8)

Rosado, A., Cordero, T., & Rodrigo, G. (2018). Binary addition in a living cell based on riboregulation. *PLOS Genetics*, 14(7), e1007548. <https://doi.org/10.1371/journal.pgen.1007548>

Roybal, K. T., Rupp, L. J., Morsut, L., Walker, W. J., McNally, K. A., Park, J. S., & Lim, W. A. (2016). Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits. *Cell*, 164(4), 770–779.

<https://doi.org/10.1016/J.CELL.2016.01.011>

Sakai, Y., Abe, K., Nakashima, S., Yoshida, W., Ferri, S., Sode, K., & Ikebukuro, K. (2013). Improving the Gene-Regulation Ability of Small RNAs by Scaffold Engineering in *Escherichia coli*. *ACS Synthetic Biology*.

<https://doi.org/10.1021/sb4000959>

Schaerli, Y., Gili, M., & Isalan, M. (2014). A split intein T7 RNA polymerase for transcriptional AND-logic. *Nucleic Acids Research*, *42*(19), 12322–12328.

<https://doi.org/10.1093/nar/gku884>

Shannon, C. E. (1938). A symbolic analysis of relay and switching circuits.

Electrical Engineering, *57*(12), 713–723.

<https://doi.org/10.1109/EE.1938.6431064>

Stanton, B. C., Nielsen, A. a K., Tamsir, A., Clancy, K., Peterson, T., & Voigt, C. a. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates.

Nature Chemical Biology, *10*(2), 99–105.

<https://doi.org/10.1038/nchembio.1411>

Sternberg, S. H., LaFrance, B., Kaplan, M., & Doudna, J. A. (2015). Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature*, *527*(7576), 110–113.

<https://doi.org/10.1038/nature15544>

Tanenbaum, A. S., & Goodman, J. R. (2005). *Structured computer organization*.

Prentice Hall. Retrieved from <https://dl.acm.org/citation.cfm?id=552473>

Tang, W., Hu, J. H., & Liu, D. R. (2017). Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation.

Tang, W., Hu, J. H., & Liu, D. R. (2017). Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nature Commu.*

Nature Communications, *8*(May), 15939. <https://doi.org/10.1038/ncomms15939>

Vigouroux, A., Oldewurtel, E., Cui, L., Bikard, D., & van Teeffelen, S. (2018).

Tuning dCas9's ability to block transcription enables robust, noiseless

knockdown of bacterial genes. *Molecular Systems Biology*, 14(3), e7899.

<https://doi.org/10.15252/MSB.20177899>

Wang, S., Su, J.-H., Zhang, F., & Zhuang, X. (2016). An RNA-aptamer-based two-color CRISPR labeling system. *Scientific Reports*, 6(1), 26857.

<https://doi.org/10.1038/srep26857>

Wong, A., Wang, H., Poh, C. L., & Kitney, R. I. (2015). Layering genetic circuits to build a single cell, bacterial half adder. *BMC Biology*, 13(1), 40.

<https://doi.org/10.1186/s12915-015-0146-0>

Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., ... Pierce, N. A. (2011). NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1), 170–173.

<https://doi.org/10.1002/jcc.21596>

Zhang, Y., Yin, C., Zhang, T., Li, F., Yang, W., Kaminski, R., ... Hu, W. (2015). CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs.

Scientific Reports, 5(1), 16277. <https://doi.org/10.1038/srep16277>

Zhen, S., Takahashi, Y., Narita, S., Yang, Y.-C., & Li, X. (2017). Targeted delivery of CRISPR/Cas9 to prostate cancer by modified gRNA using a flexible aptamer-cationic liposome. *Oncotarget*, 8(6), 9375–9387.

<https://doi.org/10.18632/oncotarget.14072>

Zuker, M., & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1), 133–148.