A Kunkel-inspired Approach to Create Light-activated Plasmids



Khoa Chung

Linacre College
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

Abstract

Plasmids are circular double-stranded DNA and have been the workhorse of molecular biology. They can also be adopted by more complex organisms to influence how their cells behave. Importantly, this means they can be leveraged to study intricate molecular networks, and gene-function relationships, and hold promise as therapeutic devices. Thus, the ability to deliver and control their activity will not only advance their utility in research fields such as developmental and molecular biology but also expedite their translation to clinical settings. In this regard, light is a convenient stimulus as it is non-invasive, and its properties and application can be adjusted to minimise cellular damage. However, examples of light-controlled plasmid are limited, with the majority displaying poor light-dependent activation, or necessitating specific sequences that limit their use.

This thesis outlines the development of a general strategy to deliver light-activated plasmids. This approach relied on photoblocking groups placed at the promoter, which could be removed in response to UV light. This strategy was used to control two ubiquitous bacterial and mammalian promoters. Cell-free reactions with the T7 promoter could be tightly repressed until the UV stimulus was applied. Likewise, light-regulated transcription and translation from the CMV promoter were also demonstrated in mammalian cells, although the photomodulation efficiency was more

modest. However, this strategy failed to work on the tac and Sp6 promoters, which is likely due to the difference in binding interactions between the promoter and their cognate RNA polymerase partner. Thus, although this strategy worked well for the T7 promoter in cell-free reactions, the promoter dependency would mean that optimisation is required when the strategy is adopted to photocage alternative promoters.

Furthermore, an outstanding challenge needs to be overcome before plasmids can be used for therapy: the inefficient delivery of large negatively charged molecules into cells. Initial studies in this regard were made, and suggestions to enable plasmid delivery were outlined. Unlike previous examples, this technology has the potential to be sequence agnostic, meaning it could be adaptable to any plasmid, greatly expanding its scope.

Acknowledgments

I would like to thank everyone who has helped me throughout my DPhil and made the journey so enjoyable. I would like to thank the Synthetic Biology CDT, the EPSRC, and the CRUK for funding my research. To Dr. Michael Booth, for all the advice and interesting scientific discussions, and for allowing me the freedom to pursue my research. To Professor Hagan Bayley, for making me feel welcome and a part of your group. To Darrell Sharp, for sparking my scientific interest in my formative years. To Dr. Ian George and Dr. Jakob Franke for giving me my foundational research skills. Oana Pelea for teaching me how to hold a pipette. To my friends, Muna Fuyal, Phitawat Namnouad, Thomas Caganek, George Papadakis, Nicolas Busatto, and Celia Toyos-Rodríguez who made the experience so special, and for being there to share in the good, and importantly the not-so-good moments. To Sonia Yang, Jorin Riexinger, Zhong Lim, Jefferson Smith, and Xingzao Wang, who inspired me through their sheer hard work, tenacity, and love for research, and who motivated me to go back to the lab whenever experiments did not go as planned. To Ivan Cheung for your unquestioning faith in me, and for being the rock that I'm not. To the Bassett-Jones, for taking me in when I first came to England. To my darling niece, baby Victoire, whose infectious smile never failed to cheer me up. To my Mum and sister, for your unconditional love and support in all I do. And last, to Dad, for all the spoken and unspoken words – I hope I've done you proud.

Abbreviations

5′-aa-dUTP = 5′-aminoallyl-dUTP

5′-ap-dUTP = 5′-aminopropargyl-dUTP

5mC = 5-methylcytidine

5mU = 5-methyluridine

^{8PV}G = 8-pyrenylvinyl deoxyguanosine

AAV = Adeno-associated viruses

AD = Activation domain

BCCP = Biotin carboxy carrier protein

BRE^{d} and BRE^{u} = B recognition element (downstream and upstream, respectively)

Bt = Biotin

BSA = Bovine serum albumin

cDNA = Complementary DNA

CFPS = Cell-free protein synthesis

CHSA = Cationic human serum albumin

CLI = Critical limb ischemia

CMV = Cytomegalovirus

COVID-19 = Coronavirus disease 2019

CRISPR = Clustered regularly interspaced short palindromic repeats

CTL = Cytotoxic T lymphocytes

DBD = DNA-binding domain

ddNTP = Dideoxyribonucleoside triphosphate

dGTP = Deoxyguanosine triphosphate

DMNPE = 1-(4,5-dimethoxy-2-nitrophenyl) diazoethane

DNA = Deoxyribonucleic acid

DNAP = DNA polymerase

dNTP = Deoxyribonucleoside triphosphate

Dox = doxycycline

dsDNA = Double-stranded DNA

dTTP = Deoxythymidine triphosphate

dUTP = Deoxyuridine triphosphate

EDTA = Ethylenediaminetetraacetic acid

ELISA = Enzyme-linked immunosorbent assay

FDA = Food and Drug Administration

GFP = Green fluorescent protein

GOI = Gene-of-interest

hATTR = Hereditary disease transthyretin-mediated amyloidosis

HEK293T = Human embryonic kidney 293T cells

HGF = Hepatocyte growth factor

HPLC = High-performance liquid chromatography

HSV-TK = Herpes simplex virus thymidine kinase

Inr = Initiator element

IPTG = Isopropyl ß-D-1-thiogalactopyranoside

IVT = *In vitro* transcription

IVTT = *In vitro* transcription/translation

LA = Light-activated

LC-MS = Liquid chromatography-mass spectrometry

LNP = Lipid nanoparticle

LOV = Light-oxygen-voltage

m6A = N6-methyladenosine

mC = M(onomeric) Cherry

MM = Mastermix

mNG = M(onomeric) Neongreen

mRNA = Messenger RNA

MRS = Multimer resolution site

mV = M(onomeric) Venus

MVS = Monovalent streptavidin

NEB = New England Biolabs

NEEL = Nick-Exonuclease-Extension-Ligation

NHS ester = N-Hydroxysuccinimide ester

NIR = Near-infrared

NLS = Nuclear localisation signal

NTP = Nucleoside triphosphate

Ori = Origin of replication

PAD = Peripheral arterial disease

PAGE = Polyacrylamide gel electrophoresis

PAMP = Pathogen-associated molecular patterns

PCR = Polymerase chain reaction

PEG = Poly(ethylene) glycol

PhyB = Phytochrome protein B

PIF3 = Phytochrome-interacting factor 3

PRR = Pattern-recognition receptors

PURE = Protein synthesis using recombinant elements

RES = Reticuloendothelial system

RNA = Ribonucleic acid

RNAP = RNA polymerase

RT-qPCR = Reverse transcription quantitative real-time PCR

scFv = Single chain variable fragment

SDM = Site-directed mutagenesis

sgRNA = Single guide RNA

shRNA = Short hairpin RNA

siRNA = Small interfering RNA

SV40 = Simian virus 40

ssDNA = Single-stranded DNA

TAE = Tris-acetate-EDTA

TBP = TATA-box binding protein

Tc = Tetracycline

TCS = Two-component system

TetO = Tet operator sequence

Tf = Transferrin

TF = Transcription factor

TfR = Transferrin receptor

TRE = Tetracycline response element

TSS = Transcription start site

tTA = Tetracycline transactivator

TTR = Transthyretin

TxRed = Texas Red

UCN = Upconversion nanoparticle

UTR = Untranslated region

UV = Ultraviolet

VEGF = Vascular endothelial growth factor

V_H = Variable light chain

V_L = Variable heavy chain

WCF = Watson-Crick-Franklin

WNV = West Nile virus

WTS = Wildtype tetravalent streptavidin

Contents

1. Introduction _____ 1

1.1 Plasmid Overview 1

1.2 Discoveries that Led to Recombinant Plasmids 2

1.3 Applications of Recombinant Plasmids 5

1.3.1 Recombinant Plasmids in Foods and Crops 5

1.3.2 Recombinant Plasmids in Research 7

1.3.3 Recombinant Antibodies for Diagnosis 9

1.3.4 Recombinant Human Insulin 12

1.4 Therapeutic Plasmids in Humans 13

1.4.1. Plasmid vs mRNA Vaccines 14

1.4.2. Approved Plasmid Vaccines 20

1.4.3 Plasmids for Gene Therapy 22

1.4.3.1 Plasmid Design Considerations 23

1.4.3.2 Approved Plasmid Gene Therapies 26

1.5 Controllable Plasmids 28

1.5.1 Regulating Plasmid Expression with Small Molecules: the Tet Systems 29

1.5.2 Regulating Plasmid Expression with Light 30

1.5.2.1 Natural Light-responsive Proteins for Plasmid Regulation .. 32

1.5.2.2 Photocaged Plasmids 34

1.6 Summary of Thesis Work..... 38

2. Development of a Kunkel-inspired Plasmid Caging Strategy _____ 40

2.1 Introduction..... 40

2.1.1 Kunkel Mutagenesis to Introduce Site-specific Mutations..... 40

2.2. Chapter Overview..... 42

2.3 Nick-Exonuclease-Extension-Ligation (NEEL) Method to Resynthesise Plasmids..... 43

2.4 Optimising the Ligation Reaction..... 46

2.5 Optimising NEEL 50

2.6 Kunkel Approach to Resynthesise Plasmids 54

2.7 Removal of Unreacted Plasmids..... 61

2.7.1 Optimising Nicking Reaction..... 61

2.7.2 ssDNA Purification..... 62

2.7.3 DpnI Degradation of Methylated DNA..... 66

3. Regulating Bacterial and Phage Promoters_____ 70

3.1 Introduction..... 70

3.1.1 Transcription Regulation within Bacteria..... 70

3.1.2 Tac Promoter..... 73

3.1.3 T7 Promoter..... 74

3.1.4 Sp6 Promoter..... 75

3.2 Chapter Overview 76

3.3 Optimising IVT 76

3.4 Optimising UV Irradiation 79

3.5 Optimising IVTT ON State 81

3.6 Optimising IVTT OFF State 86

3.7 Quantifying Plasmid Modifications 87

3.7.1 HPLC and LC-MS to Quantify Amine Modifications 87

3.7.2 Quantifying PCB Modifications using Dye-labelled MVS 90

3.8 LA-plasmids with Caged T7 Promoter 96

3.9 Caging the Tac Promoter 104

3.10 Caging the Sp6 Promoter 111

4. Regulating a Mammalian Promoter and Plasmid Delivery _____ 116

4.1 Introduction 116

4.1.1 Mammalian Transcription and the Core Promoter Elements ... 116

4.1.1.1 TATA Box and Inr 118

4.1.1.2 B Recognition Elements 118

4.1.2 Nucleic Acid Delivery 119

4.1.2.1 Problems Associated with Viral Vectors 119

4.1.2.2 Liposomes as Nanocarriers 120

4.1.2.3 LNP-based Gene Therapy and Vaccines 122

4.1.2.4 Outstanding Challenges for LNP-based Therapeutics 123

4.1.2.5 Transferrin-mediated Gene Delivery 124

4.1.2.6 Albumin-mediated Gene Delivery 126

4.2 Chapter Overview 127

4.3 Nickase promiscuity revisited 128

4.4 Lambda Exonuclease 133

4.5 Caging the TATA Box 136

4.6 Caging the Initiator and B Recognition Elements 140

4.7 Self-delivery with Biotinylated Transferrin and BSA 146

5. Conclusion and Future Directions _____ 154

6. Methods _____ 160

6.1 Materials 160

6.2 General Methods 161

6.2.1. Homologous Recombination 161

6.2.2 PCR 162

6.2.3 Agarose Gel Electrophoresis 163

6.2.4 Ethanol Precipitation 163

6.2.5. Nicking Reactions 163

6.2.6. Exonuclease Reactions 164

6.2.7. Phosphorylation Reactions 164

6.2.8. Annealing Reactions 164

6.2.9. Extension and Ligation Reactions with PCB Oligos 165

6.2.10. DpnI Reactions 166
6.2.11. Measuring Concentrations with Pico488 166
6.2.12. Binding Streptavidin 167
6.2.13. PCB-NHS Ester Reaction with Amine Oligo 168
6.2.14. HPLC Purification of PCB Oligo 168
6.2.15. Denaturing PAGE 169
6.2.16. LC-MS 170
6.2.17. Maxiprep 170
6.2.18. UV Irradiation 171
6.3. Chapter 2 Methods 172
6.3.1. Gel extraction of ssDNA 172
6.3.2 Electroelution of ssDNA Directly into Wells 173
6.3.3. G-capsule to Purify ssDNA 173
6.3.4. Dialysis Membrane to Purify ssDNA 174
6.3.5. Crush and Soak to Purify ssDNA 174
6.3.6. PCB-NHS ester Reaction with Amine Plasmids 175
6.3.7. Gyrase Reactions 175
6.3.8. Summary of Primers, PCR, and Homologous Recombination Conditions, and HPLC Details in Chapter 2 176
6.4. Chapter 3 Methods 177
6.4.1. IVT with T7 RNAP 177
6.4.2. IVTT of the T7 and Tac Promoters 178

6.4.3. IVT with Sp6 RNAP 178

6.4.4. Plasmid and amino-C6-dT Primer Digestion 179

6.4.5. HPLC to visualise digested plasmid 180

6.4.6. Binding MVS-Atto647 to PCB plasmids 181

6.4.7. Summary of Primers, PCR, and Homologous Recombination Conditions, and HPLC Details in Chapter 3 182

6.5. Chapter 4 Methods 183

6.5.1. Binding WTS to Bt-Cy5 183

6.5.2. Binding Bt-Tf and Bt-BSA to WTS-Cy5 184

6.5.3 Native PAGE 184

6.5.4. HPLC Purification of mNG 3-PCBs Oligo 185

6.5.5. Maintaining HEK293T and HeLa-GFP Cells 185

6.5.6. Plating Cells 186

6.5.7. Transfection with Lipofectamine 3000 186

6.5.8. Epifluorescence Microscopy 187

6.5.9. Transfection with B-Tf/BSA 187

6.5.10. Cell Viability Assay 187

6.5.11. RNA Extraction 188

6.5.12. cDNA Synthesis 189

6.5.13. RT-qPCR 190

6.5.14. Fluorescence Plate Reader of mNG Plasmids 190

6.5.15. Summary of Primers, PCR, and Homologous Recombination Conditions, and HPLC Details in Chapter 4 190

1

Introduction

1.1 Plasmid Overview

Deoxyribonucleic acids (DNA) encode genetic information that endows life. Across all domains, this information exists in the form of chromosomes. On the other hand, plasmids are structures more commonly found in prokaryotes and archaea, although some eukaryotes such as yeast naturally possess plasmids.[1] Like chromosomes, plasmids carry the same genetic language that ultimately gets transcribed to produce ribonucleic acids (RNA), and translated into proteins. However, there are distinct differences that also set plasmids and chromosomes apart.

Plasmids are circular, double-stranded extrachromosomal DNA that are much smaller than chromosomes and hence contain less information (Figure 1.1A).[2] They have their own origin of replication (ori) and can self-replicate independently of genomic DNA.[2] The ori also dictates their copy number, which varies in cells, whereas the chromosomal number is constant for a given cell.[3] Whereas chromosomal DNA

defines a cell, and must be passed on to daughter cells, the gene(s) encoded by plasmids are not considered vital but may confer a selection advantage, such as antibiotic resistance.[2]

Since their discovery in the 1950s,[4] plasmids have become a mainstay in biology. Their size, simplicity, and ease of manipulation have allowed scientists to study cell biology and biochemistry, understand and build complex gene networks to mimic life, and engineer novel functionalities that would otherwise be impossible. Indeed, the ability to combine and modify DNA segments from different species gave rise to recombinant DNA technology.[5] Its fruits have permeated across all walks of life and facilitated tremendous breakthroughs in science and medicine. To understand the importance of plasmids and their contributions to modern life, it is necessary to appreciate how the field of recombinant DNA came to be.

1.2 Discoveries that Led to Recombinant Plasmids

In the 1960s, scientists studying the growth of λ bacteriophage on *Escherichia coli* (*E. coli*) observed that the phage's DNA was being degraded.[6–8] Phages are viruses that infect and kill their bacterial hosts. They realised that some bacteria could produce enzymes that cleave invading DNA at specific sites (now known as restriction sites) as a defence mechanism. These enzymes were termed 'restriction enzymes' due to their ability to restrict phage colonisation. This discovery enabled scientists to cut DNA at specific sequences to produce the required DNA fragments.

Around the same time, Gellert and co-workers observed that *E. coli* extracts could join the 2 ends of phage DNA into a covalently-closed circle.[9] The reaction required a 5′-phosphate and a 3′-hydroxyl group on either end of the linear fragments, which come together to form a phosphodiester bond. They later purified the active enzyme and

named it ‘DNA-joining enzyme’, or what is now called DNA ligase. Restriction enzymes and ligases effectively gave biologists a means to combine DNA fragments (Figure 1.1B).

In the early 1970s, a series of findings catapulted the fields of molecular cloning and recombinant plasmids. It was known that some bacteria possessed R-factor plasmids (essentially plasmids containing an antibiotic resistance gene) that enabled their survival under antibiotics. Hsu *et. al.* showed that calcium chloride-treated *E. coli* in the presence of purified R-factors could transform these bacteria into drug-resistant cells.[10] Analysis of these bacteria revealed they had acquired circular DNA that was indistinguishable from the R-factor plasmids. This observation illustrated that plasmids could act as a vector to insert foreign DNA into a new host to confer phenotypic advantages.

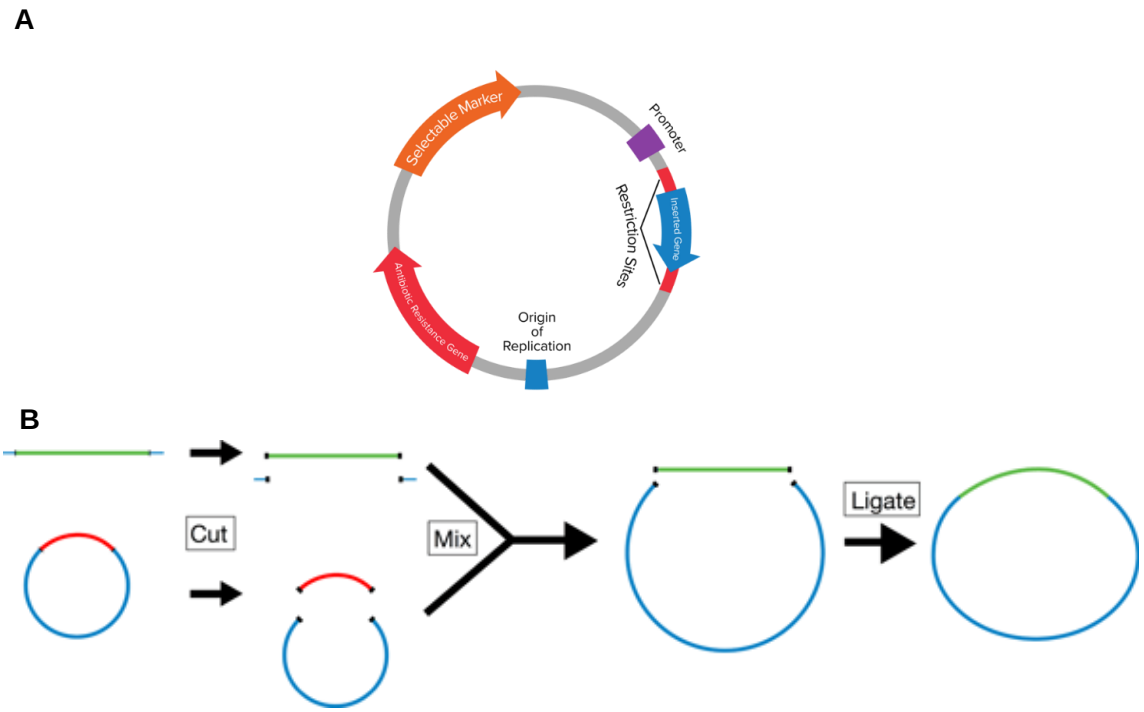


Figure 1.1. Plasmid and recombinant DNA technology

A. General features of a plasmid. B. Discovery of restriction enzymes and ligases enabled recombinant DNA technology. Figures were adapted with permission from Addgene.

Davis and colleagues showed that the R1 endonuclease could generate staggered single-stranded breaks in simian virus 40 (SV40) DNA that bore R1 recognition sequences.[11,12] The staggered cuts resulted in short overhangs which were identical and complementary. The linear SV40 DNA thus contained cohesive ends that could form hydrogen bonds and be covalently closed by *E. coli* Ligase to give the original circularised DNA. Significantly, they demonstrated that SV40 and λ DNA cut with R1 endonuclease could also be rejoined through the same cohesive ends by ligase.

Helling *et. al.* took this one step further by transforming *E. coli* with a recombinant plasmid created from two plasmid sequences deriving from different species.[13] The original plasmids either contained tetracycline or streptomycin resistance genes. When these were cut with EcoR1 endonuclease, the DNA fragments could anneal at their cohesive ends and become ligated. The colonies showed resistance to both antibiotics. Plasmid obtained from these clones had a weight similar to the combined weights of the individual plasmids. When digested with EcoR1, the resultant species were indistinguishable from the EcoR1-digested monomers.[13] Through this, it was understood that different DNAs, when cut by the same endonuclease, would produce indistinguishable ends. Furthermore, DNAs from different origins could be fused at their restriction sites and transformed into a host to give a composite, fully functional plasmid possessing the same base sequences as the parent plasmids.[13]

Last, the advent of electroporation aided the introduction of *in vitro* synthesised plasmids into bacterial cells that were otherwise resistant to traditional modes of transformation.[14] Although modern cloning techniques such as Gibson Assembly[15–18], Golden Gate[19–21], and polymerase chain assembly[22] have simplified cloning, traditional restriction-ligation cloning analogous to the steps used by these early researchers is still widely employed across molecular biology laboratories.

1.3 Applications of Recombinant Plasmids

Altogether, these pioneering studies provided the foundation for recombinant DNA technology. Scientists suddenly possessed a limitless ability to rapidly synthesise proteins of any sequences, which in turn led to developments in sectors as diverse as foods, research, and therapy. Some key demonstrations of how recombinant plasmids have affected our world are outlined below. The list is by no means exhaustive, but these examples should serve to illustrate the power and potential this technology holds.[5]

1.3.1 Recombinant Plasmids in Foods and Crops

Chymosin is a protease widely used in cheese production. It is naturally found in baby calves' stomachs, where it helps to digest milk. Prompted by 'a reduction in the veal calf market',[23] Lowe and co-workers sought an alternative for chymosin production. After cloning the gene for prochymosin (precursor of chymosin) into a bacterial plasmid, induction of the promoter led to prochymosin synthesis. Acidification and neutralisation gave active chymosin, which proved as effective as calf chymosin in milk clotting assay.[23] Today, fermentation-derived chymosin is the main source of chymosin used in the U.S. and represents the first genetically engineered additive approved by the Food and Drug Administration (FDA).[24]

Vitamin A deficiency could lead to sight-related problems and in many cases, total blindness or even death in children.[25] This problem is pronounced in countries where rice is the predominant food source since it does not contain provitamin A (β-carotene), which is converted into vitamin A inside the body. In 2000, a group led by Potrykus engineered the entire β-carotene biosynthetic pathway into immature rice embryos to produce provitamin A.[25] The transformation involved only 2 plasmids and 4 genes in

total. Drake *et. al.* enhanced the provitamin A synthesis by 23-fold by simply switching one of the genes in the pathway with a different orthologue (Figure 1.2A).[26] The authors estimated that with only 72 g of their Golden Rice, 50% of children's recommended dietary allowance of vitamin A could be delivered. This would alleviate vitamin A deficiency for 124 million children globally and potentially save 1-2 million children's lives.

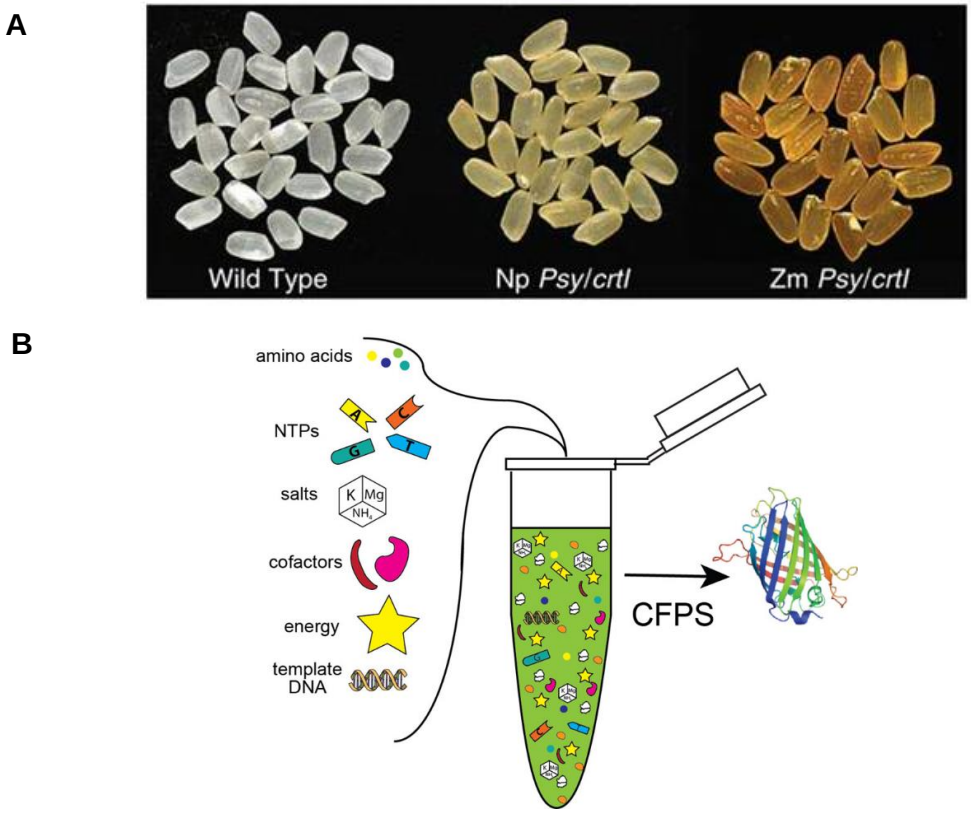


Figure 1.2. Applications of recombinant plasmids in food and research

A. Comparison between the wildtype, first (Np) and second generation (Zm) Golden Rice. The latter contained the highest amount of β-carotene giving it a distinctive orange hue (figure adapted from reference 26 with permission from Springer Nature).[26] B. CFPS enables facile *in vitro* protein production by the addition of a DNA template (plasmid or linear), an enzyme mix, and an energy solution (figure adapted from reference 27 with permission from the American Chemical Society).[27]

1.3.2 Recombinant Plasmids in Research

Recombinant plasmids enabled controlled, large-scale production of proteins in high yield and purity, which is often not possible by reliance on the native strains. Often, the desired proteins are produced in low abundance in the native host or are contaminated by undesired enzymes and nucleases. Ultimate flexibility in the selection of expression vector, strain, and vector optimisation through judicious choice of biological parts could maximise enzyme production whilst minimising strain on the host.[28] This also lowers batch-to-batch variation and processing time, ensuring reproducibility. Indeed, the majority of restriction enzymes, polymerases, and ligases routinely used in cloning are recombinant proteins.[29]

One such example where recombinant proteins have facilitated research is the Protein synthesis Using Recombinant Elements (PURE) system for facile *in vitro* protein production.[30] Transcription and translation are complex processes that involve up to 36 enzymes. To obtain proteins with the desired properties, rigorous testing with alternative DNA sequences is often carried out. In addition to plasmid construction, the plasmid must be transformed/transfected into the expression chassis, and the protein expressed and purified. Cell-free protein synthesis (CFPS) enables *in vitro* expression from the DNA template simply by the addition of the template to the pre-assembled transcription-translation mix in the presence of an energy solution, thus enabling easier and faster access to the protein products (Figure 1.2B).[27] Synthesised proteins could be immediately assayed without further purification. This massively simplifies workflow, cuts down experimental time, and enables rapid DNA prototyping. Other benefits of using CFPS include the ability to produce hard-to-express proteins (toxic,[31] misfolded,[32,33] aggregated[34]) in large quantities, incorporation of unnatural amino acids for protein labelling or to control protein expression,[35] portability and ease-

of-use through freeze-drying of the reaction components. The latter has been used to produce stable, low-cost, on-demand vaccines at the point of care.[36,37]

A method for CFPS relied on cheap cell lysates. Essentially, these are crude cell extracts containing components for *in vitro* protein synthesis. After culturing, the desired extract is separated from other components by centrifugation, whereby the necessary enzymes for CFPS are distinguished from cell membranes and organelles through their weight differences. Depending on the protein of interest, lysates obtained from different sources (*E. coli*, wheat-germ, rabbit blood, etc.[38]) are available to accommodate alternate post-translational modifications. However, the minimal purification involved in their preparation meant that extraneous enzymes present, such as nuclease, could compromise yields.[39]

In contrast to impure cell lysates, the PURExpress™ is well-defined.[40] Each individual N-terminal His-tagged enzyme involved in transcription and translation is recombinantly expressed and purified from a host, meaning PURE only contains the necessary components at a pre-determined ratio for optimal protein synthesis. Unlike cell lysates, the absence of nuclease meant that both plasmid and linear DNA templates could be used, as the latter won't be degraded by exonuclease. Since its inception, major advances have been made in the synthesis and purification of the PURE components to lower the cost and processing time. This includes the cloning of the hexahistidine sequences into the 38 genes involved in the transcription-translation machinery,[41] or co-culturing of the different hosts expressing these components, followed by bulk lysis and purification.[42–44]

1.3.3 Recombinant Antibodies for Diagnosis

Antibodies are proteins employed by the immune system to detect and overcome invasion of foreign materials. Natural antibodies are commonly made up of two identical heavy chains and two identical light chains, associated through disulfide bridges (Figure 1.3, left).[45] In turn, these chains contain a variable domain and one or more constant domains. The variable domains are responsible for the binding and thereby the specificity of a particular antibody-antigen pair.

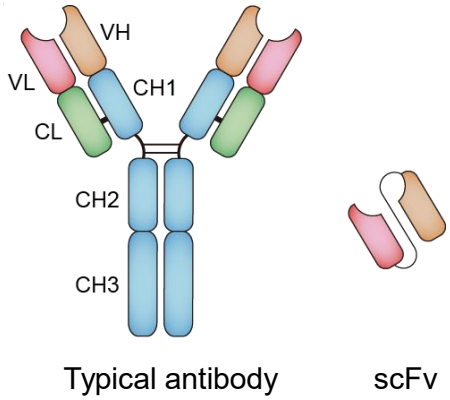


Figure 1.3. Comparison between an antibody and a scFv

Antibodies consist of multiple variable and constant domains whereas a scFv represents the minimal epitope recognition unit and contains a variable light and heavy chain connected through a short linker. Figure adapted from CreativeBiolabs.

Recombinant antibodies have found widespread use in research as well as medicine for diagnostic and therapeutic purposes.[46] Single chain variable fragments (scFv) are a type of recombinant antibodies that contain only the variable light and heavy chain regions (V_L and V_H, respectively) connected by a short peptide linker. They represent the minimal antibody unit capable of binding to the epitope (Figure 1.3, right).[47] scFv preserves the recognition capacity of the native antibody despite the removal of the constant domains.[48] Their popularity is due to the well-established and streamlined

antibody phage display method used for the selection of extremely potent scFv with affinity to virtually any target. Given the importance of recombinant antibodies and the generality of this approach, a brief explanation of the process is warranted. Selection for high-affinity scFv begins with antibody library preparation.[47]

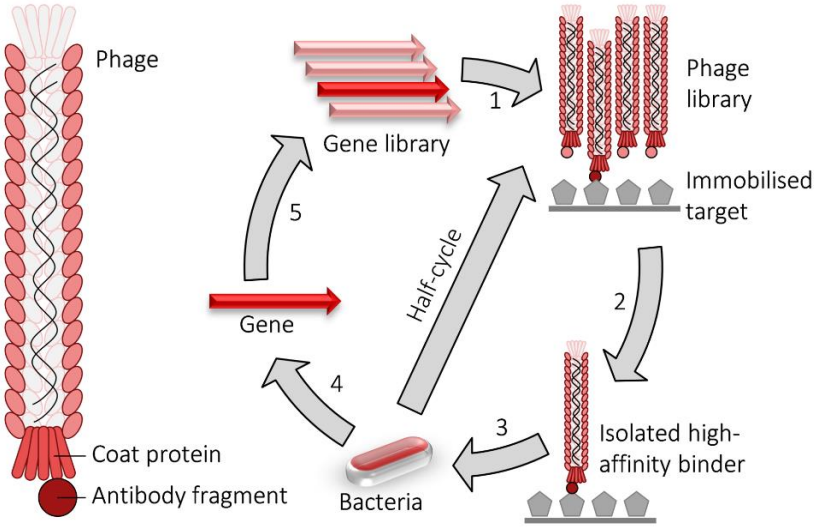
It's known that antibody-encoding genes contain conserved regions across different species. This enables amplification of the V_L and V_H domains using primers complementary to these regions. Messenger RNA (mRNA) of the antibodies is initially harvested from immune cells of naive or immune hosts (exposed to a particular antigen or pathogen). Reverse transcription produces complementary DNA (cDNA), which is then amplified by the above primers to produce the host's entire antibody repertoire.[47] The individual PCR product is then inserted into a phage display vector (Figure 1.4A). This is simply a plasmid engineered for the production of the scFv fragment fused to a capsid protein. When co-transformed with a helper phage, *E. coli* produce and release phages that display the scFv on its surface whilst the capsid protein localises to the phage's membrane.

The selection process involves immobilisation of the target antigen and iterative rounds of phage binding, washing to remove unbound phages, elution of phages containing strong-binding antibodies, and finally re-amplification in *E. coli*.[47] After 3 or 4 rounds, the pool of candidates is narrowed, whilst the strong candidates are enriched.

The selected scFv are then tested for binding to their target through enzyme-linked immunosorbent assay (ELISA, Figure 1.4B).[47] This involves sequential binding of the phage pool to immobilised antigens, followed by binding of an enzyme-linked anti-phage antibody and incubation with a substrate. Enzyme-substrate interaction

produces a fluorescent signal that denotes phage binding. *E. coli* is then re-transformed with successful binders, and individual colonies are retested with ELISA. Strong binders could then be sequenced so the base sequence of high-affinity scFv could be used to clone and express said binders.

A



B

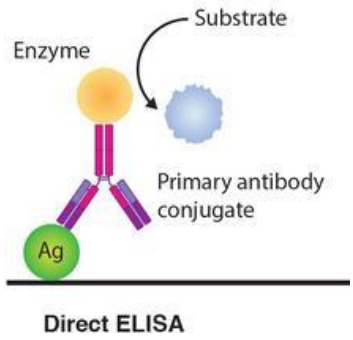


Figure 1.4. Antibody phage display workflow and testing with ELISA

A. General workflow to select high-affinity scFv binders (figure adapted from Thomas Shafee under the CC BY-NC 4.0 license terms). B. Confirming binder affinity with ELISA (figure adapted from reference 49 with permission from IntechOpen).[49] See main text (Section 1.3.3) for more details.

This technique was used to develop a recombinant antibody to diagnose the rabies virus.[50,51] Rabies infection kills up to 55,000 people globally.[52] Failure to correctly diagnose the virus is catastrophic, as it is only treatable shortly after exposure, and proves fatal once clinical symptoms begin.[52] Current diagnostic techniques employ recombinant monoclonal antibodies, but their access is costly, difficult, and time-consuming to prepare. The cell lines used are at high risk of contamination and are genetically unstable. In a comparison against a commercial rabies antibody, the recombinant scFv98H performed similarly. scFv98H was highly specific and sensitive and could bind strongly to antigens across multiple rabies strains, whilst not suffering from the production drawbacks faced by traditional antibodies.[52]

1.3.4 Recombinant Human Insulin

Insulin is a hormone naturally produced by pancreatic β-cells in humans.[53] It lowers blood glucose levels by promoting its absorption into the liver and muscle cells, where it is stored as glycogen or fats. Diabetes is typified by hyperglycaemia, a condition associated with high blood glucose levels, resulting from either a genetic autoimmune reaction (Type I) or lifestyle choices leading to low insulin secretion and loss of sensitivity to insulin (Type II).[54] Diabetic patients require insulin to maintain their glucose balance. Before recombinant plasmids, insulin was obtained from pigs' and cattle's pancreases. Eight thousand pounds of pancreas were needed for a single pound of insulin, and 56 million animals were required to satisfy the U.S.' annual demand.[55] To overcome this supply issue, in 1978, recombinant insulin was prepared and later became the first drug synthesised using recombinant DNA technology to be licensed.[53]

The initial synthesis involved expression of the A and B chain monomers separately from recombinant plasmids. After purification, their reaction to form linking disulfide bridges gave intact insulin.[56] Commercial recombinant insulin tends to exist in the oligomeric form at high concentrations, in contrast to naturally occurring insulin which exists in the monomeric form.[57] Thus, recombinant insulins are slow-acting as they first need to dissociate before entering the bloodstream. A simple workaround involved the inversion of Pro28 and Lys29 on the B chain, which reduced hydrophobicity and dimerisation. Lispro, the fast-acting insulin analogue, was prepared from a recombinant plasmid containing Lys28 and Pro29 expressed in *E. coli*. The proinsulin produced was enzymatically cleaved to give insulin.[53] Long-acting insulin was also similarly synthesised from recombinant plasmids by altering residues to raise the protein's isoelectric point.[58] These studies illustrated the flexibility and power of this technology to rapidly deliver drugs with diverse properties.

1.4 Therapeutic Plasmids in Humans

Thus far, plasmids have been shown to generate proteins that play crucial roles in human lives. These plasmids are typically transduced to an expression host, which then produces the desired proteins. The hosts are often simple organisms with quick doubling time such as *E. coli* or baker's yeast (*Saccharomyces cerevisiae*) to maximise protein production. However, as shown with Golden Rice, the recombinant plasmid could also be introduced to complex eukaryotes to provide the desired response/phenotype. In this way, plasmids can be used in humans for therapeutic purposes.

1.4.1. Plasmid vs mRNA Vaccines

In recent years, there has been a shift away from traditional vaccine methods towards nucleic acid-based vaccines. Historically, vaccines consisted of a dead or weakened form of the pathogen or a recombinant protein.[59] In comparison, DNA and mRNA vaccines do not involve the handling of pathogenic sources or arduous protein purification. Furthermore, the construct development and production of DNA and mRNA vaccines is generic, rapid, and require less optimisation.[59] Unlike protein vaccines, nucleic acids, particularly DNA vaccines, are more thermostable, less prone to denaturation, and hence do not require stringent storage conditions.[60] By using the host's translational machinery, the antigen produced would have similar folding and post-translational modifications, which is more difficult to achieve when the recombinant protein is exogenously expressed.[61]

The idea that plasmids could be used as vaccines dated back to 1990 when Felgner and co-workers showed that naked plasmids could be directly injected intramuscularly into mice to give the sustained expression of the various encoded proteins for at least two months following administration.[62] Not long after, a study demonstrated that mice injected with a plasmid encoding the influenza A nucleoprotein could generate cytotoxic T lymphocytes (CTL) when challenged with a different influenza A strain; this was the first example of a plasmid vaccine in an animal model.[63] Similar studies in the same period reported that mRNA could also elicit *in vivo* protein[62] and CTL[64] production. This sparked a surge of interest in the research community to develop nucleic acid-based vaccines.

The foremost challenge for nucleic acid therapeutics is their delivery across the cell membrane owing to the negative charges on the phosphate backbone (Figure 1.5).[59]

Hence, their formulation often contains a vector that protects and trafficks the nucleic acid cargo. This often comes in the form of positively charged chemicals (see Section 4.1.2 for more details) that can bind and effectively neutralise these charges. The positive charges also enable the association of the nanocarrier with the negatively charged plasma membrane, followed by their endocytosis.

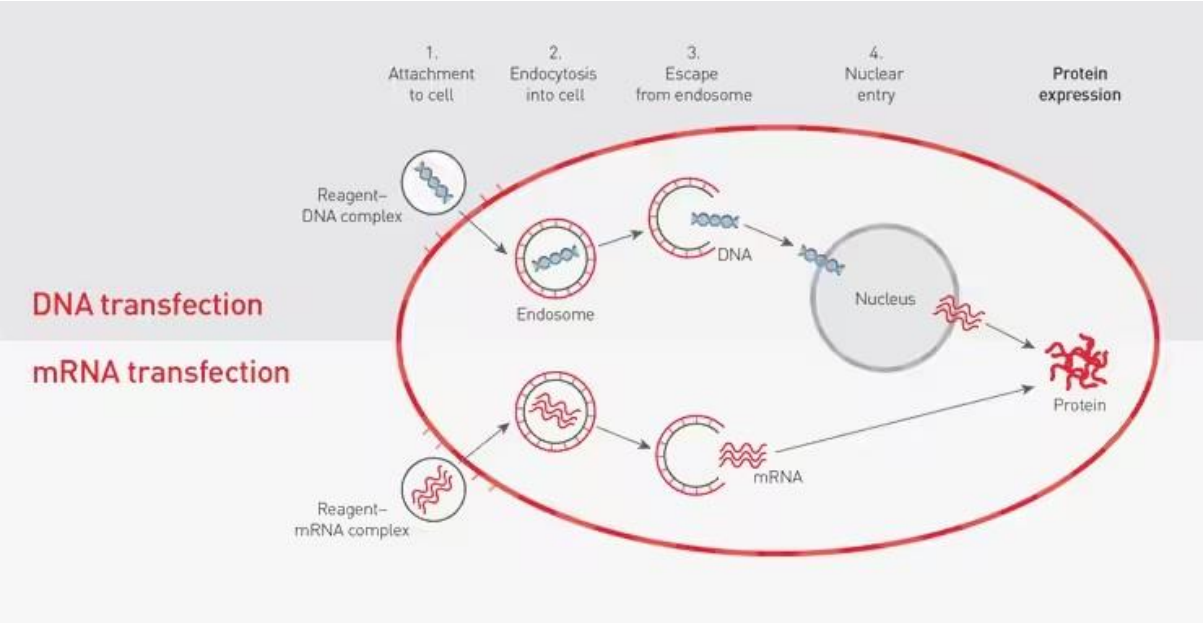


Figure 1.5. Overview of the nucleic acid delivery process

Transfection of nucleic acids often requires a vector to protect and deliver the nucleic acids into cells. Transcription and translation only occur if the cargos can escape from the endosome, withstand nuclease degradation, and enter the nucleus (in the case of linear DNA and plasmid templates). Figure courtesy of Thermo Fisher Scientific.

Once inside the cell, the nucleic acid may remain within the endosome whereby it gets degraded by lysosomes once the endosome has matured.[65] Productive expression could result only if the cargo could exit the endosomes. Here, the fates of mRNA and DNA diverge. Whereas DNA must enter the nucleus to be transcribed before it could be translated in the cytoplasm, mRNA translation could straightforwardly begin once endosomal escape has occurred.[66] In addition to the nuclear barrier, DNA vaccine dosage is typically much higher compared to mRNA (mg[67] versus hundreds of µg[68],

respectively). Despite these inherent issues, there are properties unique to plasmids that make them ideal as therapeutic agents.

Plasmids are covalently closed entities and are not exposed to exonuclease degradation the way linear templates are. Within the body, the high presence of diverse non-specific RNases can cleave RNAs before they enter cells.[69] To illustrate the importance of closed versus opened plasmids, their *in vivo* activities were compared. Although the immediately following discussion relates to plasmids, a similar extrapolation could be made for mRNA as it is also an opened linear nucleic acid. Fresh plasmid purified from bacteria exists predominantly in the supercoiled form (SC), with a minority in the nicked/open-circular (OC) and linearised form (Figure 1.6).[70] A direct comparison between SC and linearised plasmids showed that the latter retained only 40-50% transfection efficiency in mammalian cells. This decrease was attributed to the removal of the linearised template by nucleases within the cellular milieu.[70] In a different study, the authors showed expression from OC and linear templates was decreased by as much as 80% relative to the SC plasmid.[71]

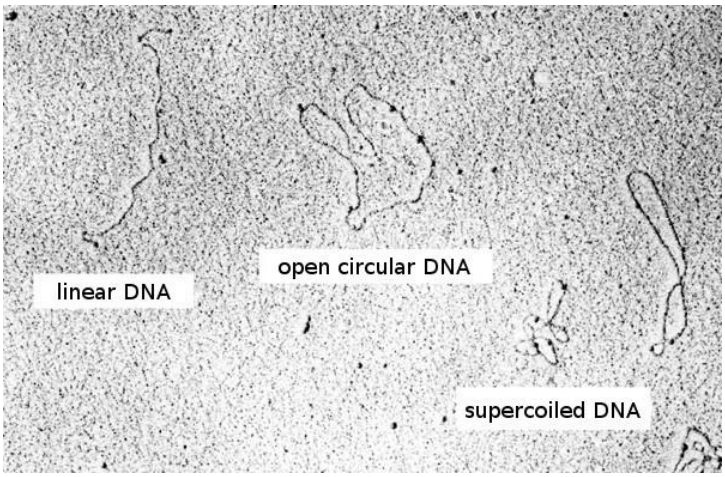


Figure 1.6. Different plasmid conformations

Electron micrograph showing the linear, OC, and SC plasmid conformations. Figure adapted from Wikipedia Commons under the CC BY-SA 3.0 license terms.

To understand the cause for this discrepancy, the different DNA topologies were injected directly into the nucleus. This time, all DNA templates performed similarly.[71] In contrast, when the DNAs were injected into the cytosol, once again SC plasmid significantly outperformed its counterparts. Incubation of SC plasmid with cytoplasmic cell lysate showed that although 50% of the SC plasmid had been converted to the OC form within 20 mins, template nicking slowed down beyond this point, and SC plasmid could still be detected after 160 mins. Last, the authors also measured the size of the different DNA templates. The hydrodynamic radius of the OC form was twice that of SC and linear DNAs.[71]

These observations suggested that expression was not dependent on the DNA topology, but rather the presence of DNA in the nucleus. However, to reach the nucleus, the template must withstand the nuclease present within the cytoplasm. Sustained presence and the small, compact size of SC plasmid could enable it to find its way to the nucleus.[71] Although linear DNA was similar in size to SC plasmid, it was more stretched out and could become entangled within the mesh of the actin cytoskeleton. This would be exacerbated by its susceptibility to exonucleases.

In terms of production, mRNA is transcribed *in vitro* from a DNA template, whilst plasmids are obtained through batch fermentation. As the former does not involve any cellular work, its regulatory process would be simpler,[72] although this is offset by its manufacturing cost.[73] In comparison, bacterial fermentation and plasmid purification from *E. coli* are relatively simple, inexpensive, and produce large amounts of the plasmid.[60]

Compared to DNA, mRNA is more immunostimulatory which may prove toxic.[59] However, immunogenicity is not necessarily undesired, as stimulation of the immune

system is needed to initiate an immune response. Indeed, by leveraging RNA's ability to induce immunogenicity, non-coding mRNA has been used as an adjuvant in rabies vaccine[74] and cancer trials.[59]

However, overstimulation could lead to undesired outcomes including RNA degradation or abortive translation,[75] or adverse effects on the patient. Although plasmids themselves are also immunologically active, their effects are fewer and better defined.[59] The instability of RNA and its inflammatory effects probably accounted for the lacklustre interest in RNA therapy at the outset.

Renewed enthusiasm for RNA-based vaccines came with the introduction of modified nucleosides into the RNA sequence. Specifically, 5-methylcytidine (5mC), N6-methyladenosine (m6A), 5-methyluridine (m5U), and pseudouridine modifications dampened RNA immunostimulatory effects through reduced cytokine release.[76] Pseudouridine in particular could enhance protein production and mRNA stability.[77]

Theoretically, both mRNA and plasmid benefit from an amplification effect compared to proteins. Multiple protein copies can result from a single mRNA unit; this effect is even more pronounced for plasmids as they can create multiple copies of mRNA.[59] In reality, this augmentation is balanced by the poor transfection property of nucleic acids. It is estimated that only 1 in 10,000 mRNA molecules could successfully escape the endosome, and this lowly figure still yet takes into account RNA instability in the cytoplasm.[75] For plasmids that entered cells, it's predicted 0.1% would end up in the nucleus.[78]

However, low gene expression does not make nucleic acid poor vaccines. Activation of the immune system by antigen leads to the proliferation of antibodies, and hence fewer copies of the antigens may yet be conducive to efficacy. Liu reasoned that after

the initial dose (prime), higher-affinity antibodies are generated when antigens are scarce. Second vaccine administration (boost)[79] could lead to the expansion of these strong binders.[59] Given their greater nuclease resistance, plasmid's persistence within cells would inevitably be longer – its extrachromosomal presence in muscle after 19 months has been reported.[80]

mRNA	Plasmid
Expressed in cytoplasm	Expressed in nucleus
Effective at lower dose (µg)	Higher dosage required (mg)
Exposed ends, prone to exonuclease degradation	Covalently-closed, resistant to exonucleases
More immunogenic	Less immunostimulatory
Produced using *in vitro* transcription	Batch fermentation
Less thermostable and more prone to denaturation	More thermostable, hence easier storage

Table 1.1. Comparison between mRNA and plasmid platforms for vaccines

The often-cited argument against the use of DNA in therapy is the possibility of insertional mutagenesis.[81] However, extensive testing with various DNA vaccines has alleviated the concern over DNA integration.[82,83] This risk is minimised by the removal of sequences that could lead to recombination and integration with genomic DNA during the design of the plasmid vector.[84] Though the possibility of integration still exists,[85] this risk is relatively low. The authors proposed that for a worst-case scenario, an integration frequency of 1.3×10^{-9} events per gene would still be 3 orders of magnitude less than the naturally occurring gene-deactivating mutations.[85,86] In a different study, mice injected with three different plasmids still retained $1.0\text{-}4.0 \times 10^3$ plasmid copies/µg host DNA after six weeks, which dropped to $0.2\text{-}0.8 \times 10^3$ after six months. Despite their prolonged presence, no integration event was detected.[83] At these levels, recombinant plasmids fall within the FDA's requirement for DNA

vaccines: that the frequency of plasmid integration should be less than spontaneous mutation.[87] These differences between plasmid and mRNA vaccines are summarized in Table 1.1.

1.4.2. Approved Plasmid Vaccines

Given the early successes with plasmid vaccines in animal models, efficacy in humans has generally remained low. Efforts to enhance their potency have included: co-administration with immunostimulatory adjuvants (or the adjuvant encoded on the plasmid itself)[88], heterologous prime-boost vaccination,[89] or reengineering of the plasmid parts (alternate promoters, additional CpG islands to raise the immune response [repeating sequences of cytosine followed by guanine in the 5′-3′ direction; these motifs are commonly found in pathogens],[90] codon optimisation,[91] etc.).

As of 2019, five plasmid-based vaccines have been approved for animal use. Of these, a vaccine against the West Nile virus (WNV) for horses was especially promising.[92] The low effectiveness of DNA vaccines in humans was believed to be due to the inefficient antibody-inducing effect of plasmids, and the relative size difference between humans and typical animal models. However, the equine WNV vaccine addressed these issues by generating sufficient antibodies to neutralise the virus in horses.[59]

The coronavirus disease (COVID-19) resurged interest in nucleic acid vaccines. Dozens of COVID-19 plasmid-based vaccines were in clinical trials.[60] Needle injections, as well as physical methods such as electroporation and jet injection, were the main modes of delivery for plasmids (Figure 1.7).[60] Amongst these, ZyCoV-D became the world's first DNA/plasmid vaccine approved for human use for COVID-19.[93,94] The vaccine uses a high-pressure jet of fluid to puncture the skin to deliver the

vaccine into the dermal layer. This enables better cell uptake and hence increases antibody production.[95,96]

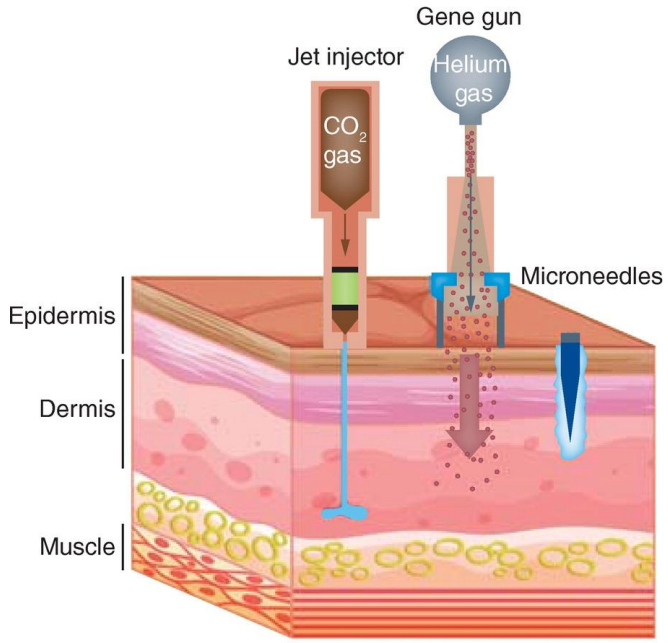


Figure 1.7. Physical methods for gene delivery

By delivering nucleic acids physically, the gene therapy can penetrate deep into the skin and cells, thus requiring a lower dose and exhibiting higher efficacy. Figure adapted from Future Oncol. (2007) 3(5), 569-574 with permission of Future Medicine Ltd.[60]

Compared to the two major mRNA vaccines from Pfizer and Moderna that showed 95% protection against symptomatic infection, ZyCoV-D exhibited only 67% efficacy.[93,94] However, the comparison is not straightforward, as ZyCoV-D trials were performed when the Delta variant, a more infectious COVID-19 strain, was the predominant strain, whereas the mRNA vaccines' figures were derived from an earlier, less transmissible variant.[93,94] Despite this, and in agreement with previous reports on DNA vaccines, ZyCoV-D is less potent. The plasmid vaccine requires three 2-mg doses, whereas the mRNA vaccines only use two doses of 30-100 µg.[93]

ZyCoV-D's licensure is momentous for DNA vaccines. Similar to the jet injector used for ZyCoV-D, gene guns can propel plasmid-coated microparticles directly to the nucleus using a jet of pressurised gas.[97] Plasmids delivered in this manner required much smaller doses (1-10 μg) to elicit immune responses in animal and human trials.[98] Clearly, the overriding problem with DNA vaccines has been the delivery of the nucleic acid to the site of action.[99] However, the reliance on a physical device for delivery is not ideal. This significantly increases the cost, as well as limits distribution as the specialised device is impractical and less available compared to the conventional needle injection.[60,93]

Meanwhile, there are numerous plasmid vaccines in trial to combat a range of diseases including cancer,[100] diabetes,[101] human immunodeficiency virus (HIV), influenza virus, and Zika virus.[94,102] In clinical trials against WNV[103] and Ebola and Marburg viruses[104], the vaccines were well tolerated and the majority of subjects could produce neutralising antibodies. In sum, the burgeoning field of plasmid vaccines is growing exponentially with the realisation that the plasmid platform is safe, affordable, flexible, and can rapidly produce vaccines against sudden epidemics. ZyCoV-D's approval would hopefully bring needed confidence in this modality and act as an impetus for more plasmid vaccines to come.

1.4.3 Plasmids for Gene Therapy

Aside from vaccines, there are many ways plasmids could be used to treat diseases. By supplying a copy of a normal gene, sufficient proteins could be generated. This is particularly relevant for genetic diseases resulting in defective proteins such as cystic fibrosis,[105] or autoimmune diseases whereby the body's cells attack its proteins, exemplified by the deterioration of insulin in Type 1 diabetic patients.[106] Alternatively,

plasmids can encode for proteins that play a central role in eliminating the pathogen or tumour cells. Said proteins can kill their target, and these in turn can come in different formats: cytotoxic peptides,[107] bacterial toxins,[108] and proapoptotic signals.[109] These proteins can also play an intermediary role by converting a substance into an active ingredient, for example, when co-administered with an inactive prodrug.[110] Finally, plasmids encoding for short hairpin RNA (shRNA)[111] or small interfering RNA (siRNA)[112] can inhibit the translation of disease-causing genes by degradation of the target mRNA through the RNA interference pathway.

1.4.3.1 Plasmid Design Considerations

Unlike plasmid vaccines that can benefit from immunostimulation and transient properties, plasmid gene therapies should provide long-lasting or even permanent protein production without triggering the body's immune response.[59] As mentioned in Section 1.4.1, plasmids are naturally immunoactive as the body can recognise bacterial DNA sequences. Bacterial DNA is characterised by rich unmethylated CpG motifs.[113] Mammalian DNA contains much lower CpG dinucleotides; and when available, they are usually cytosine-methylated.[99] CpG islands are thus classified as pathogen-associated molecular patterns (PAMPs), and their presence is detected by various pattern-recognition receptors (PRR). These receptors then initiate a signalling cascade that alerts the immune system to mount a response.[113,114]

One strategy to avoid PRR is to minimise the CpG sequences.[99] However, this is not always possible as these may form part of the introns, or they may have a regulatory role. The alternative approach is to methylate the plasmid. Although this can be done *in vitro*, the associated cost is prohibitive.[99] Ertl *et. al.* showed that by simply transforming their plasmid into the *E. coli* strain ERI1821 which harboured a plasmid

encoding SssI methylase, their desired plasmid was sufficiently methylated to enable prolonged transgene expression within mice.[115] This represents a facile approach to repress the host's immune responses against plasmids.

Plasmids contain features that are necessary for the selection, survival, and propagation of the bacteria but do not contribute to the therapeutic effect of the gene of interest (GOI). Once fermentation stops, these components are no longer required and their extraneous presence may prove detrimental to the plasmid's performance. Larger plasmids are known to decrease transfection efficiency and cell viability.[116–118] Furthermore, genes conferring survival advantage may be horizontally transferred to the gut bacteria through conjugation.[119,120] These antibiotic-resistance genes could also interfere with transgene expression and elicit immune responses.[99] It has been shown that non-coding sequences, regardless of origin, exceeding 1 kb in length could promote gene silencing.[121] In this case, transferring the same non-coding sequences into the 3′-untranslated region (UTR) of the expression cassette did not abrogate gene expression.[121] Thus, it would be advantageous to remove these accessory sequences in the plasmid design step.

Minicircles comprise solely the eukaryotic expression cassette (promoter, GOI, additional regulatory elements) that drives transcription and translation of the GOI.[122] This minimal unit is obtained through recombination events, dictated by the multimer resolution sites (MRS), from a parental plasmid to selectively remove components not involved in the GOI expression; the latter forms the miniplasmid (Figure 1.8).[123] As the minicircle does not possess an ori, it cannot replicate on its own. Hence, the gene encoding the necessary recombinase must be tightly controlled.[124] It is placed under an inducible promoter, and its expression should only happen once enough plasmid has been produced – towards the end of the fermentation.

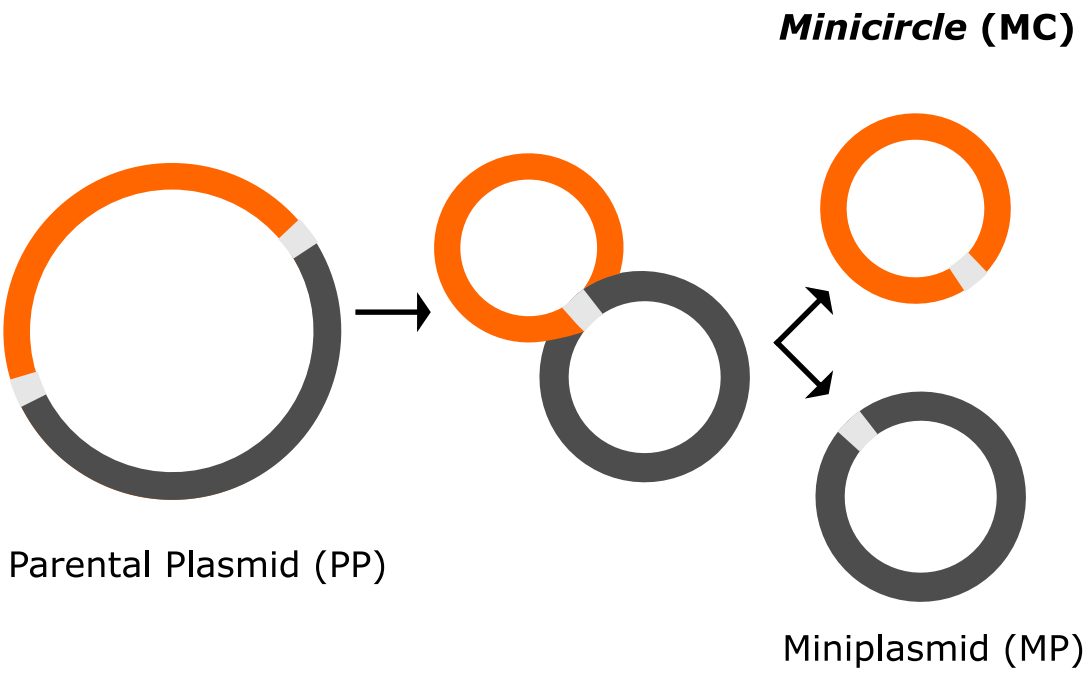


Figure 1.8. Plasmid recombination to form a minicircle and a miniplasmid

The parental plasmid contains recombination sites. During fermentation, induction of the recombinase leads to plasmid recombination that generates a minicircle and a miniplasmid. Figure adapted from PlasmidFactory.

Once the 2 species are formed, it is necessary to isolate the minicircle from the miniplasmid. This has been achieved by placing the lac operator sequence within the minicircle.[123] The minicircle-miniplasmid mixture was then applied to a column containing an immobilised LacI repressor. The latter bound strongly to the operator on the minicircle, whilst the miniplasmid was washed away. The minicircle was then eluted by the addition of isopropyl ß-D-1-thiogalactopyranoside (IPTG), which preferentially bound to LacI.[123] Sustained and increased *in vivo* gene expression has been observed from minicircles, which the authors ascribed to the higher extracellular and intracellular stability.[122]

Other simple changes to the plasmid include the removal of secondary structure-forming sequences and careful selection of the polyadenosine sequence, both of which contribute to plasmid instability in cells.[125,126] Co-administration with nuclease

inhibitors also enhanced the plasmid's nuclease resistance.[127,128] Problematic sequences such as direct or inverted repeats, or sequences with high similarity to the host's genome may give rise to unexpected rearrangement events and should be avoided where possible.[129] Their removal also decreases the plasmid size and minimises potential inflammatory sequences as discussed above.

1.4.3.2 Approved Plasmid Gene Therapies

Between 2010-2020, two plasmid gene therapies entered the market.[130] These treatments utilise a naked plasmid encoding a hormone to stimulate angiogenesis – thus they may be regarded as gain-of-function therapies. Neovasculgen was granted approval in Russia in 2011 for the treatment of peripheral arterial disease (PAD) and represents the first plasmid gene therapy.[131] In 2019, Collategene was given conditional, short-term approval to treat the same disease in Japan.[132]

PAD occurs when a build-up of fatty deposits in the arteries in the legs restricts blood flow to this region.[133] Patients suffer from intermittent claudication (pain in the thigh and calf), tissue loss, and ulcers that do not heal.[133] If PAD worsens, this can lead to critical limb ischemia (CLI). CLI is characterised by severe loss of blood flow and can result in gangrene of the ischemic limb. It is estimated that CLI affects 0.05-0.1% of the global population annually,[134] and is responsible for 30,000-40,000 amputations annually in Russia alone.[131]

CLI is often treated by surgical intervention to restore blood flow (revascularisation) to the region, either through physical removal of the plaque, or stent or bypass insertion.[133] However, revascularisation effectiveness is also temporary, and surgery is not always possible when the occlusion is advanced or the affected site is large.[133]. Without revascularisation, amputation is often the sole recourse, which in itself can

lead to significant loss of quality of life and even death. Overall, CLI patients have a 50% mortality rate in the first 5-year period.[133]

Neovasculgen and Collategene are plasmids that encode for the vascular endothelial growth factor (VEGF) and hepatocyte growth factor (HGF), respectively. Despite their different target proteins, conceptually, they are both angiogenesis factors and lead to similar outcomes (Figure 1.9).[133] Both are injected at the site of ischemia to promote the formation of blood vessels and increase blood flow. Desirable clinical outcomes include a reduction in rest pain, recovery of ulcers, decrease amputation rate, and a better quality of life.[133]

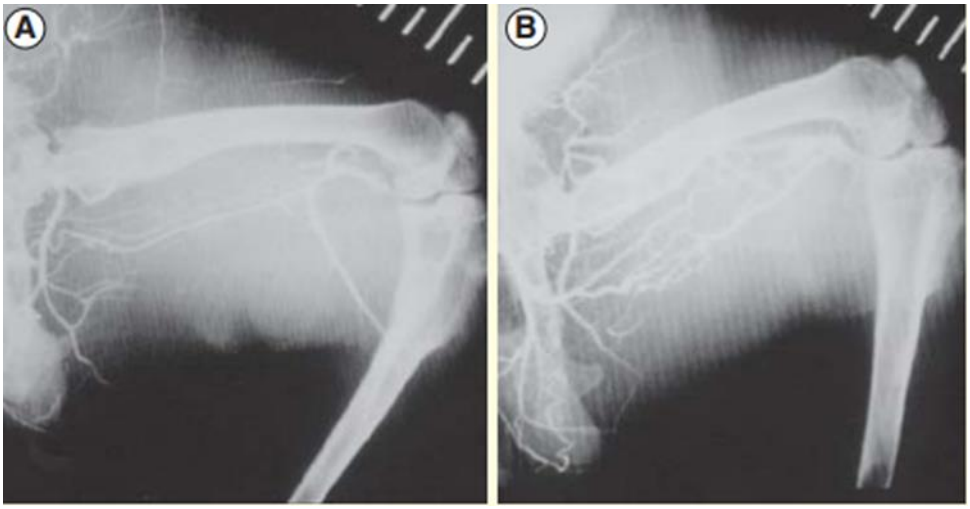


Figure 1.9. Neovascularisation induced by Collategene

Angiograms of rabbit arteriosclerosis rabbit model administered with A. Control plasmid and B. Collategene. Figure adapted from reference 133 with permission from Taylor & Francis.[133]

As both VEGF and HGF are implicated in carcinogenesis, malignant tumour growth and adverse reactions towards neovascularisation were the primary concerns. However, neither drug elicited any specific side effects, the majority of which arose from the underlying disease itself.[131,133] Pain-free walking distance (PWD), and transcutaneous oxygen pressure (TcPO2, a measure of skin oxygenation) were used

as indicators for angiogenesis. For Neovasculgen, the PWD, and TcPO2 increased by 263%, and 26%, respectively, after the 5-year study period,[131] whilst the control group's values changed by -13% and +15%.[131] The amputation-free rates at the end of the study were 95% and 67% for the treated and placebo groups.

For Collategene, the most recent Phase III trial used ischemic ulcer size as a determinant for efficacy. After 12 weeks, 67% of patients' ulcers had healed.[133] Despite multiple clinical trials with diverse viral and non-viral-based therapies in the previous 2 decades, no treatment had been available for CLI. Collectively, these data suggest both plasmid therapies could induce angiogenesis to restore blood flow to CLI patients, offering an alternative to amputation and those patients for which revascularisation is not an option.

1.5 Controllable Plasmids

Although plasmids are powerful tools to understand biology and have the potential to transform medicine, their encoded genes shouldn't be always expressed. To appreciate the causative relationships between one gene and another, or its wider impacts at the cellular/tissue levels, the expression must be carefully controlled in a spatiotemporal manner.[135] Likewise, in the context of therapy, it is necessary that the therapeutic agent activity, especially if it were a cytotoxic peptide or deregulatory RNA, be confined within the intended tissue/tumour. Their systemic expression may lead to unexpected complications.

1.5.1 Regulating Plasmid Expression with Small Molecules: the Tet Systems

Many natural and engineered gene expression control systems have been exploited to regulate gene expression. The archetypal system commonly used to regulate expression in eukaryotic cells is the Tet ON/OFF system.[136–138] In nature, TetR is a transcription factor (TF) involved in bacterial resistance to the tetracycline (Tc) antibiotic, or its related analogue doxycline (Dox). In the absence of Tc, TetR is bound to the Tet operator sequence (TetO). TetA, which acts as a membrane protein to remove toxins such as Tc, is repressed. TetR is sequestered when bound to Tc, enabling the expression of TetA that then pumps Tc out of the cell (Figure 1.10).[139]

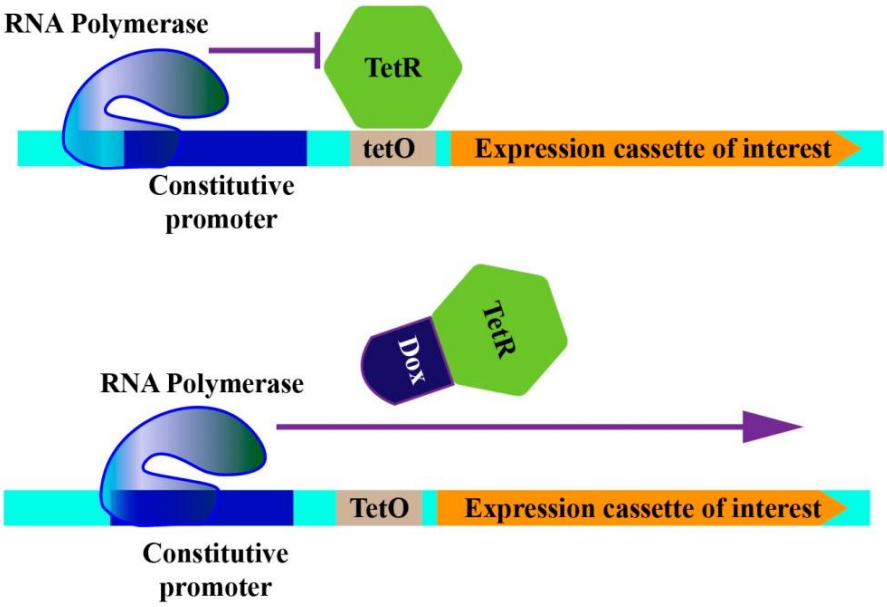


Figure 1.10. The TetR system

In the absence of Tc or Dox, TetR binds to TetO and inhibits transcription. When TetR is sequestered by either antibiotic, transcription is activated. Figure adapted from reference 140 under the CC BY 4.0 license terms.[140]

TetR acts as a repressor as it inhibits transcription from TetR-regulated promoters. However, in the Tet ON/OFF system, the transcriptional repression domain in TetR is

switched to a transcriptional activation domain to create a tetracycline transactivator (tTA).[137] Several repeats of the TetR/tTA binding sequence, TetO, are placed upstream of a promoter; collectively, this sequence is termed the tetracycline response element (TRE) and is where tTA binds. The binding of tTA to TRE enables the expression of the GOI controlled by this promoter. In Tet-OFF systems, Tc association with tTA prevents it from binding to TRE, and transcription is repressed; the opposite is true for Tet-ON systems (Figure 1.11).

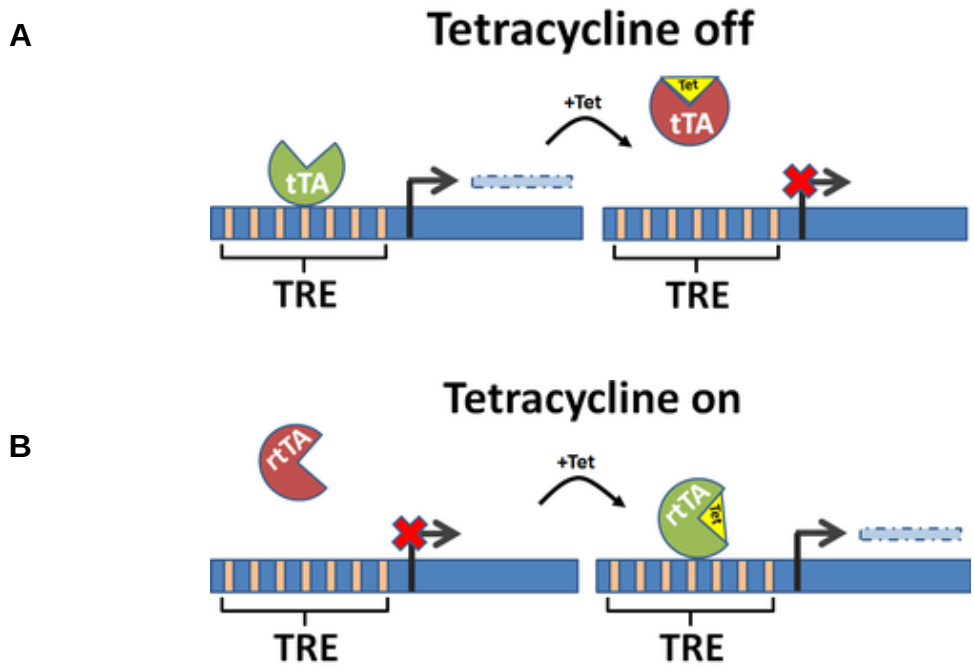


Figure 1.11. Tet ON/OFF systems

A. In the Tet-OFF system, binding of tetracycline (depicted as Tet here) to the tetracycline transactivator (tTA) sequesters tTA and represses gene expression. B. In the Tet-ON system, Tet binding to the reverse tetracycline-controlled transactivator (rtTA) enables its binding to TRE and activates gene transcription. Figures were adapted with permission from Addgene.

1.5.2 Regulating Plasmid Expression with Light

The Tet and related small molecule-inducible systems possess excellent gene regulation and provide a good platform for *in vitro* or bacterial studies. However, the use of small molecules as a trigger comes with many limitations that restrict its

application to more complex model organisms, or where localised control is desired. Temporal and spatial control is impossible with this method due to the diffusion of small molecules across the system, and the inability to direct their presence to a specific site. This also meant that their cellular concentration is low, necessitating excess amounts of the compounds. The residual presence of these molecules in the culture medium or the host could also contribute to basal leakiness.[141] Furthermore, these molecules may be metabolised or interact with unintended cellular pathways to confound experimental results. Antibiotics such as tetracycline and doxycycline involved in the Tet system could not be used in humans for their toxicity, and damage to the gut microbiome.

In contrast to small-molecule inducers, light overcomes the above setbacks and represents an attractive stimulus as it possesses excellent spatiotemporal control as the timing, duration, and site of irradiation are user-defined. Additionally, it is non-invasive and its intensity and wavelength can be adjusted to modulate the penetration whilst minimising cellular damage. Depending on the light-sensitive group, reversible control is possible by alternate applications of light at different wavelengths.[142–144]

Light has been used to remotely activate or repress gene expression through a multitude of natural and synthetic systems.[145,146] Using these technologies, it is possible to regulate transcription and translation using wavelengths across the UV-visible spectrum.[145] The following is a discussion of how these strategies have been adapted to control expression from plasmids, but linear DNA and RNA, and proteins have also been subjected to photoregulation.[145]

1.5.2.1 Natural Light-responsive Proteins for Plasmid Regulation

Natural light-responsive proteins can be found in photosynthetic organisms including algae, plants, and cyanobacteria.[145] These proteins have been exploited through innovative engineering approaches to exert light-controlled expression of the GOI. Broadly, these systems can be divided into 2 categories: following light induction, the photosensitive proteins can either directly participate in transcriptional activation, or they can initiate signalling cascades that also ultimately lead to gene expression.[145]

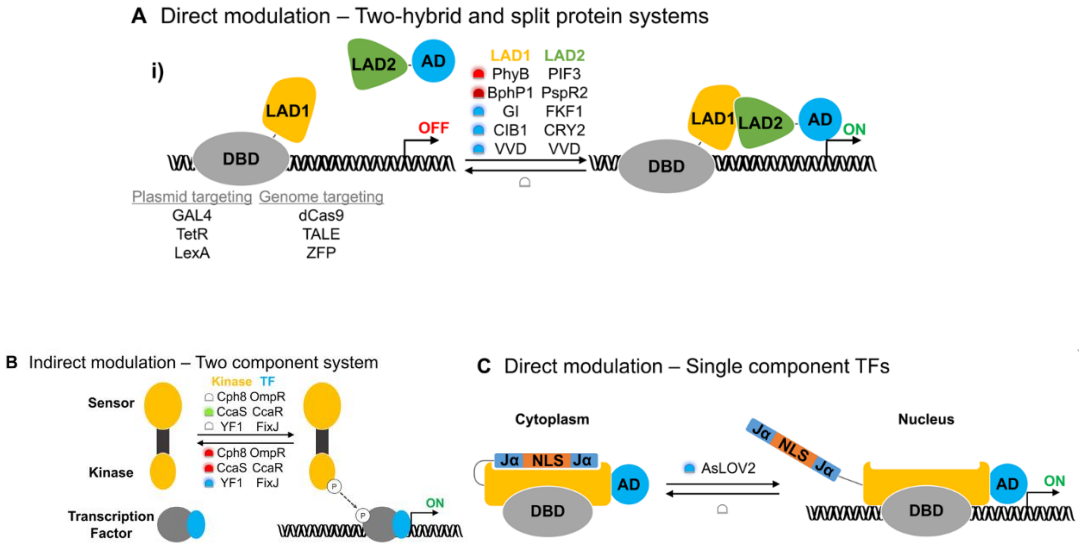


Figure 1.12. Regulating plasmid gene expression using light-responsive proteins

A. Light-activated domains (LAD) dimerise or dissociate in response to different light wavelengths, which in turn activate/repress transcription B. Light absorption by the sensor domain of histidine kinase triggers phosphorylation of the kinase domain. Subsequent transfer of the phosphate group to the transcription factor activates transcription. C. Blue light caused a conformational change of the Jα helix to expose the NLS. NLS-mediated nuclear translocation enabled transcription. Figure adapted from reference 145 under the CC BY-NC 4.0 license terms.[145]

The first use of such systems to regulate gene expression involved a two-component system (TCS).[147] This TCS relied on phytochrome protein B (PhyB) and phytochrome-interacting factor 3 (PIF3). Under short red light pulses, PhyB underwent a conformational change that enabled the 2 proteins to rapidly dimerise, whilst far-red

light disrupted this binding.[147] When these proteins were fused to the DNA-binding domain (DBD) and activation domain (AD) and transformed into yeast[147] or mammalian cells,[148] conditional gene expression only occurred when red light was applied. Far-red light-induced dissociation abolished expression as expected (Figure 1.12A).

The LOV domain of YtvA has also replaced the oxygen-sensing domain of a histidine kinase to create YF1, a TCS that could phosphorylate its response regulator FixJ in the absence of blue light (Figure 1.12B).[149] Phosphorylated FixJ could drive expression from its cognate promoter. A related system whereby phosphorylated FixJ acted as a transcriptional repressor was also created.[150] Thus, blue light could be used to repress or activate gene expression through a signalling cascade.

Such TCS are common and can be prepared by switching either the light-sensitive proteins (BphP1/PspR2,[151] FKF1/GI,[152] etc.) or DNA-interacting proteins (GAL4,[147,148,152], LexA,[153] etc.). Alternatively, a one-component system has been demonstrated using the light-oxygen-voltage domain of *Avena sativa* (AsLOV2) as the light-sensitive component.[154] Here, a nuclear localisation signal (NLS) was fused to the C-terminus of the Jα helix of AsLOV2. In the dark, Jα was folded and hidden away, sequestering NLS from importins, which could bind to NLS and transport it into the nucleus. Blue light caused Jα to unfold, revealing NLS and enabling the construct to enter the nucleus.

DBD and AD have been fused to AsLOV2/NLS to activate gene expression (Figure 1.12C).[154,155] In this manner, a nuclear export signal (NES) could also be fused to AsLOV2. The AD was switched with a repressor domain to give a transcriptional activator. When irradiated with blue light, exposure of NES led to nuclear export of the

construct, hence transcription of the GOI.[156,157] Although some of these systems possessed excellent photomodulation such as the YF1/FixJ TCS shown here, their therapeutic utility is hindered by the requirement for continuous light exposure,[150,158–164] complex and multi-component systems,[147,158–161,163,164] niche promoters,[149,150] and lengthy optimisation.[158,165]

1.5.2.2 Photocaged Plasmids

Instead of relying on photosensitive proteins, light-controlled gene expression could take place directly on the nucleic acids. Photocaging involves covalent linkage between a light-reactive blocking group to either nucleic acids or proteins. This disrupts transcription, and translation, or perturbs the protein activity until photolysis with a suitable wavelength liberates the active molecule. The first use of a photocaged nucleic acid to enable light-dependent activation was demonstrated in 1998.[166] The field has grown considerably since then with abundant examples of caging strategies.[145]

Whilst the majority of these examples were applied to short RNA oligonucleotides or linear DNA templates, there is a dearth of photocaged plasmids in the literature. This is despite several advantages that plasmids possess over linear templates including their resistance to exonucleases, and amplification of mRNA from plasmids.[59] The higher stability also meant they would make better templates *in vivo.* Their neglect by the research community is probably accounted for by the ability to site-specifically synthesise short oligonucleotides with defined modifications, which is not feasible for plasmids due to their size.

Early examples of photocaged plasmids resorted to bulk plasmid reactions to randomly insert photolabile groups across the plasmid. This gave a heterogenous

mixture of products, where each plasmid contained a different number and placement of the caging group.[166–168] Haselton *et. al.* installed ~270 1-(4,5-dimethoxy-2-nitrophenyl) diazoethane (DMNPE) photoblocking groups to the plasmid's phosphate backbone.[166] The Anderson group used an analogous caging strategy. Their DNA template was placed within a liposome carrier containing cell lysate. This was the first example of a light-induced synthetic cell.[167]

Zhang *et. al.* attached their DMNPE-plasmid to lanthanide-based upconversion nanoparticles (UCN). UCNs absorbed light in the near-infrared (NIR) region and emitted high-energy UV to uncage embedded plasmids.[168] NIR has much higher tissue penetration compared to wavelengths in the UV-Vis region and is also less toxic than UV light.[169,170] Despite the excessive photoblockers used across these systems, their caged plasmids showed significant leakiness, which was presumably due to the random placement of the caging moieties on the phosphate backbones. Some plasmids invariably possessed fewer caging groups than others, whilst some may not contain any caging groups within the important promoter/regulatory elements or the coding sequence. Moreover, expression fold change after UV uncaging was moderate across the above examples (<5-fold), reflecting the need to remove hundreds of caging groups per plasmid molecule.

Okamoto and colleagues also modified their entire plasmid but with a coumarin-based caging group.[171] Once again, their method showed aberrant expression in the absence of light and poor photoactivation. The Nagamune group was the first to site-specifically cage their plasmid.[172] To achieve this, a forward primer containing a reactive hexylamine attached to the phosphate group was used in a polymerase chain reaction (PCR) with a regular primer to resynthesise the plasmid. A biotinylated coumarin was appended onto the plasmid via reaction at the amine, and subsequently bound to

streptavidin. As only a single streptavidin was used, the caged plasmid retained 60-70% expression.[172]

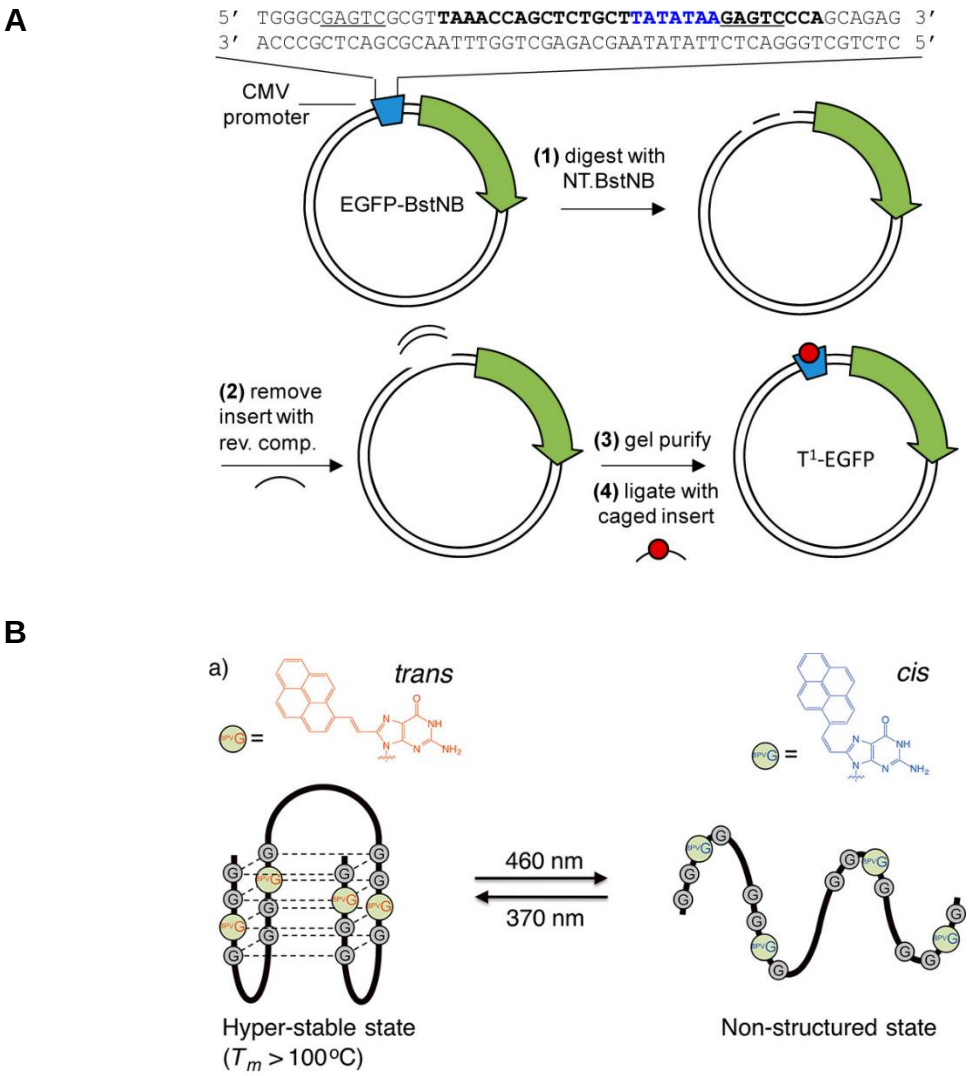


Figure 1.13. Selected literature examples of photocaged plasmids

A. Deiters *et. al.* caged the CMV promoter by annealing a single-stranded DNA (ssDNA) fragment containing photoresponsive groups at the TATA box, which sterically blocked the interaction between RNAP and the promoter. Blue sequence represents the TATA box, and underlined sequences represent the nickase's recognition sequence. B. Ogasawara ligated a dsDNA fragment containing photoreversible ^{8PV}G downstream of a promoter. Depending on the configuration, ^{8PV}G could stabilise/destabilise the G-quadruplex. G-quadruplex formation acted as a blockade to stall RNAP (both figures adapted from references 173 [A] and 174 [B] with permission from the American Chemical Society).[173,174]

To control their plasmids, Deiters *et. al.* cloned two nickase recognition sites flanking the TATA box, an important transcriptional regulatory element, within the cytomegalovirus (CMV) promoter.[173] The plasmid was nicked, and the nicked strand was removed by annealing it to an excess of the reverse complement. An identical sequence containing up to 3 thymines caged with the UV-labile groups 6-nitropiperonyloxymethylene (NPOM) was ligated into the backbone to give the caged plasmid (Figure 1.13A). Using this system, their group was able to repress expression to the background level. After photorelease, expression was restored.[173]

Ogasawara made use of G-quadruplexes to control gene expression.[174] Guanine-rich sequences can form secondary structures known as G-quadruplexes. These are a series of planar guanine tetrads, which in turn are formed through the association of four guanine bases joined by Hoogsteen hydrogen bonding.[174] Monovalent cations such as potassium or sodium can further stabilise these structures by occupying the central position within each tetrad. G-quadruplexes are often found near the promoter regions and they have been implicated in transcriptional regulation.[174]

Ogasawara proposed that by stabilising G-quadruplexes, transcription could be repressed by stalling the RNA polymerase's progress (RNAP, Figure 1.13B). This is different from the methods of gene control discussed thus far, whereby transcription was repressed by preventing RNAP from docking to its promoter. A sequence for a stable G-quadruplex was cloned downstream of a T7 promoter. This sequence contained up to four 8-pyrenylvinyl deoxyguanosine (^{8PV}G) which could undergo reversible photo-induced cis-trans isomerisation.[174] The trans form promoted G-quadruplex formation and resulted in premature transcripts. G-quadruplexes were destabilised when ^{8PV}G was in the cis form, which enabled RNAP to continue transcript elongation.

With a single G-quadruplex, the photomodulation efficiency, defined as the ratio of ON/OFF after photo switching, was 5.8-fold between the cis/trans forms in an *in vitro* transcription (IVT) reaction.[174] With the addition of a second quadruplex, a 23-fold difference was observed. This system was further tested in zebrafish embryos, which showed a 68-fold difference in reporter activity. Furthermore, expression was reversible by repeated cis-trans interconversion.[174]

Of the seven caged plasmid examples, only Ogasawara's, and to an extent, the Deiters' methods gave a meaningful ON/OFF ratio.[174] However, to create the modified oligo, their approaches required modified phosphoramidites that are not commercially available and difficult to synthesise. This made the strategy inaccessible to many laboratories. They also relied on precise placements of nickase recognition sites or a specific G-quadruplex-forming sequence at or around the promoter sequence. This necessitated an additional cloning step to insert the sequence and may interfere with the promoter activity. Significantly, there has been no previous example of a photocaged plasmid system with coupled cellular delivery.

1.6 Summary of Thesis Work

This thesis aims to develop a general approach to creating light-activated (LA) plasmids. Ideally, the strategy should be modular, have a tight OFF state, be sequence agnostic, and not require niche DNA synthesis so it could be widely adopted by the research community to remotely control any plasmid sequence. The following chapters outline how this has been achieved:

Chapter 2 delineates the optimisation process to resynthesise plasmids to incorporate photocaging modifications. The strategy took inspiration from the Kunkel method for

site-directed mutagenesis. Using this approach, plasmid resynthesis, caging, and uncaging was demonstrated.

Chapter 3 applies the photocaging strategy to control expression from a ubiquitous phage T7 promoter. Efforts to improve cell-free reactions and quantify modifications led to the revision of the caging protocol. The improved technique enabled site-specific modifications on the plasmid, which also benefitted the photomodulation efficiency. Attempts to recapitulate these findings on the tac and Sp6 promoters were not successful but revealed a possible alternative means to control these promoters.

Chapter 4 expands this caging technique to regulate gene expression from the CMV promoter in mammalian cells. Photo-dependent activation was measured at the transcription and translation stages. Initial efforts to enable plasmid self-delivery with transferrin and albumin were unsuccessful likely due to protein aggregation.

2

Development of a Kunkel-inspired Plasmid Caging Strategy

2.1 Introduction

2.1.1 Kunkel Mutagenesis to Introduce Site-specific Mutations

Site-directed mutagenesis (SDM) introduces specific changes to the plasmid DNA sequence to alter the structure and function of the encoded protein. Researchers use SDM to study the effect of particular mutations on protein activity and to introduce specific properties into the newly mutated protein.

Kunkel devised a method for SDM to increase the rate of mutagenesis and thereby circumvent the need for mutant selection (Figure 2.1).[175] The sequence to be mutated is cloned into a plasmid vector derived from f1 phage (a phagemid), which is then co-transformed with the M13KO7 helper phage into *E. coli.* The helper phage assists in the packaging and secretion of the single-stranded phagemid into the culture medium.

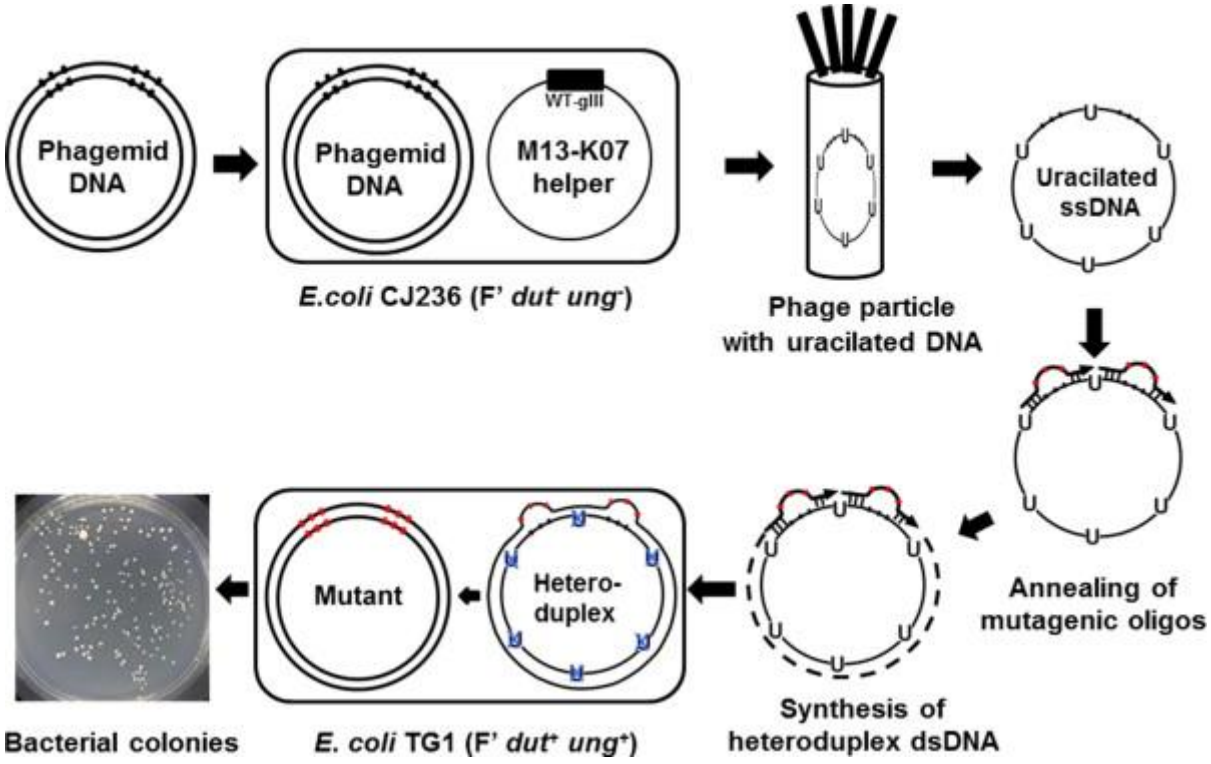


Figure 2.1. Summary of Kunkel mutagenesis

Phagemid containing the desired sequence to be mutated is transformed into a *dut*$^-$ and *ung*$^-$ deficient *E. coli* strain. A helper virus is also infected to amplify phage particles. Uracil-containing ssDNA is used as the template, and primer(s) containing the mutation(s) are annealed, extended, and ligated *in vitro*. The resultant plasmid is then transformed into a *dut*$^+$ and *ung*$^+$ *E. coli* strain. The Uracil-containing strand is removed, and the final plasmid contains the desired mutations. Figure adapted from reference 176 with permission from Elsevier.[176]

The chosen *E. coli* strain lacks the genes encoding uracil deglycosidase (*udg*$^-$) and dUTPase (*dut*$^-$). These enzymes are involved in DNA repair by preventing the introduction of deoxyuridine triphosphate (dUTP) into the bacterial chromosome.[175] Specifically, dUTPase deficiency results in a build-up of intracellular dUTP. This

excess dUTP competes with deoxythymidine triphosphate (dTTP) during DNA replication. Without uracil deglycosidase, the misincorporated dUTP does not get removed, with the end product being a uracil-containing ssDNA.

A primer containing the desired mutation is then annealed to this ssDNA template and extended *in vitro* to give the dsDNA plasmid.[175] This plasmid contains a mutated strand and a uracil-containing strand. The plasmid is then transformed into an *udg*+ and *dut*+ *E. coli* strain. These enzymes specifically remove the uracil-containing strand and replace it with a normal thymine-containing strand. The final heteroduplex plasmid contains the intended mutation.[175]

2.2. Chapter Overview

In this chapter, a method to resynthesise a plasmid encoding the yellow fluorescent reporter protein, mVenus (mV), was described. This method comprised four enzymatic reactions, and these reactions were sequentially optimised. After extensive testing, it was revealed that failure to resynthesise the plasmid was due to a promiscuous nickase and the choice of DNA polymerase (DNAP). By changing these enzymes, and repurposing the Kunkel mutagenesis approach, LA-plasmids were generated, which exhibited UV-dependent photorelease. Leftover, unmodified plasmids could express without the UV stimulus. Thus, various gel purification methods were attempted to remove these unmodified plasmids. However, recovered plasmid materials were impure and in low yields. A short DpnI treatment was sufficient to get rid of unreacted plasmids to leave only the desired LA-plasmids.

2.3 Nick-Exonuclease-Extension-Ligation (NEEL) Method to Resynthesise Plasmids

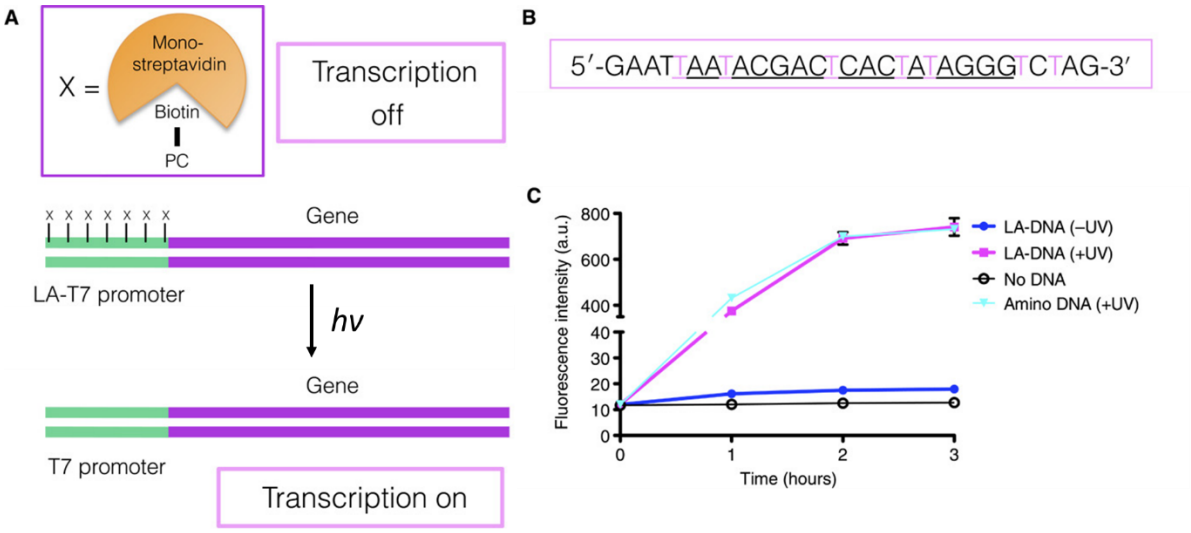


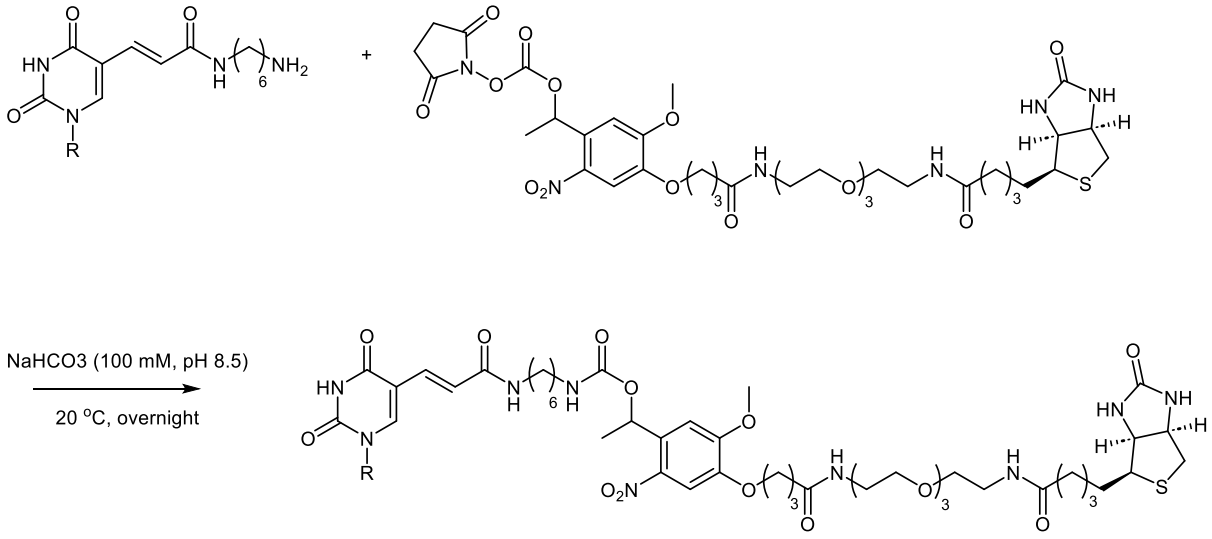
Figure 2.2. Schematic illustrating the LA-DNA caging strategy

A. PCB was attached at the T7 promoter and then bound to streptavidin. LA-DNA blocked T7 RNAP from transcribing until UV was applied. B. The sequence of the amine oligo. The T7 promoter is underlined, and T indicates amino-C6-dT modifications. C. *In vitro* protein expression from amine- or LA-DNA with/without the UV stimulus. Figure adapted from reference 177 under the CC BY-NC 4.0 license terms.[177]

To externally control plasmid expression, the method devised should satisfy the following criteria to maximise its utility. It should be conceptually simple and facile, which permits its use by any laboratory. Importantly, the strategy should not rely on niche phosphoramidites that may limit its transferability. The strategy should also be universal to any plasmid, meaning it should not require any particular plasmid sequence to work and thus is readily adaptable. Last, it should possess a tight OFF state in the absence of the stimulus.

To this end, the light-activated DNA (LA-DNA) method as developed by Booth *et. al.* was chosen as the starting point (Figure 2.2A).[177] In this approach, an oligo containing

7 modified thymines, each possessing a hexylamine (amino-C6-dT) modification, at the T7 promoter was placed directly upstream of the GOI (Figure 2.2B). Amines were nucleophilic groups and acted as a means to attach photocleavable biotin (PCB). The amine-containing oligo was reacted with a small molecule through a substitution reaction at the N-hydroxysuccinimide ester (NHS ester), and purified using high-performance liquid chromatography (HPLC, Scheme 2.1). The small molecule contained 2-nitrobenzyl at one end that acted as the UV-responsive group. The other end contained biotin that could bind to streptavidin proteins.



Scheme 2.1. Reaction between amino-C6-dT on the amine oligo and PCB-NHS ester

An oligo containing 7 modified thymines (amino-C6-dT) was reacted with PCB-NHS ester in presence of sodium bicarbonate to give the PCB oligo. R = the DNA sequence in Figure 2.2B (for the T7 promoter). See Section 6.2.14 (Methods) for more details.

The resultant PCB oligo was then used as a primer in PCR to generate linear PCB DNA. As the amine modifications were situated on the non-Watson-Crick-Franklin face (WCF), base-pairing was not disrupted. PCB DNA was then bound to seven monovalent streptavidin (MVS) to give LA-DNA. Without the ultraviolet light (UV) stimulus, cooperative steric hindrance afforded by bulky streptavidin blocked the T7

RNAP from interacting with its promoter, and minimal fluorescent reporter was detected (Figure 2.2C). UV uncaging removed PCB and streptavidin, enabling binding of the T7 RNAP to its promoter to restore protein expression.

To create UV-responsive plasmids, the same biochemical modifications as those in the LA-DNA technology (aliphatic amine modifications, PCB groups, and streptavidins) were incorporated into the sense strand of the plasmid to disrupt RNAP binding. To install the light-sensitive and streptavidin modifications in plasmids, a multistep method involving four enzymatic reactions was adopted. Initially, the plasmid was nicked by recognition of a unique restriction site on the sense strand by a nickase. Although this necessitated a nickase recognition site, a strategy to overcome this sequence dependency with the Cas9 nickase is discussed in Chapter 5. The nicked plasmid was reacted with an exonuclease, which removed the nicked strand. The ssDNA, corresponding to the antisense strand, acted as a template for primer annealing. One-pot extension followed by ligation gave the re-ligated plasmid (Figure 2.3). This strategy was dubbed NEEL for Nicking, Exonuclease digestion, Extension, and Ligation.

Depending on the choice of the primer (with or without amino modifications) and the presence of amino dUTP in the extension reaction, the re-ligated plasmid outcome may differ. Without modifications in the primer and only regular deoxyribonucleoside triphosphates (dNTP) used in the extension reaction, the resultant plasmid should be identical to the native plasmid, and was referred to as the ‘reformed plasmid’. When the primer contained amino modifications (amino-C6-dT), and/or amino dUTP (either 5'-aminoallyl-dUTP [5'-aa-dUTP] or 5'-aminopropargyl-dUTP [5'-ap-dUTP]) was used in the extension reaction, ‘amine plasmid’ was formed. The amine plasmid was then reacted with PCB-NHS ester to give the ‘PCB plasmid’ (Figure 2.4A and 2.4B). Finally,

PCB plasmids were then bound to streptavidin to create the 'LA-plasmids'. LA-plasmids were expected to repress gene expression until UV irradiation uncaged the plasmids (Figure 2.4C). Throughout this thesis, these terms were used consistently to refer to the respective biochemically modified plasmids.

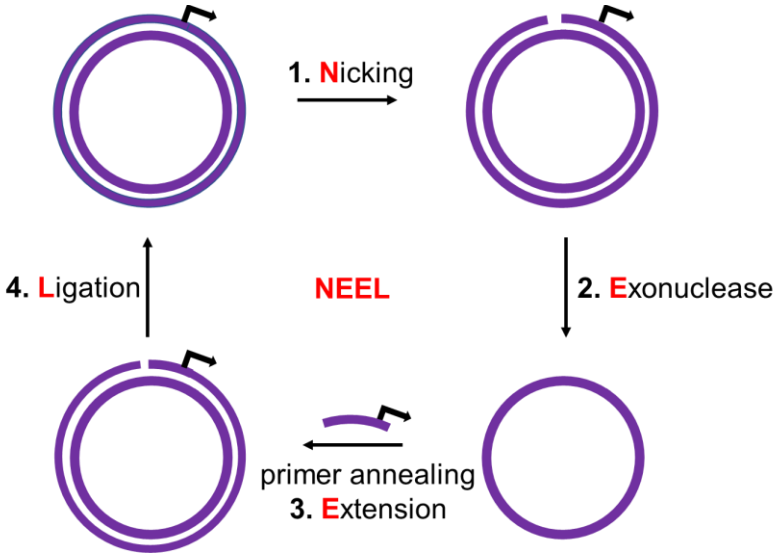


Figure 2.3. Overview of NEEL

Our plasmid resynthesis strategy involving four enzymatic steps: Nicking-Exonuclease digestion-Extension-Ligation. This strategy was dubbed NEEL.

2.4 Optimising the Ligation Reaction

Initially, the mCherry (mC) plasmid expressing a red fluorescent protein was chosen to optimise NEEL as it contained a recognition site for the Nt.BbvCI nickase and its expression could be quantified using CFPS. Transduced linear plasmids are prone to degradation by endogenous exonuclease abundant within cells,[70,71] therefore, LA-plasmids must be covalently closed to increase their lifetime once inside cells. Thus, particular focus was given to the ligation reaction at the outset.

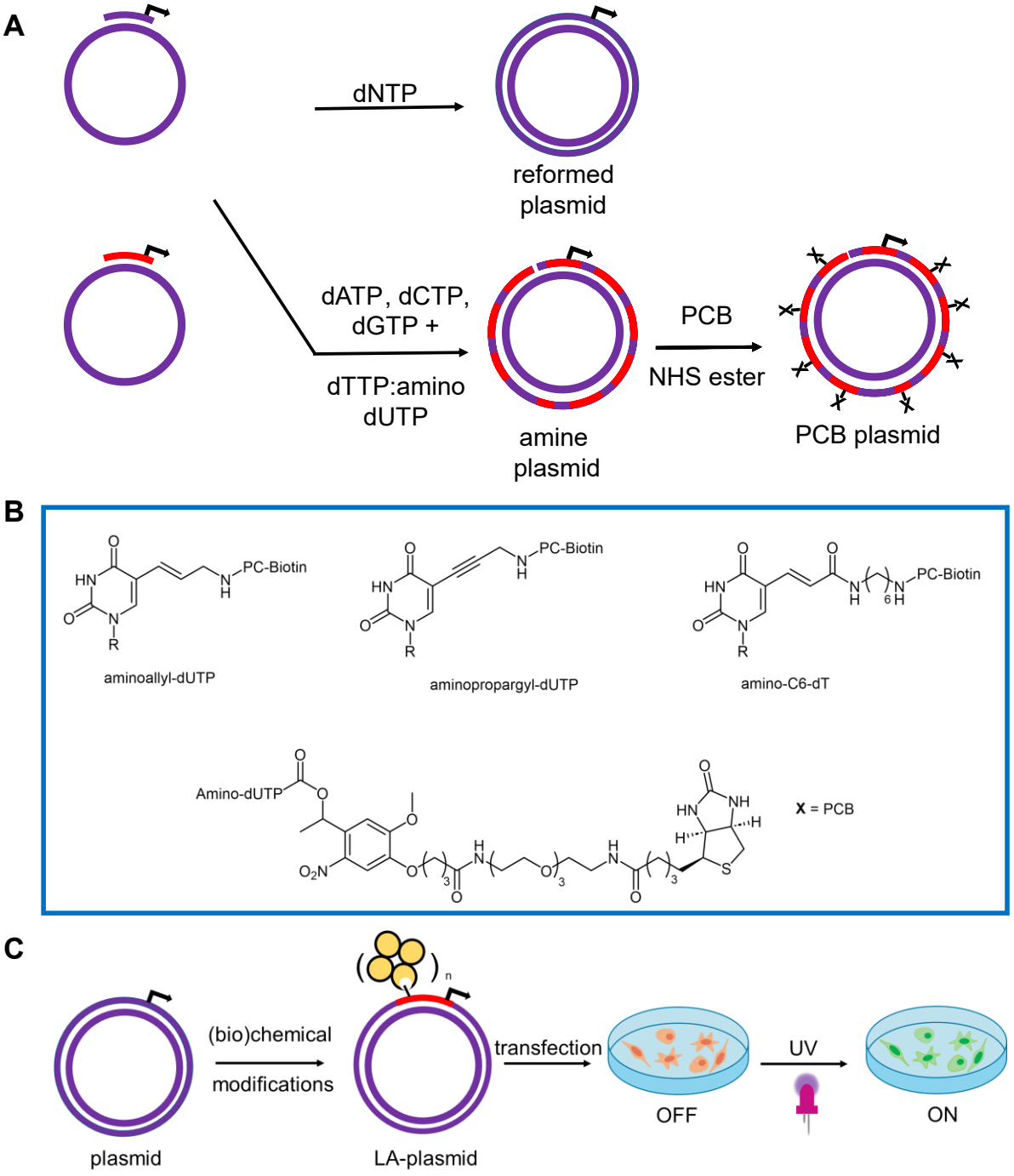


Figure 2.4. Incorporating biochemical modifications into LA-plasmids

A. A regular primer could be annealed and extended with regular dNTP to give the reformed plasmid. Alternatively, when regular triphosphates (A, C, G) and varying ratios of dTTP:amino dUTP (either 5'-aminoallyl-dUTP or 5'-aminopropargyl-dUTP) and/or amine primer (containing amino-C6-dT) was used, the amine plasmid was formed. Reaction with PCB-NHS ester gave the PCB plasmid. B. Structures of amine and PCB modifications. C. LA-plasmid activity is expected to be repressed until UV is applied. R = the deoxyribose triphosphate.

A typical ligation reaction between vector and inserts involves 2 steps: hybridisation between the vector and inserts and enzymatic joining of the 3'-hydorxyl to the 5'-phosphate functional groups. By far, the first step is much more inefficient as the DNA ends must come together and remain associated long enough for the ligase to act.[178] The first step is facilitated at low temperatures, which also helps to stabilise the hydrogen bonds between complementary nucleotides. The optimal temperature for T4 ligase is 37 ˚C.[178] Hence, the ideal reaction temperature is a compromise between the enzyme's optimal temperature, and that which holds the DNA ends in place. In NEEL, the situation is simplified. Once the DNAP has extended the primer across the ssDNA template, the plasmid strands would already have been fully hybridised, precluding the inefficient hybridisation step. For this reason, it was envisaged that ligation in NEEL would be more facile compared to conventional ligation reactions.

This reaction was investigated by first nicking the mC plasmid. The nicked plasmid was then re-ligated at 25 ˚C for 1 h and analysed on an agarose gel. The gel was visualised by post-staining with GelRed, a DNA-intercalating dye (Figure 2.5A). The reaction yielded little ligated product, denoted by the faint presence of ladder bands and a prominent band for the nicked plasmid starting material. The ladder bands represented different conformations of the supercoiled plasmid (their identities were subsequently verified in Figure 2.6). As these plasmids were partially relaxed, they migrated slower relative to the fully supercoiled native plasmid. Poly(ethylene) glycol 8000 (PEG) is a large polymeric molecule that acts as a crowding agent and effectively increases the reagent concentrations, but its addition did not improve the reaction efficiency.

The inefficient ligation was thought to be due to the DNA ends moving freely at room temperature. The reaction temperature was lowered to 16 ˚C to better stabilise the

nucleotides surrounding the ligation junction. As the enzyme operated sub-optimally at this temperature, the reaction time was also extended (Figure 2.5B).

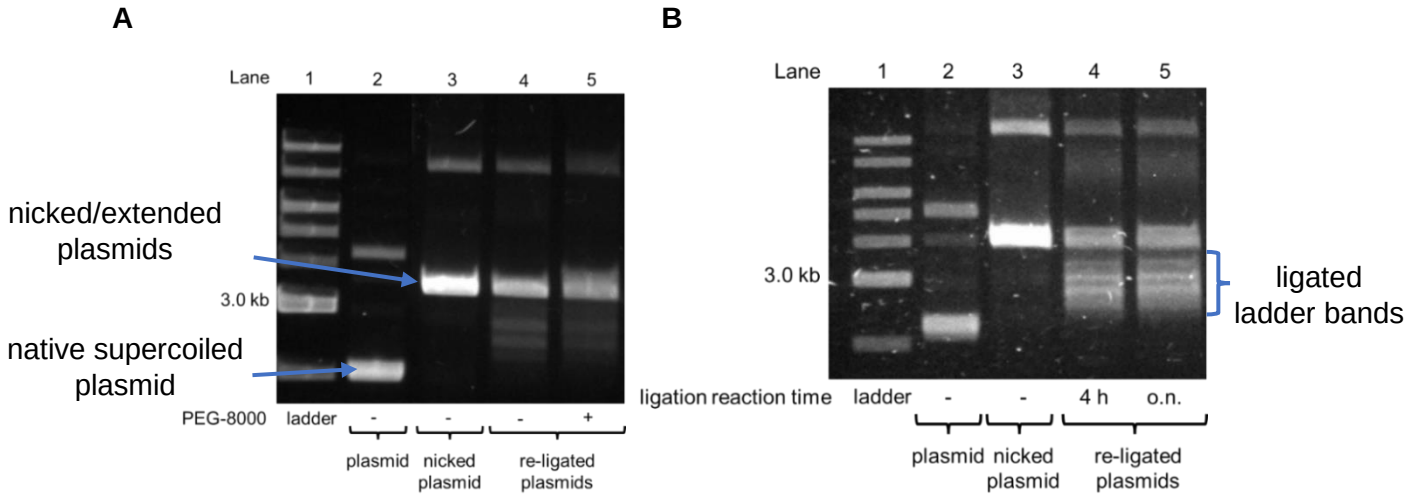


Figure 2.5. Optimising ligation reaction with the mC plasmid by varying reaction time, temperature and PEG addition

A. Ligation at 25 ˚C did not work well, and PEG 8000 did not improve ligation. B. Lowering ligation temperature to 16 ˚C enhanced ligation, whilst prolonging the reaction time did not.

Decreasing the temperature increased the ligated products, although a prolonged reaction, or lowering the temperatures below 16 ˚C did not improve the reaction (Figure 2.5B and 2.6). The presence of ligated plasmids was confirmed by treating the ligation products with exonuclease. As the nicked plasmids were opened, these bands were removed, whilst the ladder bands corresponding to ligated products remained.

The less-commonly used Taq Ligase specifically seals nicks on dsDNA but it could not drive the reaction to completion. Nonetheless, the ligation was again verified by the reaction of the ligated product with gyrase. This enzyme converts covalently closed dsDNA into supercoiled DNA. Thus, the ladder bands were supercoiled and the resultant plasmid looked identical to the native supercoiled plasmid (Figure 2.6).

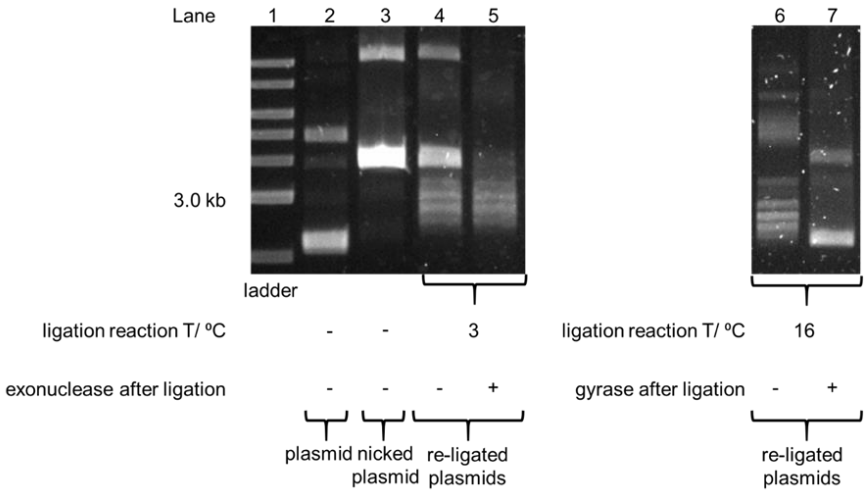


Figure 2.6. Confirming ligation reaction products with the mC plasmid

Lowering the temperature beyond 16 ˚C (lanes 4 and 5) or using Taq Ligase (lanes 6 and 7) did not enhance the reaction. Exonuclease (lane 5) or gyrase (lane 7) after ligation confirmed plasmids had been ligated.

2.5 Optimising NEEL

Having shown ligation worked, NEEL was attempted to resynthesise the mC plasmid. To this end, the ssDNA was formed by sequential nicking and exonuclease reactions. Unlike PCR, where repeated primer annealing and polymerisation steps are performed for 25-35 cycles, only a single round of annealing, followed by concomitant extension and ligation was required in NEEL. This means that repeated cycles of elevated temperatures and rapid cooling for denaturation and primer annealing were redundant.

For each plasmid molecule, only a few thousand nucleotides needed to be incorporated (corresponding to the length of the ssDNA). For polymerases with a high polymerisation rate (>100 nt/s) such as T4 and T7 DNAP, it should take ~20 s to fully resynthesise a 3 kb mC plasmid. These alterations were expected to render the reactions facile.

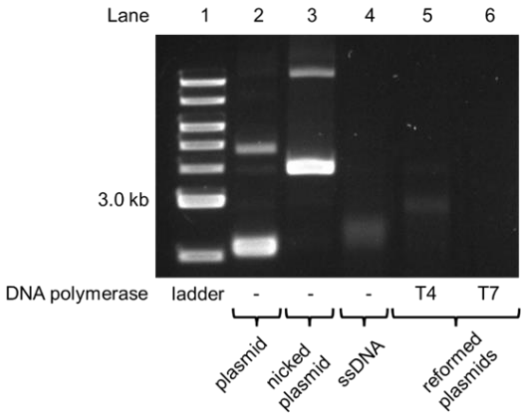


Figure 2.7. NEEL did not give the reformed mC plasmid

NEEL attempts with the T4 and T7 DNA polymerases did not give the extended or ligated products.

Initial attempts at NEEL using either T4 or T7 DNAP failed to give the reformed plasmid (Figure 2.7). After testing with different primers, and high-fidelity and processive DNAPs (Pfu, Phusion, DreamTaq, results similar to Figure 2.7), mC could not be resynthesised. The problem was narrowed down to the nickase promiscuous activity after it was noted that the ssDNA band was inconsistent and faint across different plasmid preparations. If the nickase was cutting at both strands or was able to recognise and nick sequences other than its cognate target, then both plasmid strands may be destroyed after the exonuclease reaction.

This hypothesis was tested with plasmids encoding for either mC or mV, using either the Nt.BbvCI, Nt.AlwI, or Nt.BspQI nickase. Nt.AlwI has a 5-nt recognition site (GGATC) as opposed to the typical 7-nt recognition sequence for most other nickases. This meant Nt.AlwI could nick at multiple sites across either plasmid and acted as the positive control.

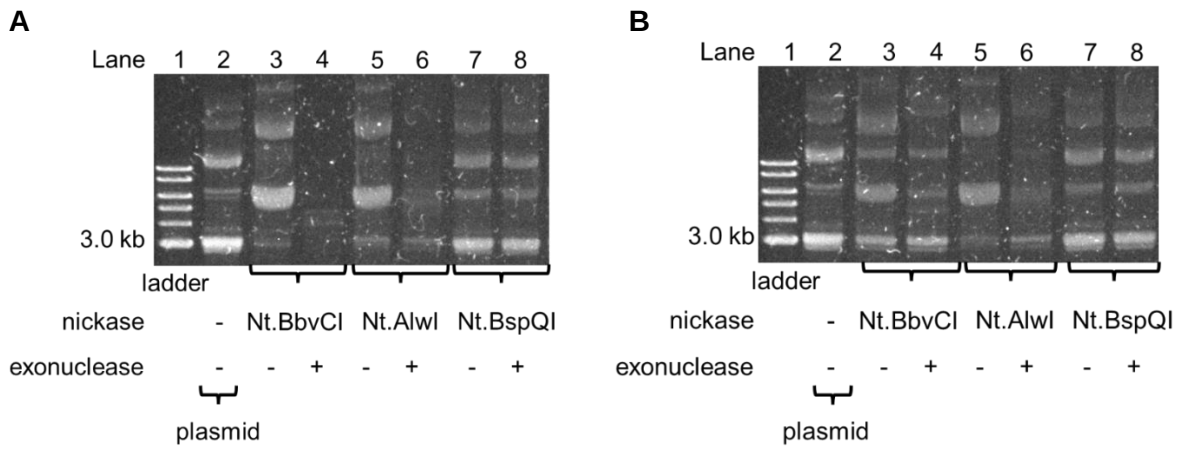


Figure 2.8. Nt.BbvCI was a promiscuous nickase

Testing various nickases' specificity with A. mC and B. mV plasmids.

Of the remaining nickases, the mV plasmid did not contain a recognition site for either nickase and should remain intact, whereas the mC plasmid should only be recognised by Nt.BbvCI. Figure 2.8 shows that whilst the mC plasmid was nicked by Nt.BbvCI as expected, this nickase also nicked the mV plasmid. Neither mC nor mV plasmid was nicked by Nt.BspQI. This meant that the Nt.BbvCI nickase had been acting promiscuously and could account for the lack of extension and ligation products.

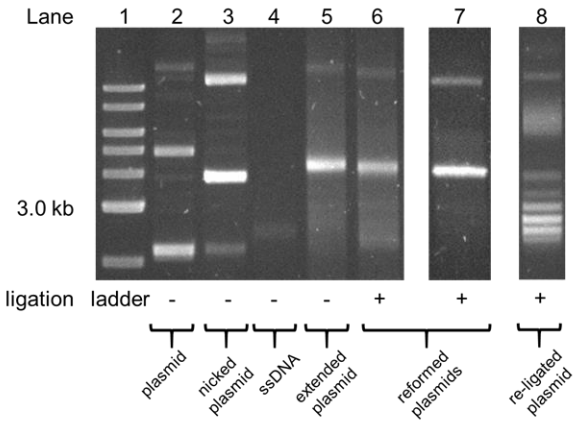


Figure 2.9. The mV_nicks extended plasmid could not be ligated

Using the more specific Nt.BspQI, the extended plasmid could be made. However, ligation was still ineffective when either DreamTaq (lane 6) or Phusion Polymerase (lane 7) was used. Plasmid in lane 8 was nicked and re-ligated and acted as the ligase control.

The BspQI and BbvCI recognition sites were cloned into the mV plasmid to give mV_nicks (Tables 6.2-6.5). This plasmid was again subjected to NEEL resynthesis (Figure 2.9). Without ligation, the fully-extended plasmid should appear identical to the nicked plasmid. When the mV_nicks plasmid was nicked with Nt.BspQI and the nicked strand removed with an exonuclease, the ssDNA template could furnish the extended product. However, ligation was again ineffective, in contrast to prominent ligated bands when the plasmid was nicked and immediately re-ligated. This suggested there was also a problem in the extension step. Specifically, the polymerase employed in lane 6 was DreamTaq, which possessed 5'-3' exonuclease activity. It was thought that the enzyme could simultaneously extend at the 3'-end whilst remove bases at the 5'-end, which could prevent ligation. Repeating NEEL with Phusion DNAP that lacked 5'-3' exonuclease activity did not enable ligation (lane 7).

Although it was clear the extension reaction was problematic, initially, the problem could not be overcome despite extensive optimisation. The search for a suitable DNAP was refocussed, specifically on the enzyme's ability to use modified triphosphates as substrates. It was envisaged that the amino modifications should be placed across the plasmid. Hence, the polymerase must tolerate these modifications. Five different DNAPs were assessed for their ability to incorporate either 5'-aa-dUTP or 5'-ap-dUTP (both amino dUTPs used in this thesis were commercially available, Figure 2.4B) in the absence of dTTP. Extension with Phusion mastermix (MM) was carried out with regular dNTPs to act as the positive control. Although T7, Pfu, and DreamTaq could tolerate 5'-ap-dUTP, only T7 successfully gave the extended plasmid in the exclusive presence of 5'-aa-dUTP (Figure 2.10).

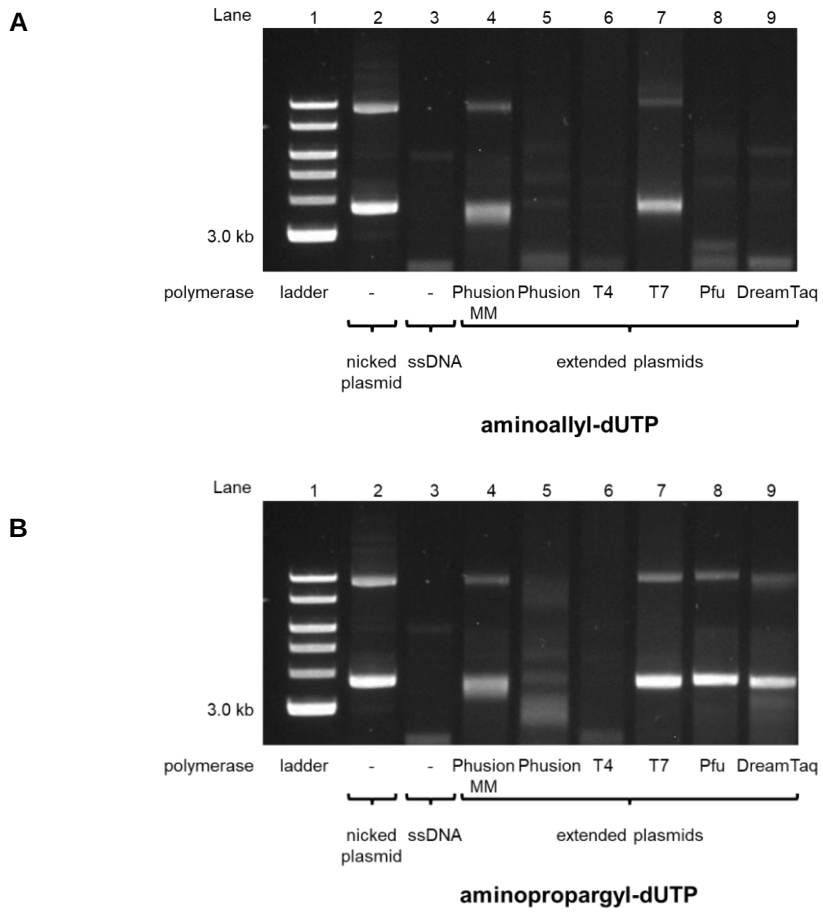


Figure 2.10. Extension for the mV_nicks plasmid with amino dUTPs using different DNAPs

Different polymerases were tested in the extension reaction with the exclusive presence of **A.** 5'-aa-dUTP or **B.** 5'-ap-dUTP in place of dTTP. Reaction with Phusion MM was carried out with regular dNTPs and acted as the positive control.

2.6 Kunkel Approach to Resynthesise Plasmids

Kunkel mutagenesis (Section 2.1.1) is a well-known approach for SDM. Since its inception, the method has been modified and adapted by different laboratories to improve the mutagenesis efficiency, consistency, and ease of use. Conceptually, Kunkel mutagenesis and NEEL are similar in that both attempt to resynthesise the

plasmid from a ssDNA template through the same primer annealing, extension, and ligation steps. To repurpose Kunkel SDM for the LA-plasmid strategy, four Kunkel protocols were replicated and compared for the production of the ligation product.[176,179–181]

Of the four methods, only those from Kay[176], and Sidhu *et. al.*[181] gave appreciable ligated products, again confirmed through the exonuclease, and gyrase tests (Figure 2.11). Notably, all four methods employed T4 as the ligase, and T7 or Sequenase as the polymerase in the extension step (Table 2.1). T7 DNAP is frequently used for SDM due to its high polymerisation activity, high processivity, and lack of strand displacement.[182] It also has low discrimination against dideoxyribonucleoside triphosphates (ddNTP), the chain-terminating nucleotides in Sanger sequencing. Although it could incorporate ddNTP in the low presence of dNTP, it is not suitable for sequencing purposes due to its high 3'-5' exonuclease activity.[183] This becomes problematic when dNTP concentration falls as the reaction progresses, resulting in no net DNA synthesis, or the DNA strand becomes degraded. Sequenase is an engineered T7 DNAP whose exonuclease activity is abolished, finding its widespread use in dideoxy-sequencing.[183]

	Woodgate	Kay	Sidhu	Varshney
Reaction time/ mins	20	180	overnight	25 ºC for 5 mins
Temperature/ ºC	33	20	20	37 ºC for 30 mins
Polymerase	T7	T7	T7	Sequenase
[Polymerase]/ U/ µL	0.15	0.05	0.10	0.22
[Ligase]/ Weiss U/ µL	0.03	0.05	0.10	0.03

Table 2.1. Summary of the different Kunkel SDM reaction conditions to resynthesise the mV_nick plasmid

The major distinguishing features between the four methods are the extension and ligation conditions. Despite using a similar polymerase and ligase, Kay's and Sidhu's

protocols required an incubation temperature of 20 ˚C for 3 h or overnight, whereas Woodgate and Varshney *et. al.* incubated their reactions at 33 and 37 ˚C for 0.5 h. As T7 DNAP operates optimally at 37 ˚C, presumably the former methods were optimised for ligation, which necessitated lower temperatures and a prolonged reaction time (Section 2.3), whereas the latter focussed on extension. Thus, although all methods could affect the extended products, only those favouring ligation conditions gave the ligated plasmids as the major products. Having demonstrated Kunkel SDM could also be adapted to generate the amine plasmids (Figure 2.12A), Kay's method was employed hereafter as it required a shorter incubation period.

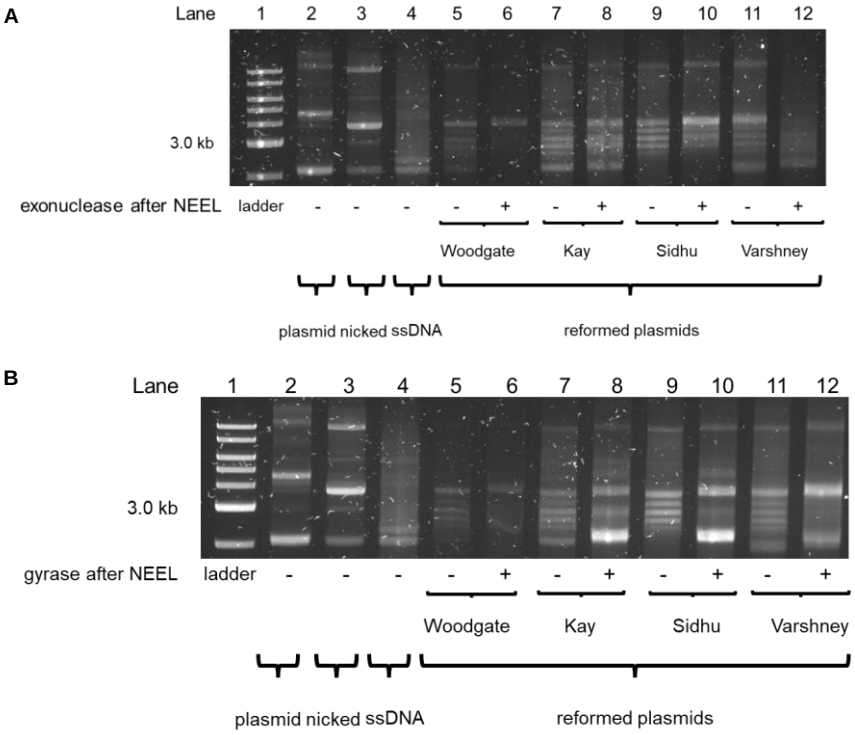


Figure 2.11. Comparison between different Kunkel protocols to resynthesise mV_nicks plasmid

Only methods from Kay, and Sidhu *et. al.* yielded the reformed plasmids as confirmed by **A.** exonuclease and **B.** gyrase treatments after NEEL.

Next, the identity of the dinucleotides at the ligation junction was studied. Sixteen primers with varying lengths, GC contents, annealing temperatures, and locations, representing all combinations of dinucleotides at the 5' and 3' positions where ligation occurred were tested for ligation efficiency (Table 6.1, Figure 2.12B). The reformed plasmids were afforded without discrimination with regard to the base identities at the junction, or the individual primer's features.

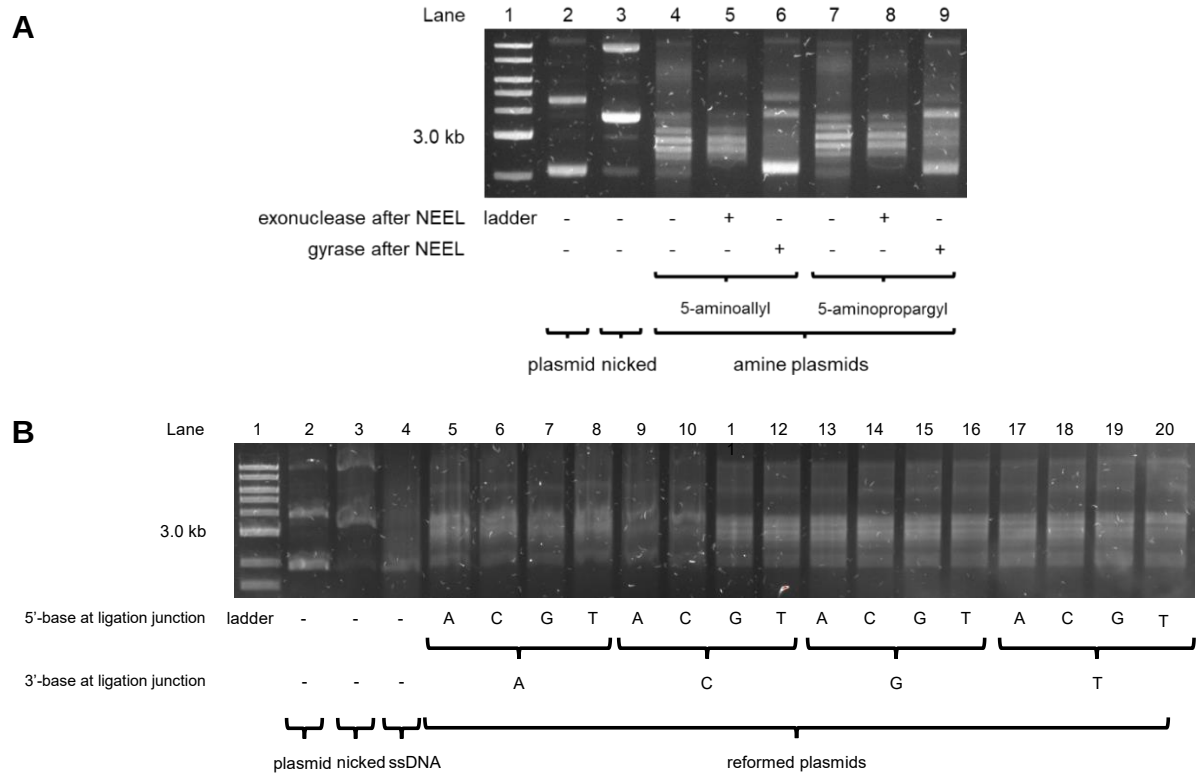


Figure 2.12. NEEL to incorporate amino bases into mV_nicks plasmid, and confirming dinucleotides at the ligation junction did not affect the ligation reaction

A. NEEL was performed in the exclusive presence of 5'-aa- or 5'-ap-dUTPs. B. Primers to generate different dinucleotides at the ligation junction were used for NEEL.

The final steps to deliver LA-plasmids were the reaction with PCB-NHS ester and streptavidin binding. The former reaction was facile and could be catalysed by sodium bicarbonate at room temperature. Regarding streptavidin, the ratio between wildtype tetravalent streptavidin (WTS) and biotin was explored. In the absence of dTTP in the extension reaction, it was assumed that every amino dUTP would be incorporated in

place of dTTP. For simplicity, in a stoichiometric WTS:biotin ratio, it was also assumed that each WTS was bound to a single biotin. This would be true for MVS, as each streptavidin could only bind to a single biotin. Without steric constraint, a 1:1 MVS:biotin ratio would give a homogeneous population of LA-plasmids (Figure 2.13, left).

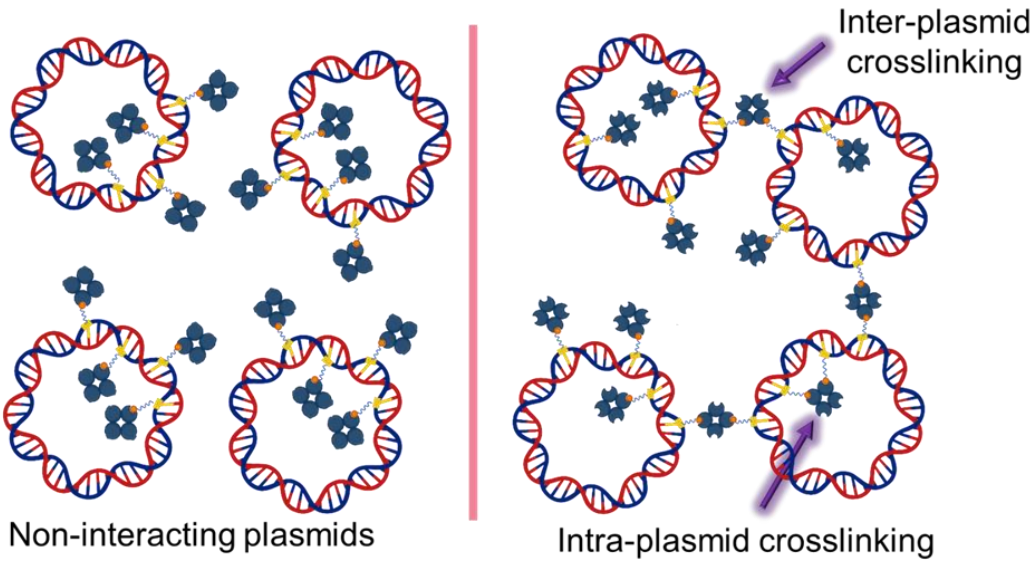


Figure 2.13. Monovalent vs. tetravalent streptavidins

MVS could only bind to a single biotin to give a more homogeneous non-interacting LA-plasmid population (left). WTS could give rise to inter- and intra-plasmid crosslinking to give heterogeneous LA-plasmid populations.

For WTS, the situation was more complex. Each WTS contained 4 active subunits, any number of which could bind to biotin. Thus, a single WTS molecule could bind to multiple biotins on the same plasmid (intra-plasmid crosslinking) and/or biotins on different plasmids (inter-plasmid crosslinking, Figure 2.13, right). This would give rise to a heterogeneous population of differently interacting plasmid aggregates. Indeed, PCB plasmids bound to WTS gave a large smear (Figure 2.14). UV irradiation (365 nm, max setting for 1 h) partially uncaged these aggregates and shifted this smear

toward the supercoiled plasmid. Without PCB, streptavidin could not bind the plasmids (lanes 7 in **A** and **B**). In contrast, the absence of reactive amine groups meant plasmids could not react with PCB-NHS ester, and hence could not bind streptavidin (lane 8 in **B**). Altogether, the LA-plasmid strategy was viable and required NEEL to install amine modifications. These amines acted as attachment points for the PCB bridges that ultimately anchored streptavidin.

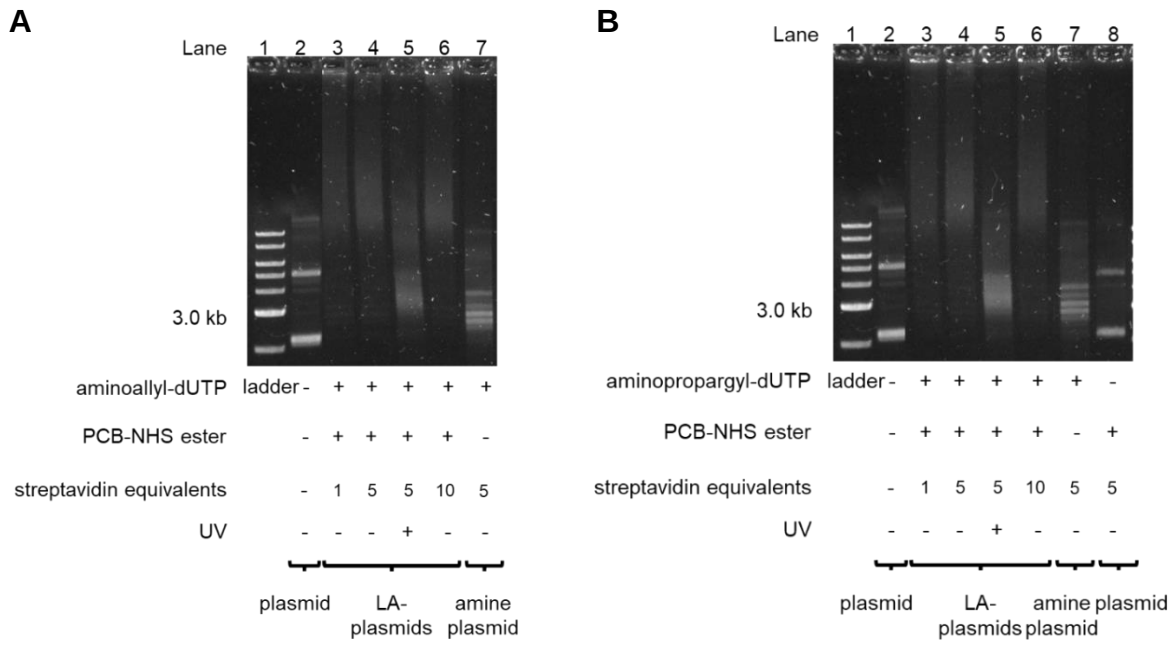


Figure 2.14. Uncaging mV_nicks LA-plasmids

WTS bound to PCB plasmids to give LA-plasmids. UV uncaging partially removed streptavidin. Negative controls without amines or reaction with PCB-NHS ester showed that streptavidin only bound when plasmids contained both amines and PCB. This was demonstrated for plasmids obtained from the extension with A. 5'-aa-dUTP and B. 5'-ap-dUTP.

Around this time, it was also determined that bands appeared differently on the gel depending on the time of staining. Thus far, gels had been post-stained, as this avoided any interference the dye might have on DNA migration and offered more uniform bands. The effect could be seen for the ladders in the pre- and post-stained gels (Figure 2.15). GelRed interaction with DNA in the pre-stained gel caused the

bands to slightly curve upwards, producing a ‘smile’ effect. In contrast, bands in post-stained gels ran more evenly.

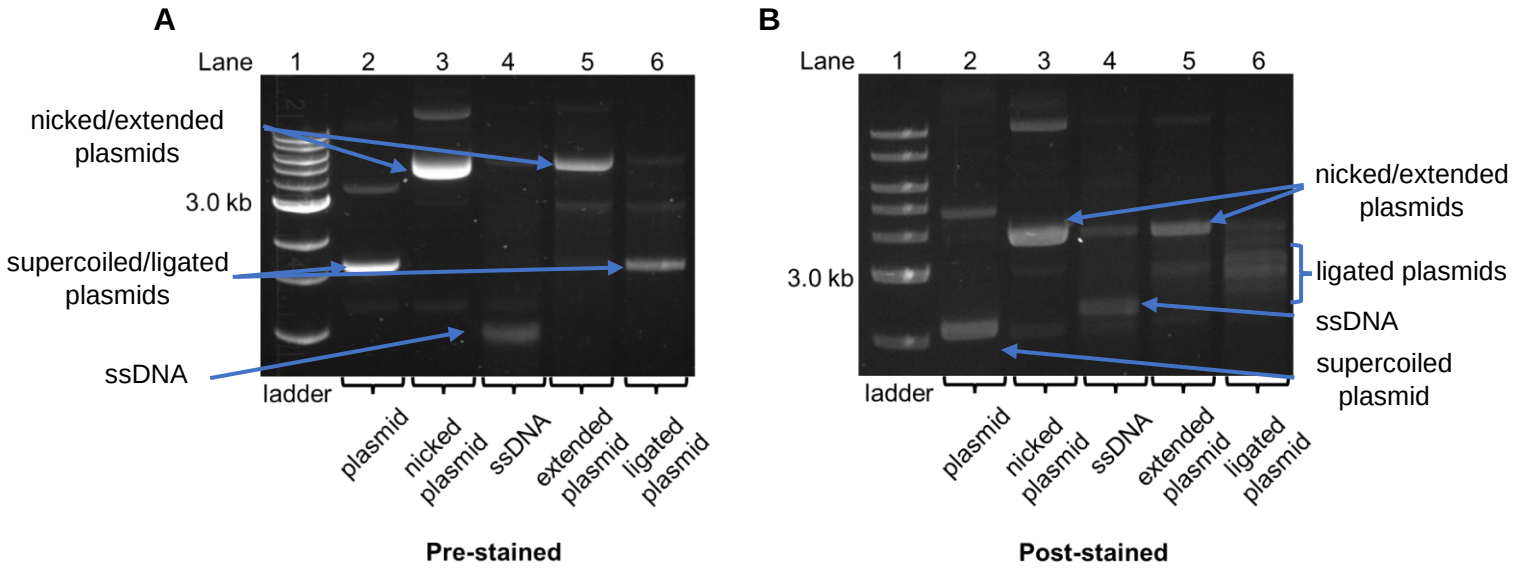


Figure 2.15. Comparing NEEL of the mV_nicks plasmid between the pre- and post-stained gels.

Plasmids were resynthesised and loaded on a A. pre-stained or B. post-stained gel.

In terms of band migrations, the most notable differences were the relative positions of the ssDNA compared to the supercoiled plasmid, and the disappearance of ladder bands after ligation in the pre-stained gel. As previously mentioned, these bands represented the different supercoiled plasmid conformations. When pre-stained, these bands converged into a single band that co-migrated with the supercoiled plasmid. Hypothetically, GelRed intercalation with the reformed plasmids could compact these species to a size similar to that of the actual supercoiled plasmid, giving them a similar appearance on gels. Unless otherwise stated, gels were pre-stained hereafter as bands were more intense and easier to distinguish.

2.7 Removal of Unreacted Plasmids

With a crude method to synthesise LA-plasmids, each step had to be fine-tuned. It was evident from previous figures that the nicking reaction was incomplete, e.g. in Figure 2.15B, the supercoiled plasmid was still present in lane 3. Unreacted native and nicked plasmids would not be caged and would contribute towards leakiness. There were several possibilities to address this issue: optimising the nicking reaction to ensure completion, the ssDNA could be purified from the unreacted plasmid, or degradation of the unreacted plasmid whilst leaving other plasmids intact.

2.7.1 Optimising Nicking Reaction

Nt.BspQI was a salt-sensitive nickase, and this was a facile first step for optimisation. After ethanol precipitation to remove salt contaminants, purified DNA was nicked at increasing DNA concentrations (Figure 2.16). The DNA concentration did not impact the nicking efficiency as the conditions appeared identical. At first glance, the reaction appeared to finish when high-purity DNA was used. However, when more reaction was loaded on the gel, it was evident there was still unreacted native plasmid.

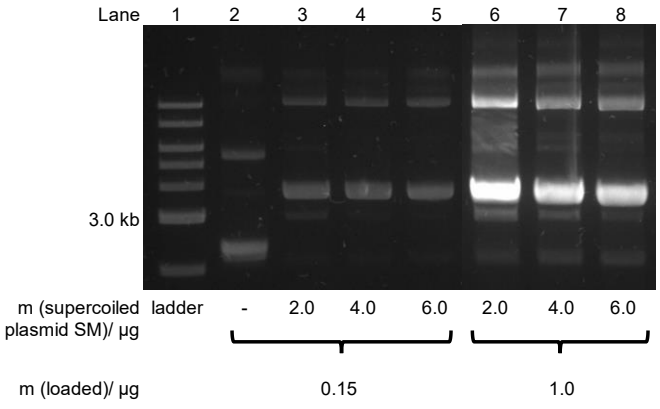


Figure 2.16. Pure template enhanced nicking efficiency but the reaction was still incomplete

Ethanol-precipitated mV_nicks plasmid was nicked at different DNA concentrations and either 0.15 or 1.0 µg of each reaction was analysed on an agarose gel.

The reaction likely reached a plateau when the majority of starting material had been converted into product. At this stage, the likelihood of enzyme-substrate interacting was low, explaining the identical product profile across different starting material concentrations. As the nicking reaction was unlikely to progress beyond this plateau, attention was turned to separating the ssDNA from unreacted plasmids.

2.7.2 ssDNA Purification

After the exonuclease reaction, the solution contained predominantly the desired ssDNA, alongside small amounts of the native plasmid. The only way to separate the ssDNA from these leftover materials was through extraction from the gel. Under electrophoretic conditions, the different species could be distinguished through their differences in size, shape, and charge. Resolved ssDNA bands could be excised from the gel, and the DNA separated from the agarose pieces.

Gel extraction is frequently used in restriction-ligation cloning to purify the insert and vector PCR products from unspecific amplified products. The yield and purity of the recovered DNA are usually low but are satisfactory for cloning purposes. Repeated gel extraction attempts with modifications aimed to address these issues still gave yields <20%, and the low 260/230 purity ratio prohibited downstream applications.

Ethylenediaminetetraacetic acid (EDTA) found in tris-acetate-EDTA (TAE) buffer used to prepare agarose gel could chelate divalent magnesium ions which were required for T7 DNAP and T4 Ligase to function. TAE was replaced with ammonium acetate and the ssDNA bands were purified using the ‘crush and soak’ method.[184] In this method, the gel slabs containing the DNA were crushed with a pestle and incubated overnight in the crush and soak buffer to extract the DNA. The buffer was then collected and re-purified. Small agarose pieces were often carried over which if not

removed properly, could inhibit NEEL altogether (Figure 2.17). Even when these agarose contaminants were removed, the reaction was still inefficient.

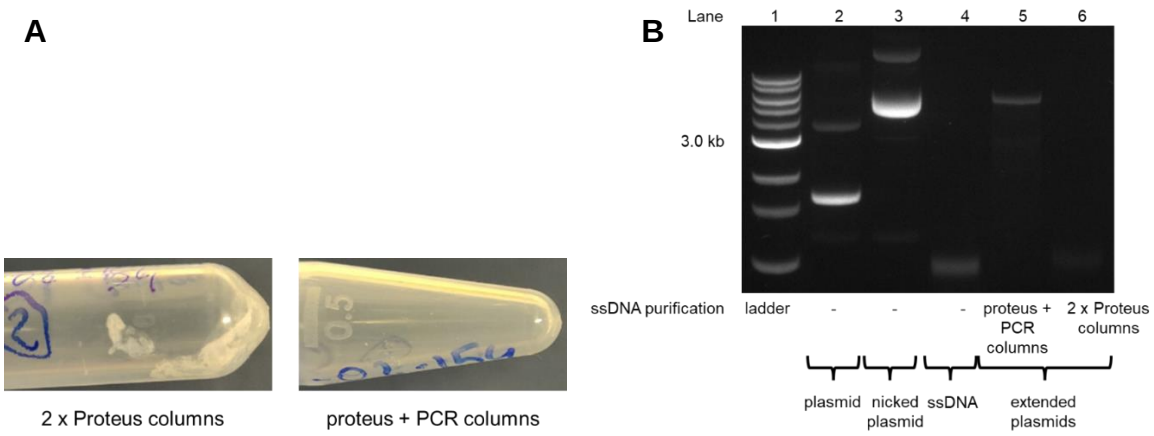


Figure 2.17. Extension using mV_nicks ssDNA obtained from the 'crush and soak' method

A. Crush and soak carried over agarose pieces which necessitated further purification. B. The low yield and contamination inhibited downstream reactions.

Contaminants were derived from the agarose gel slab. To overcome this, a method was devised to avoid DNA extraction from the gel. The DNA mixtures were allowed to run on an agarose gel as normal until the different species had separated (Figure 2.18A). The current was stopped, and the gel slab immediately in front of the ssDNA band was removed to create a well. The well was loaded with a minimal volume of buffer, and the current was resumed momentarily to allow the ssDNA to run into the well. The buffer was then removed and further purified.

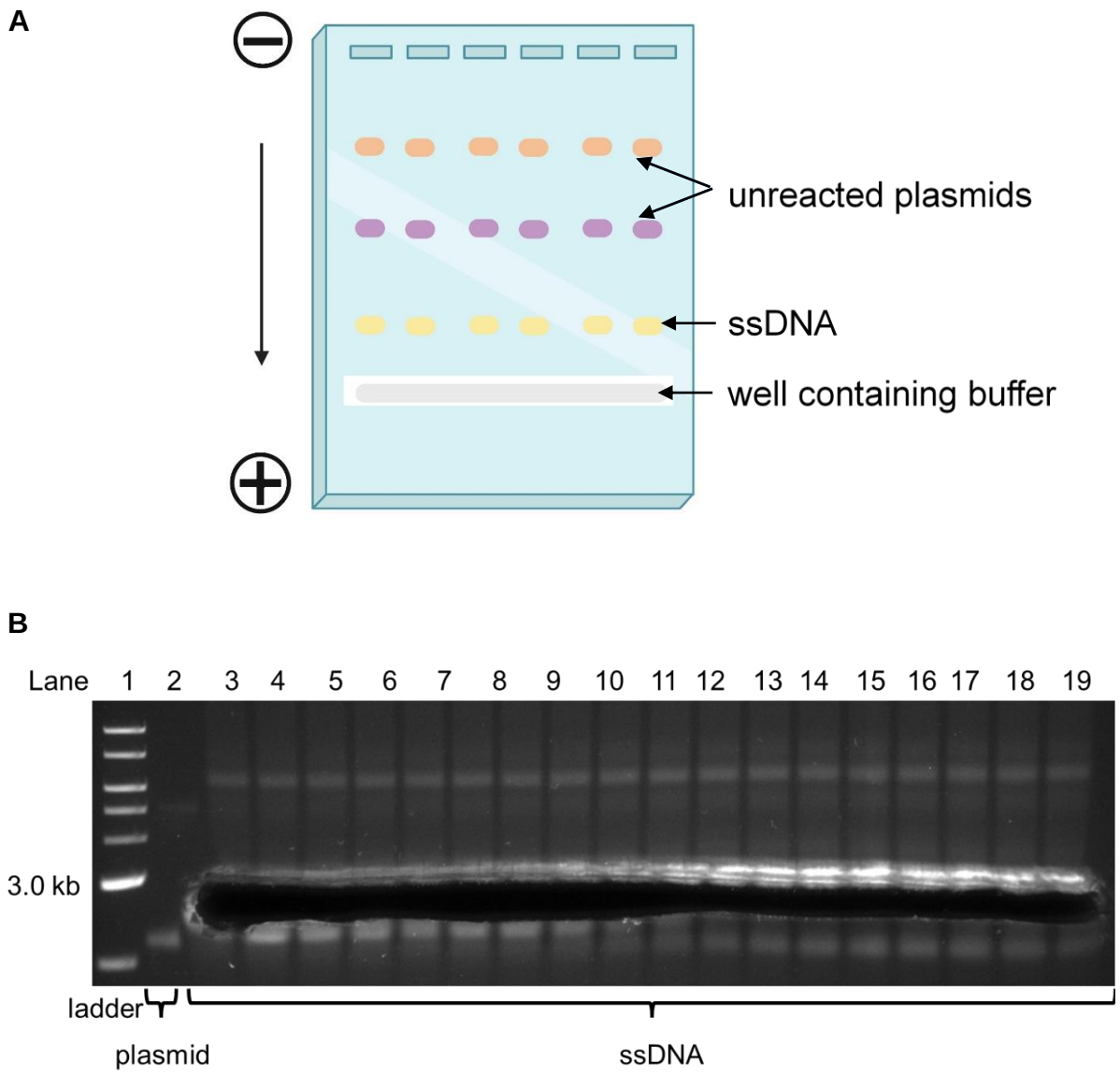


Figure 2.18. Isolating mV_nicks ssDNA without extraction from the gel

A. Schematic illustrating how ssDNA could be separated by electroelution into a well containing buffer. B. After ssDNA had been recovered, leftover ssDNA could still be seen.

In reality, this method was not so straightforward. The DNA often overran past the well so that the current had to be reverted, only for the DNA to overrun in the other direction. Often many back-and-forth rounds of currents were applied before the bands ended inside the well. Furthermore, not all ssDNA could be recovered as the bands could migrate non-uniformly (Figure 2.18B). The recovered material still required further

purification, and the final yield and purity were not better than conventional gel extraction.

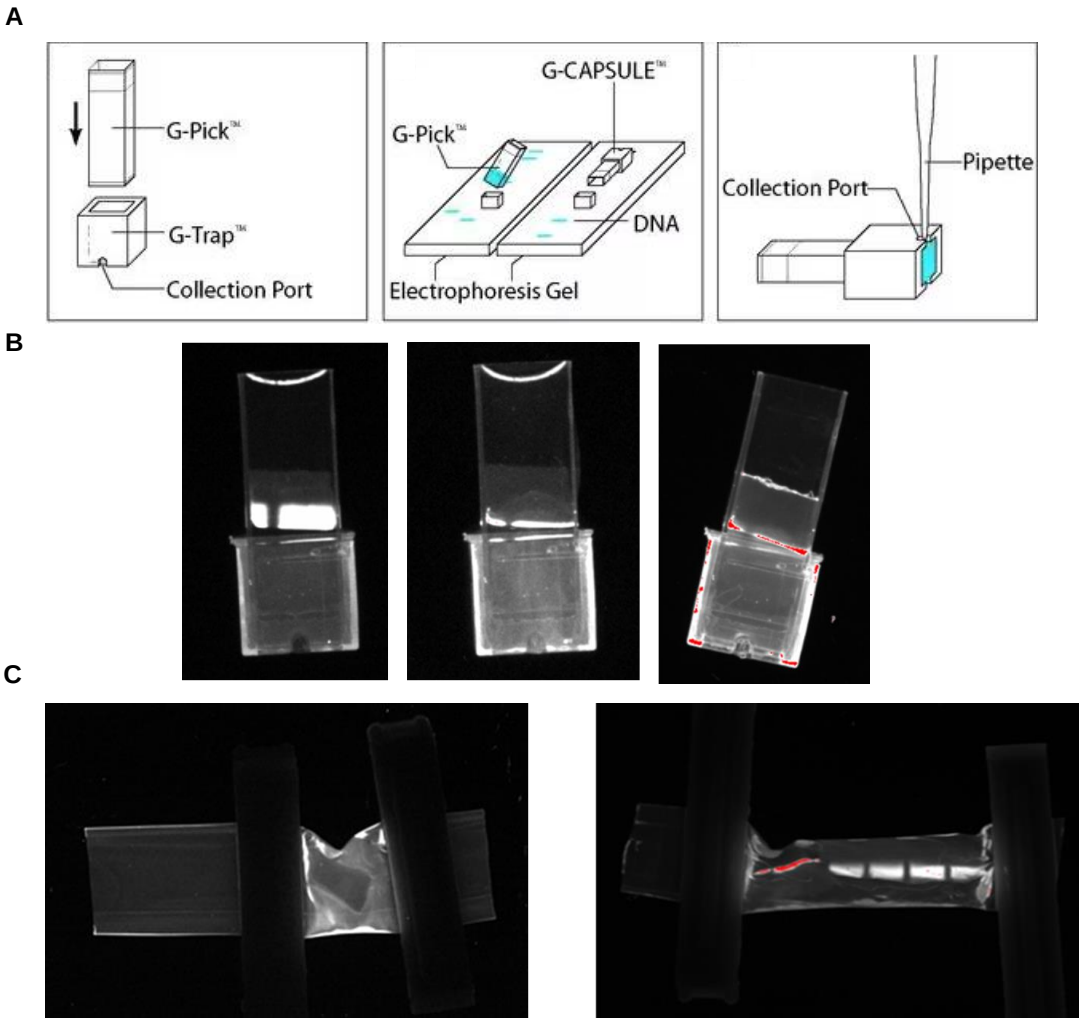


Figure 2.19. Electroelution of mV_nicks ssDNA using the G-Capsule or a dialysis membrane

A. Schematic illustrating how G-Capsule functioned. Figure adapted from G-Biosciences. B. DNA electroeluted with a G-Capsule and C. a dialysis membrane. When the method worked, DNA was visible in the solution surrounding the gel slab (left). Sometimes, DNA remained within gel slabs despite the applied current (right).

Two related approaches also required the electroelution of the DNA bands into the buffer. The first involved a G-Caspule™, which consisted of 2 parts: G-Pick™ and G-Trap™. The pick was used to excise the gel bands. It was then attached to the trap,

which was pre-filled with buffer. The current was then applied to enable the DNA to move into the trap whilst the gel slab remained in the capsule (Figure 2.19A and 2.19B). The second method was similar but instead used a dialysis membrane. The G-Capsule was not a viable method for the purification of large amounts of DNA, as each band had to be electroeluted individually. On the other hand, the dialysis membrane was unreliable. When it worked, the DNA could be seen in the solution, but oftentimes the DNA was stuck within the gel slabs (Figure 2.19C, left vs right). Notwithstanding, both methods also suffered from the same problems relating to yield and purity. Clearly, purification from the gel was not a viable method to eliminate native plasmids. Therefore, a means to preferentially degrade these unreacted plasmids was sought.

2.7.3 DpnI Degradation of Methylated DNA

DpnI is an endonuclease that recognises and cleaves methylated GATC sequences.[185] Bacteria methylate their DNA as a means to regulate DNA-protein interactions, specifically to control DNA expression and degradation.[185] Miniprepped plasmid DNA is fully methylated and can act as a substrate for DpnI. Indeed, the enzyme is typically used after PCR to degrade the plasmid template whilst leaving PCR products intact, as the latter are synthesised *in vitro* and therefore are unmethylated. In cloning, this minimises false positive colonies that may have taken up the original plasmid.

NEEL plasmids were hemi-methylated as they contained one methylated parental strand and an unmethylated resynthesised strand. Although DpnI could cleave hemi-methylated DNA, the cleavage activity is much slower.[186] This bias towards fully

methylated DNA as substrate could be leveraged to selectively degrade the leftover native plasmid, whilst leaving the majority of NEEL products intact.

To minimise the degradation of NEEL plasmids, mild conditions were desired. Native mV plasmid was digested with decreasing time and enzyme concentration (Figure 2.20). The manufacturer's protocol (0.4 U μL^{-1} for 60 mins) fully cleaved the plasmid. Higher-order structures appeared and became more intense as the time and/or enzyme concentration decreased further. These could represent concatamers of degraded DNA, which formed more readily when there was less enzyme to degrade them. Nevertheless, the supercoiled plasmid was absent at a DpnI concentration as low as 0.02 U μL^{-1}. For subsequent experiments, 0.06 U μL^{-1} DpnI and 5 mins were used.

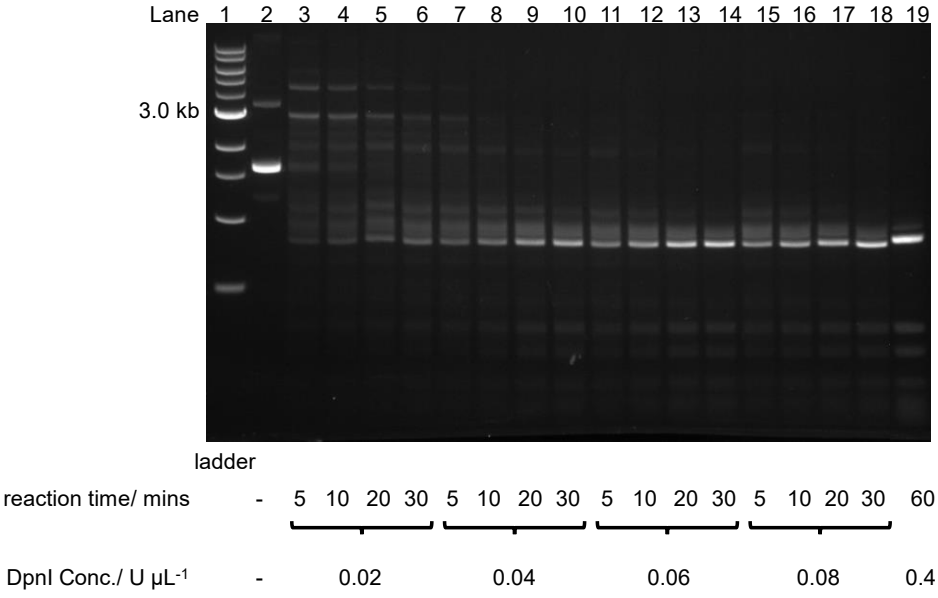


Figure 2.20. Optimising DpnI reaction conditions by varying time and DpnI concentration to remove unreacted mV_nicks plasmids

The reaction was completed at 0.02 U/µL DpnI, but milder conditions gave concatamers of degraded products.

Supercoiled and PCB plasmids appeared at the same position on the gel, hence PCB plasmid could mask the supercoiled plasmid. To determine if DpnI had removed methylated supercoiled plasmids, these plasmids had to be visually distinct from the other plasmid types. As streptavidin only bound to PCB plasmids to produce bands with slower mobilities on gels, and any supercoiled plasmids would not bind and remain at the same position, their presence could be detected.

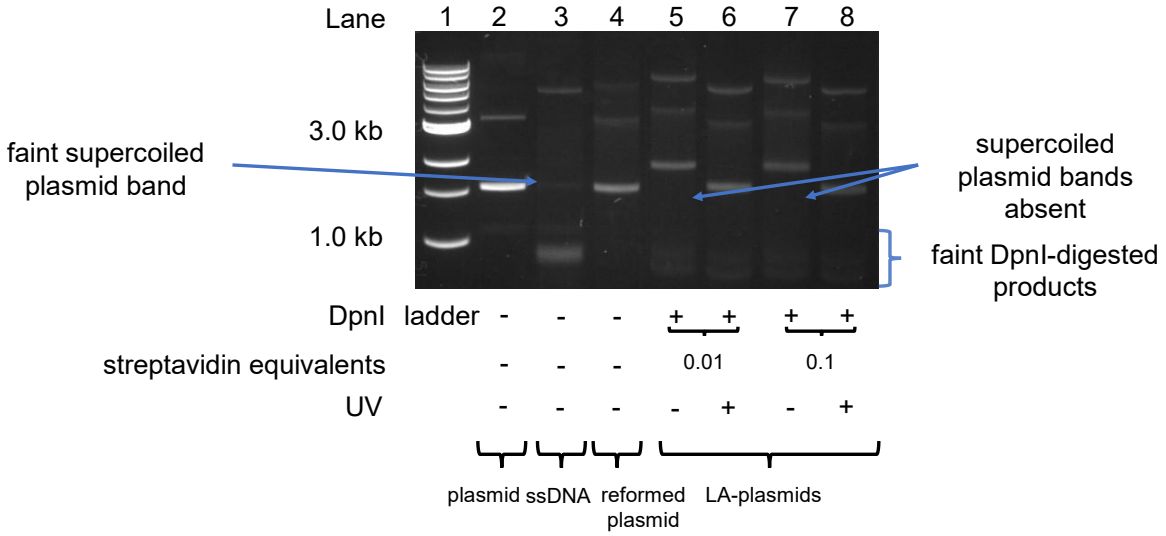


Figure 2.21. DpnI removed unreacted mV_nicks supercoiled plasmid

Shifted LA-plasmid bands showed that all unreacted supercoiled plasmid had been degraded. UV uncaging restored band positions. Plasmids were re-synthesised with the regular triphosphates and 5'-ap-dUTP and without dTTP presence.

The amine plasmids were reacted with DpnI followed by PCB-NHS ester and bound to streptavidin. This time, MVS was used to ensure a homogeneous LA-plasmid population (Section 2.6). This would generate a single band, instead of a smear, so any unreacted supercoiled plasmid could be easily visualised. Before DpnI, the leftover supercoiled plasmid could be observed as a very faint band after the exonuclease reaction (Figure 2.21, lane 3). Expectantly, the MVS LA-plasmids band was gel-shifted, and nothing was detected at its original position, indicating

supercoiled plasmid had been degraded. Only a small amount of supercoiled plasmid was left over from an incomplete nicking reaction, so the cleaved product bands were expected to be dim. These were faintly visible for all conditions that had been DpnI-treated and would have comprised degraded supercoiled plasmid as well as a small amount of degraded amine/PCB plasmids. Unlike Figure 2.14, where uncaging only partially removed PCB, here UV irradiation of MVS LA-plasmids removed most if not all PCB, evident by the band positions.

Overall, the NEEL method that made use of 4 enzymatic steps (nicking, exonuclease digestion, extension, and ligation) to resynthesise plasmid was developed. By repurposing Kunkel's SDM protocol, and using a more specific nickase and T7 DNAP, the mV_nicks plasmid was successfully reformed. The presence of amino dUTP in the extension reaction allowed its incorporation into the plasmid. The reaction with PCB-NHS ester followed by streptavidin binding gave LA-plasmids. Streptavidin was removed by UV-induced cleavage of the photoresponsive group. Finally, to ensure that all plasmid was caged with streptavidin, any leftover unreacted plasmid was removed through their preferential reaction with DpnI.

3

Regulating Bacterial and Phage Promoters

3.1 Introduction

3.1.1 Transcription Regulation within Bacteria

Bacterial transcription is a highly regulated process.[187] Initially, RNAP associates with a σ factor to form the holoenzyme. The σ factor enables the binding of the holoenzyme to the promoter sequence. Bacteria contain multiple different σ factors, and each is capable of directing the holoenzyme to a subset of promoters with varying sequences.[187] Upon binding, the DNA strands around the transcription start site (TSS, from -10 to +2) are unwound. The holoenzyme uses the antisense strand as the

template for transcription. The resulting transcript will have the same base sequence as the sense strand, except U replaces T. Transcription initiation begins with the formation of a phosphodiester bond between the initiating and adjacent nucleoside triphosphate (NTP). Chain elongation follows until the holoenzyme encounters the transcription terminator site or reaches the end of the DNA template.

The two most important regulatory sequences are the hexameric -35 and -10 regions, which are found 35 and 10 base pairs upstream of the TSS, respectively. The consensus sequences at these regions are known, and promoters bearing the consensus sequences enable more effective transcription, whereas deviations from the consensus sequences generally lead to fewer transcription events.[188] The amount of RNAP and σ factors within bacteria are limited. To transcribe the 4,000-5000 genes within bacteria, different promoters compete for the binding of the holoenzyme, which results in unequal distribution of different gene products.[189]

There are multiple mechanisms in play to partition the transcriptional machinery between the competing promoters. Within the context of photocaging, the -35 and -10 sequences are the most relevant feature to exert control, and also the easiest to modify, primarily due to their important regulatory roles and proximity to one another. By aligning 112 *E. coli* promoter sequences, Hawley and McClure established the identities for the consensus sequence: 5'-TTAACA-3' and 5'-TATAAT-3' for the -35 and -10 regions, with the five-prime most thymine located at -36 and -13 positions, respectively.[188]

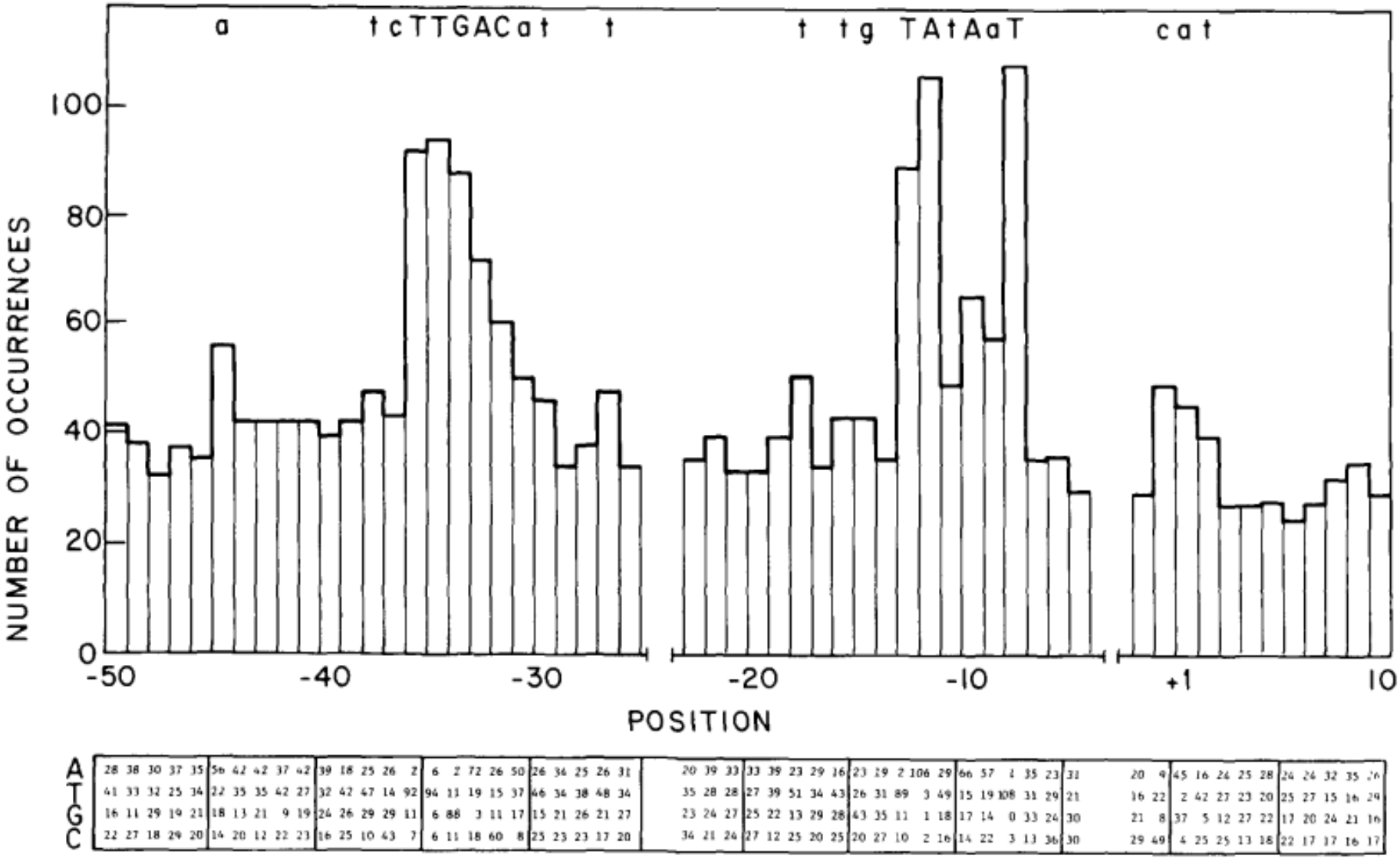


Figure 3.1. Histogram showing distribution of each base within the -50 to +10 region of the *E. coli* promoter

The number of appearances of each base at each position were tabulated and summarised below the histogram. Bases with at least 39% occurrence were listed above the columns; capitalised bases were >54% conserved. Figure adapted from reference 186 with permission from the Oxford University Press.[186]

Figure 3.1 shows the distribution of each base between the -50 to +10 sequence for the promoters under investigation.[188] TTGAC is highly conserved at the -35 region, with a ~10 base stretch of A-T rich sequence further upstream. Within the -10 region, four out of the 6 bases are strongly conserved. Further downstream, a thymine is also frequently found at the +2 position. Mutagenesis of the conserved bases within these regions typically led to transcription downregulation.[188] For our purposes, highly conserved thymines at these regions are desirable, as they can be replaced with the amino dUTPs to allow photocaging.

3.1.2 Tac Promoter

The tac promoter is a synthetic hybrid promoter developed by Vasser *et. al.*[190] It is designed to be an inducible promoter with better expression from its parent promoters, the bacterial lac UV5 and trp promoters. The lac UV5 promoter possesses a consensus sequence at the -10 region but not the -35 region whereas it is the opposite case for the trp promoter. The tac promoter was created by fusing the sequence upstream of the -20 position from the trp promoter with the sequence downstream of the -16 position from the lac UV5 promoter.[190] Thus, the tac promoter has complete homology at the determinant -35 and -10 regions and expressed 11.8 and 3.5 times as strongly as the parental lac UV5 and trp promoters, respectively.[190] Analogous to the lac promoter, the tac promoter could be repressed by the lac repressor and re-activated by IPTG. The improved expression and ability to regulate expression make this promoter highly popular for the regulated overproduction of proteins in *E. coli* chassis.[191]

3.1.3 T7 Promoter

The T7 promoter forms the most widely used expression system for recombinant proteins in *E. coli* and is also commonly found in CFPS systems, making it one of the most important promoters within synthetic biology.[192] Like the tac promoter, its activity could be regulated by IPTG when the expression vector is transformed into an *E. coli* strain harbouring a lac-regulated T7 RNAP cassette.[192]

Transcription from phage promoters such as T7 is relatively simple compared to *E. coli* promoters and can be performed straightforwardly with the T7 RNAP without additional σ factors or transcription factors.[193] A comparison of 17 T7 promoters revealed a highly conserved 23 base pair sequence at the -17 to +6 positions. This sequence comprised the binding (-12 to -5) and initiation domains (-4 to +5).[194] Promoter activity was still retained after deletion of the -13 to -17 region, albeit suboptimal.[195] Mutagenesis of the bases within the -22 to +6 sequence indicated 4 bases at the -11 to -7 region that are indispensable for promoter activity. Of interest is the thymine at -8, which abolished activity if replaced with an adenine.[196]

Durbin *et. al.* determined that the glutamine at residue 758 of T7 RNAP made important contacts with this thymine and the surrounding bases that are crucial for promoter binding.[197] In particular, the replacement of glutamine by other amino acids decreased or abrogated expression from the T7 promoter. On the other hand, wild-type RNAP had no activity when this thymine was changed to any of the other 3 bases. The identification of this crucial thymine, along with 6 other thymines within the consensus sequence provides ample opportunities to cage this promoter.

3.1.4 Sp6 Promoter

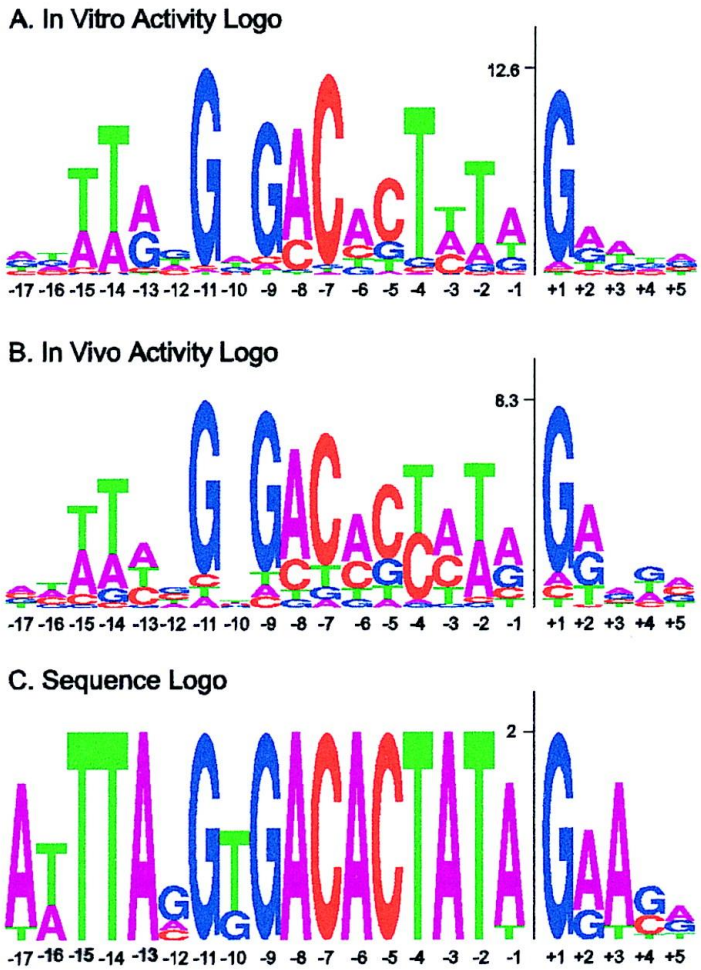


Figure 3.2. Logos summarising the relative contribution of each base within the -17 to +5 region of the Sp6 promoter

The contribution of each base to the promoter's activity at each site was tested A. in vitro and B. in vivo and graphically summarised. C. Relative frequency of each base as determined by aligning 11 Sp6 promoter sequences. For A-C, larger heights indicate higher importance. Figure adapted from reference 198 with permission from the National Academy of Sciences.[198]

The phage Sp6 promoter shares many characteristics with the T7 promoter: the high specificity and affinity between the Sp6 RNAP and its promoter sequence, the simple reaction components,[199] and the ability to regulate Sp6-encoded gene expression[200] also make this a promoter of choice for *in vitro* and *in vivo* protein production. Kang and co-workers constructed a mutant library for every possible single-base substitution

within the known Sp6 consensus sequence and assessed their *in vivo* and *in vitro* activities (Figure 3.2).[198]

Analogous to the T7 promoter, the Sp6 promoter could be divided into an RNAP binding (-17 to -5) and initiation region (-4 to +3). Changes to positions -17, -16, -12, and -10 within the former did not correlate with any change in activity.[198] The 4 most important positions at -11, -9, -8, and -7 were intolerant to mutations, supporting the consensus sequence. Of interest to our caging system were the thymines at -15, -14, -4, and -2, which reduced promoter activity to <30% when replaced with the other 2 or 3 bases.[198]

3.2 Chapter Overview

This chapter applied the caging strategy developed in Chapter 2 to regulate the expression of mV from the T7 promoter. The overarching goal was to maximise the ON/OFF ratio by examining light-dependent expression in cell-free transcription and translation. The presence of excess photocaging groups across the plasmid was detrimental to the photomodulation efficiency. By performing NEEL with a purified PCB oligo, both the number and location of modifications were pre-defined, which permitted tight photoregulation from LA-plasmids. The caging strategy was deployed to control the tac and Sp6 promoters, but more fine-tuning is needed before this goal can be realised.

3.3 Optimising IVT

With a method to install amines across the plasmid, it was necessary to fine-tune the number of modifications. Thus far, re-synthesised plasmids exclusively contained amino-modified thymines in place of the regular thymine nucleobases. As there were

666 thymines on the sense strand, this also equated to 666 amino-modified thymines. It would not be feasible to remove so many modifications (Figure 2.14), or harsh UV conditions would be required that would be damaging to biological systems. Additionally, in the case of WTS, extensive inter-plasmid crosslinking would form aggregates which would prevent gene delivery.

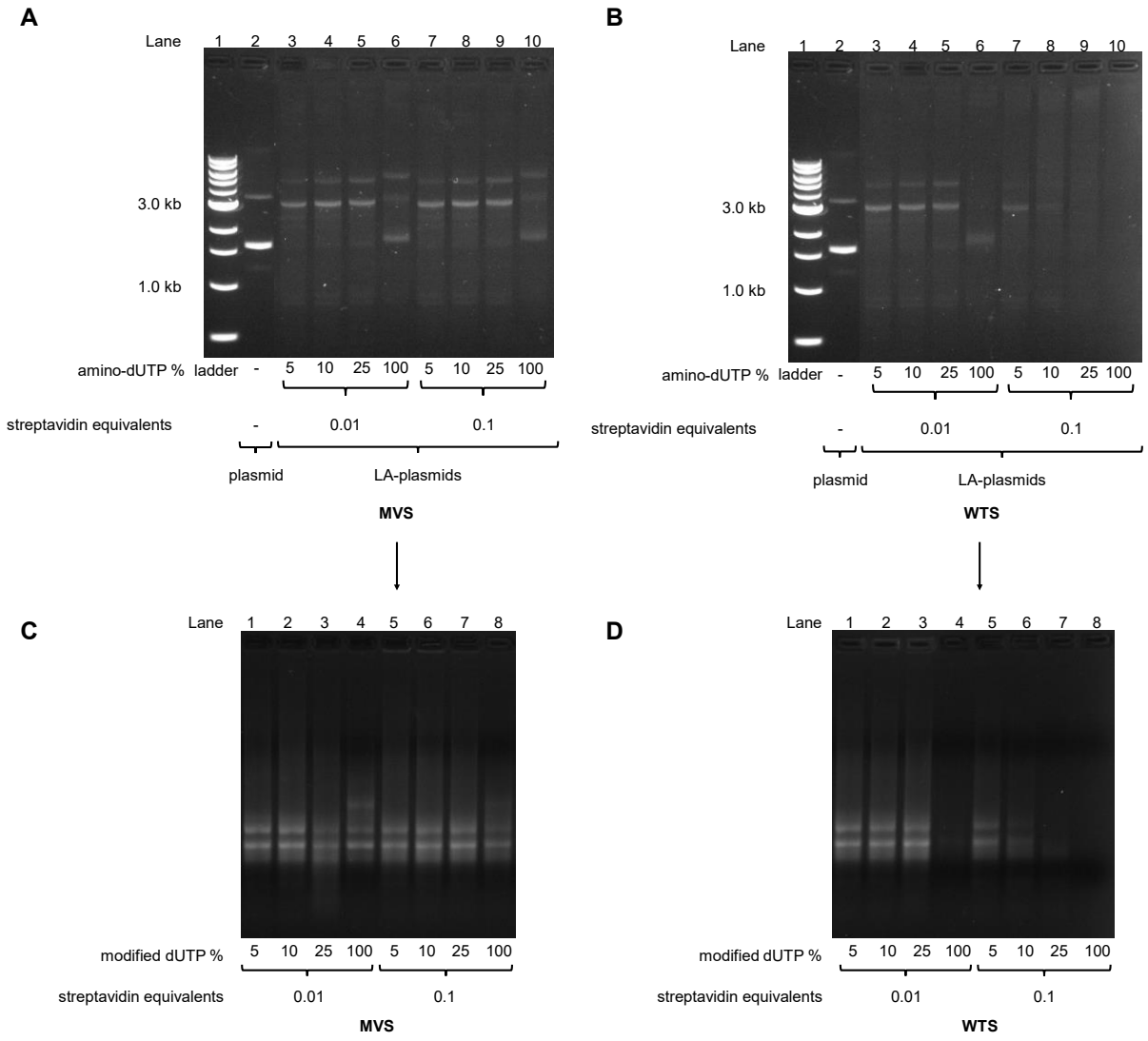


Figure 3.3. Minimising amino modifications in mV_nicks LA-plasmids

Varying amounts of 5'-ap-dUTPs were used in the extension reaction. The PCB plasmids were then bound to different equivalents of **A.** monovalent streptavidin and **B.** tetravalent streptavidin. **C-D.** IVT products from the plasmids obtained in **A** and **B**, respectively.

To minimise the amino modifications, different ratios of dUTP and dTTP were added to the extension-ligation reactions. It was assumed that T7 DNAP exhibited no discrimination for the regular versus modified triphosphate, and the statistical distribution would be maintained between the dUTP:dTTP ratio in the reaction, and the ratio on the resultant amine plasmids. For instance, in a 5% amino dUTP extension reaction, each plasmid should contain ~33 amines. Moreover, the streptavidin equivalents were also lowered, as it had been demonstrated that streptavidin binding and removal were effective with as low as 0.01 streptavidin equivalent (Figure 2.21).

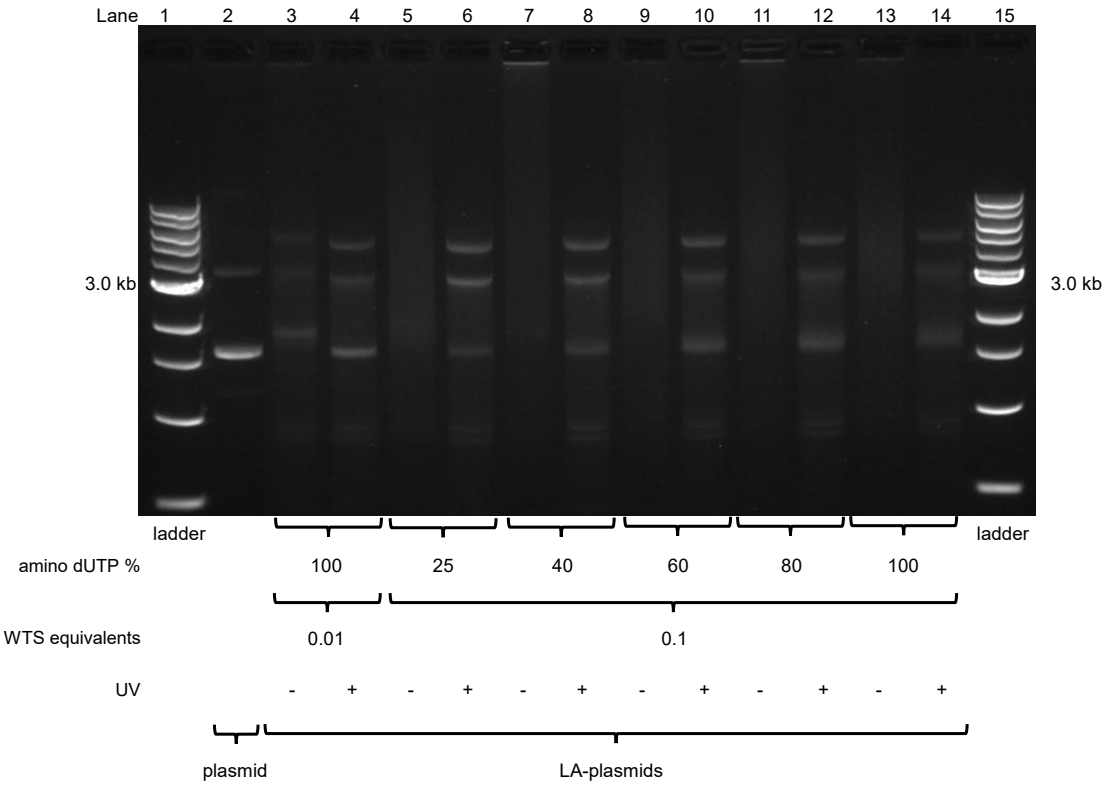


Figure 3.4. Effect of lowering WTS on uncaging mV_nicks LA-plasmids

Uncaging WTS LA-plasmids containing fewer 5'-ap-dUTP modifications and WTS equivalents led to the formation of distinct bands.

Under these conditions, only 0.1 WTS produced smears at above 10% 5'-ap-dUTP modifications (Figure 3.3A and 3.3B). When these plasmids were used as IVT templates, a higher presence of WTS led to decreasing levels of transcripts, whereas

none of the MVS LA-plasmids could suppress IVT (Figure 3.3C and 3.3D). For MVS, the number of streptavidin was too few and transcription was unaffected. Despite possessing a similar number of streptavidin, WTS' ability to crosslink either stalled RNAP and/or precluded its interaction with the T7 promoter.

Unlike the smear observed in Figure 2.14, uncaging WTS LA-plasmids with fewer modifications produced distinct bands (Figure 3.4). However, it was clear that increasing the modifications produced gradually retarded band mobilities and increased smeariness. Unless otherwise stated, subsequent experiments employed 25% 5'-ap-dUTP and 0.1 WTS equivalent.

3.4 Optimising UV Irradiation

Thus far, LA-plasmids were irradiated at the max UV setting for 1 h. These conditions would be too damaging to be biologically useful. After rigorous testing with different wavelengths, lamp distance, irradiation time, and intensity, a collimated lamp illuminating 365 nm wavelength with an irradiance of 2.5 (for 15 mins) or 7.8 mW cm^{-2} (for 5 mins) could efficiently cleave PCB (Figure 3.5A and 3.5B). *In vitro* transcription/translation (IVTT) reactions with the native plasmid were assembled and irradiated with these UV settings. In these initial experiments, only a single repeat was performed for each condition. This was because the reactions involved in LA-plasmid synthesis were not fully optimised for high yields, and many conditions were assessed simultaneously. Furthermore, at this stage, it was desirable to determine qualitative answers to quickly prototype different reaction conditions, rather than absolute quantitative values.

The reporter protein expression was subsequently measured using a fluorescent plate reader (Figure 3.5C). There was a clear decrease in protein expression as the intensity

and time were increased. The higher irradiance was markedly harmful, and an intensity of 2.5 mW cm^{-2} was chosen for further optimisation.

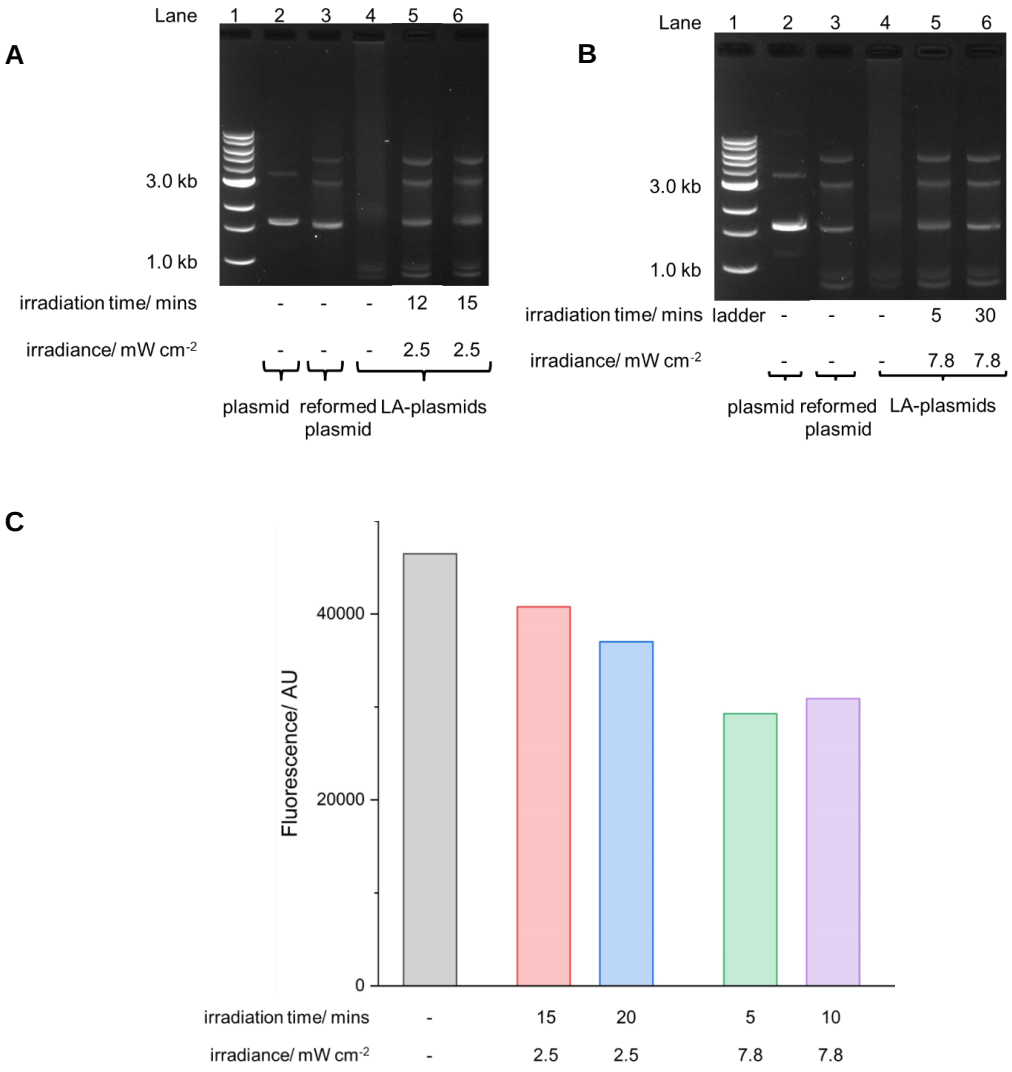


Figure 3.5. Optimising irradiation conditions to uncage mV_nicks LA-plasmids

Plasmids were re-synthesised with the regular triphosphates and 1:3 ratio of 5'-ap-dUTP:dTTP, reacted with PCB-NHS ester then bound to WTS. LA-plasmids were uncaged using the following irradiance. A. 2.5 mW cm^{-2} and B. 7.8 mW cm^{-2} with the indicated durations. C. IVTT reactions using native plasmid as the template were irradiated with varying UV irradiance and time (n = 1).

3.5 Optimising IVTT ON State

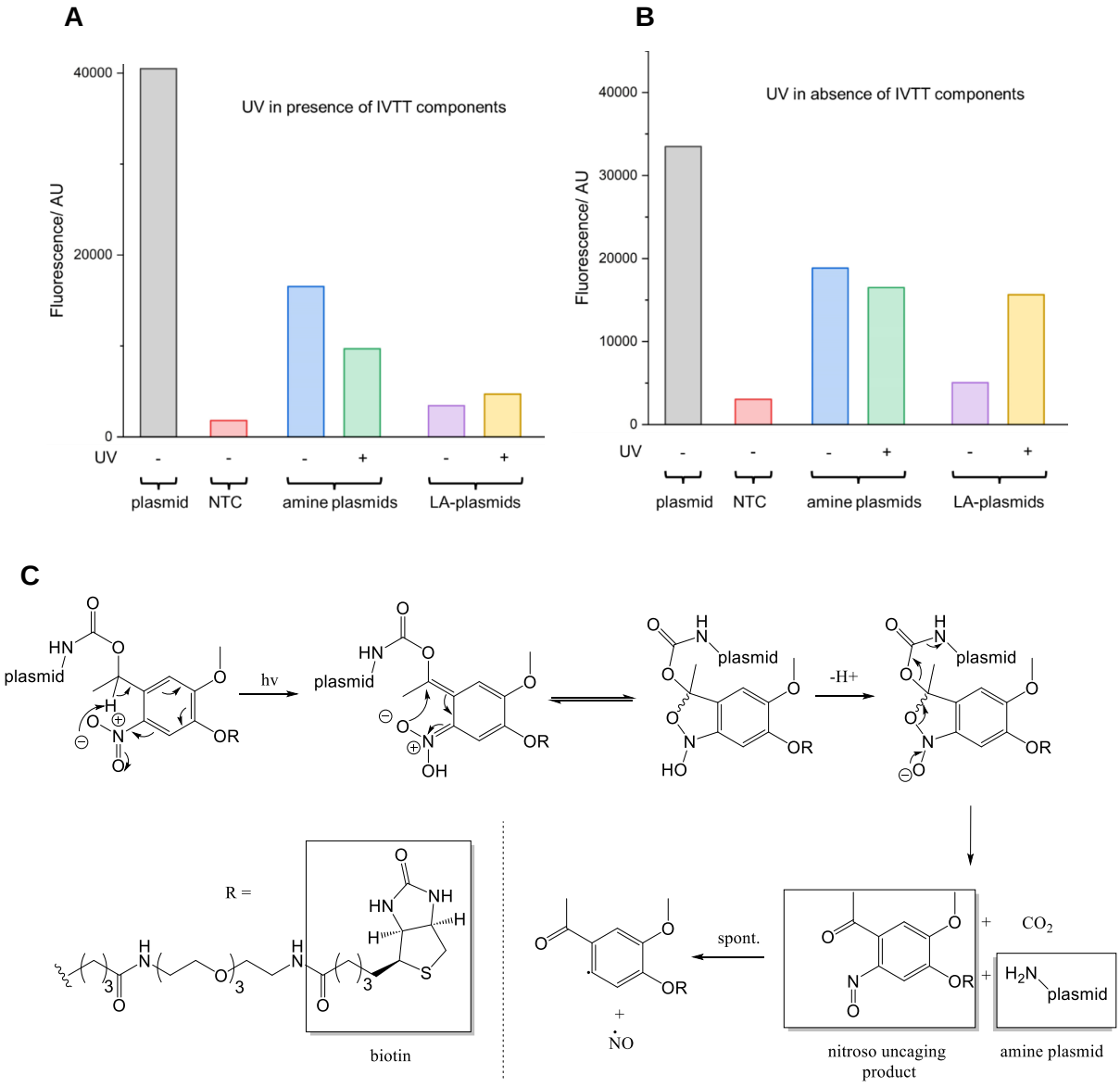


Figure 3.6. Uncaging mechanism and its effect on IVTT components for the mV_nicks plasmids

Plasmids were re-synthesised with the regular triphosphates and 1:3 ratio of 5'-ap-dUTP:dTTP, reacted with PCB-NHS ester then bound to WTS. UV was applied either A. in the presence or B. absence of IVTT components and fluorescent reporter expression was measured after a 3 h reaction. C. 2-nitrobenzyl UV-uncaging mechanism and nitroso-compound decomposition to give nitric oxide free radical (*n* = 1).

IVTT reactions with the modified plasmids were irradiated using these UV conditions, and their expression was compared against the native plasmid. When only amines

were incorporated, fluorescence decreased to half that of the native plasmid (Figure 3.6A). This suggested the presence of amine modifications alone could affect expression; UV slightly dampened expression further. The 'amine plasmid +UV' condition acted as the best-case scenario, as removal of PCB groups from LA-plasmids would reveal the amine bases. Thus, the uncaged LA-plasmids resembled the amine plasmids. Although caged LA-plasmids achieved a promising OFF relative to the no-template control (NTC), uncaged LA-plasmids did not perform much better, and their fluorescence was much lower than the amine plasmid control condition.

Uncaging led to the release of a nitroso compound (R-N=O). These compounds have a low C-N bond dissociation energy and are unstable. They readily decompose to give the more stable nitric oxide (NO) free radical by-product (Figure 3.6C).[201] These species are incredibly reactive and have a very short lifetime.[202] It was thought that NO radicals and/or UV interaction with the IVTT components, particularly the abundant enzymes present, would be detrimental to transcription and translation.

To investigate the effect of UV and/or the uncaging by-products on expression from the caged plasmids, the LA-plasmids were UV-irradiated before the addition of the IVTT system. Due to the short-lived nature of the free radicals (a few seconds),[202] they would be immediately quenched, and could no longer react with enzymes. Figure 3.6B shows that when LA-plasmids were uncaged in the absence of the IVTT components, their expression was as high as the amine plasmid control, which indicated that UV and/or the potential presence of free radicals could have dampened gene expression upon removal of the PCB group. Although the exact reason responsible for the lowered activity was not studied further, one way to decouple the effect of UV and free radicals was to make use of radical scavengers such as antioxidants. Their presence in IVTT reactions would rapidly quench any potential radical formed as a consequence of the

uncaging reaction. If expression was still negatively impacted, then the main culprit would be UV damage. Nevertheless, the 25% 5'-ap-dUTP (~167 amino modifications) used in the extension reaction was still significant, and should be decreased further.

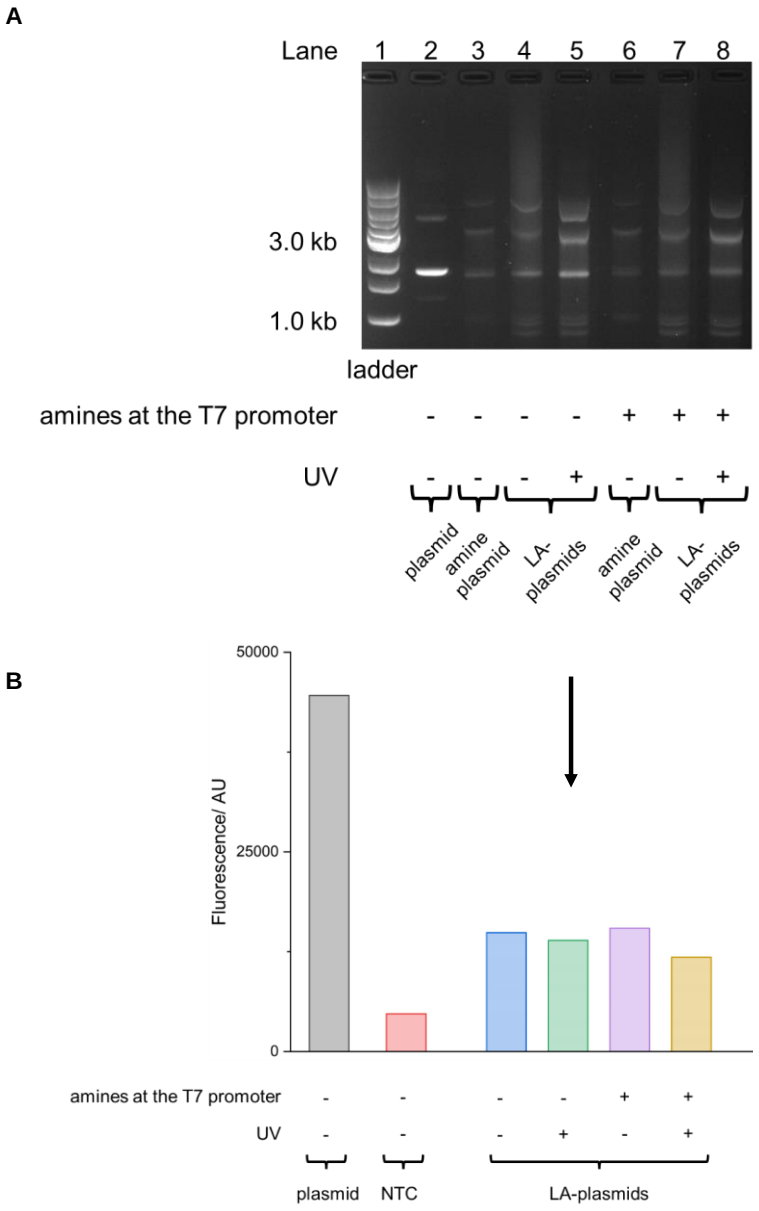


Figure 3.7. Caging the T7 promoter of mV_nicks plasmid with 2 modifications

A. Plasmids were extended with primers containing either 0 or 2 amino-C6-dT modifications at the T7 promoter and the regular triphosphates and 1:19 ratio of 5'-ap-dUTP:dTTP, reacted with PCB-NHS ester then bound to WTS. B. LA-plasmids from (A) were used in IVTT (n = 1).

As transcription requires binding of the T7 RNAP at the T7 promoter, abolishing this interaction could effectively repress expression altogether. Initially, primers containing

either 0 or 2 amino-C6-dT modifications at the T7 promoter (all primers containing amino-C6-dT modifications used in this thesis were commercially available) were extended in the presence of 5% 5'-ap-dUTP (corresponding to ~33 or 35 amines). Under these conditions, an irradiation time of 10 mins was enough to restore plasmid bands (Figure 3.7A). However, the few modifications were not enough to suppress expression (Figure 3.7B). It was also shown that the DpnI-digested plasmid did not contribute to the OFF state (Figure 3.8); thus, any leakiness would have derived from aberrant LA-plasmid expression.

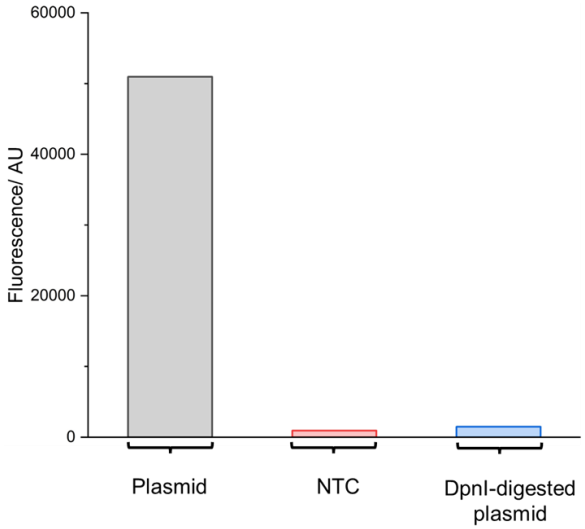


Figure 3.8. The DpnI-digested mV_nicks plasmid was not responsible for leakiness in the OFF state

Expression from DpnI-digested plasmid was not dissimilar to NTC (n = 1).

The LA-DNA system developed by Booth *et. al.* installed 7 amino modifications at the T7 promoter without additional modifications in the DNA template. This adequately shut down the expression to a level comparable to the NTC. To recapitulate their findings, the same primer containing 7 amino-C6-dT (Table 6.6) was used in NEEL and extended with either 0 or 5% 5'-ap-dUTP (7 or ~40 amino modifications in total). The combination of lowering modifications and caging the T7 promoter gave a minor increase in expression after UV uncaging (Figure 3.9). Notwithstanding this

encouraging result, it was clear the system was leaky, and attention was turned to improving the OFF state.

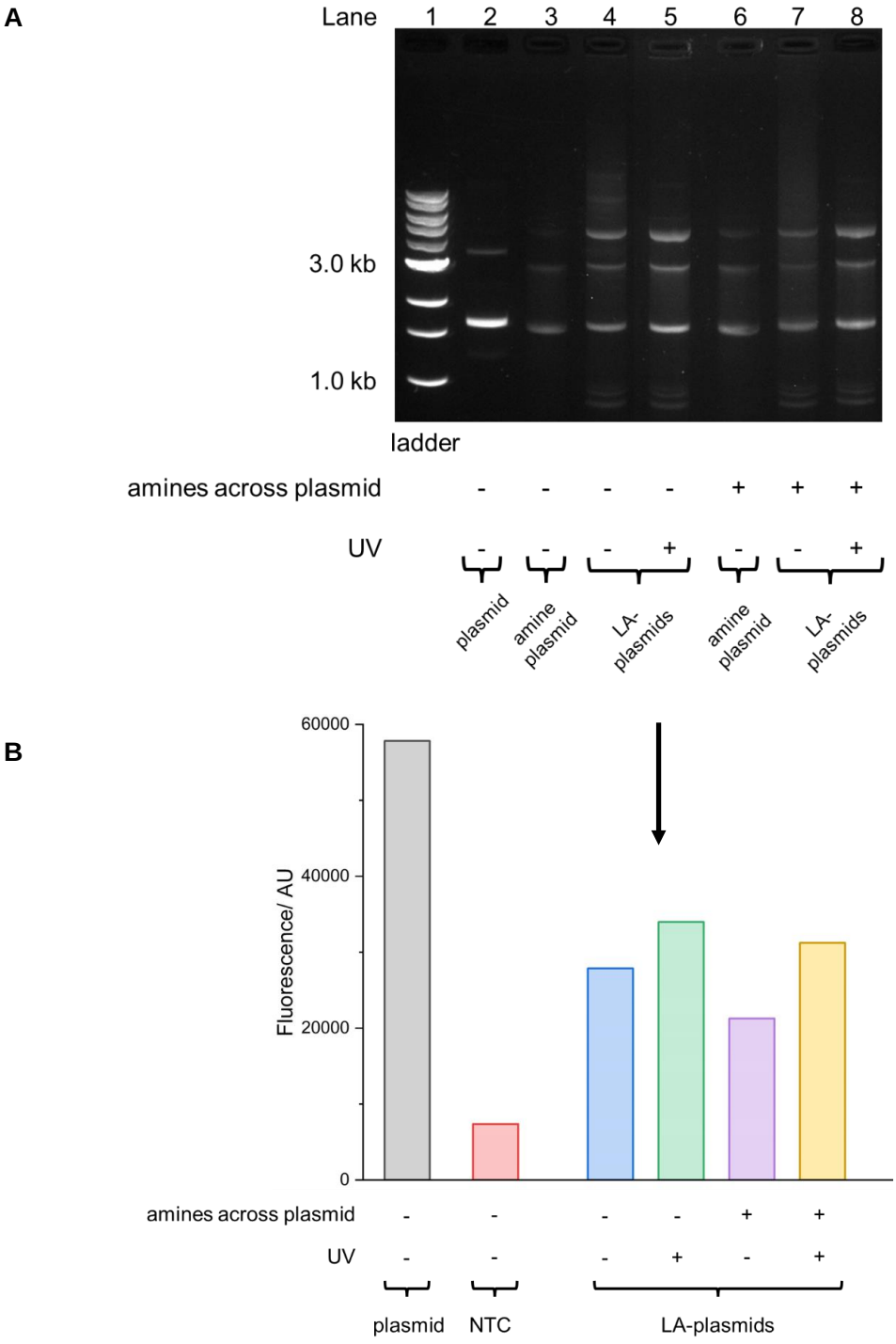


Figure 3.9. Caging the T7 promoter of the mV_nicks plasmid with 7 amino modifications A. Plasmids were extended with primers containing 7 amino-C6-dT modifications at the T7 promoter, the regular triphosphates and either no 5'-ap-dUTP, or 1:19 ratio of 5'-ap-dUTP:dTTP, reacted with PCB-NHS ester then bound to WTS. B. The corresponding LA-plasmids were used in IVTT (*n* = 1).

3.6 Optimising IVTT OFF State

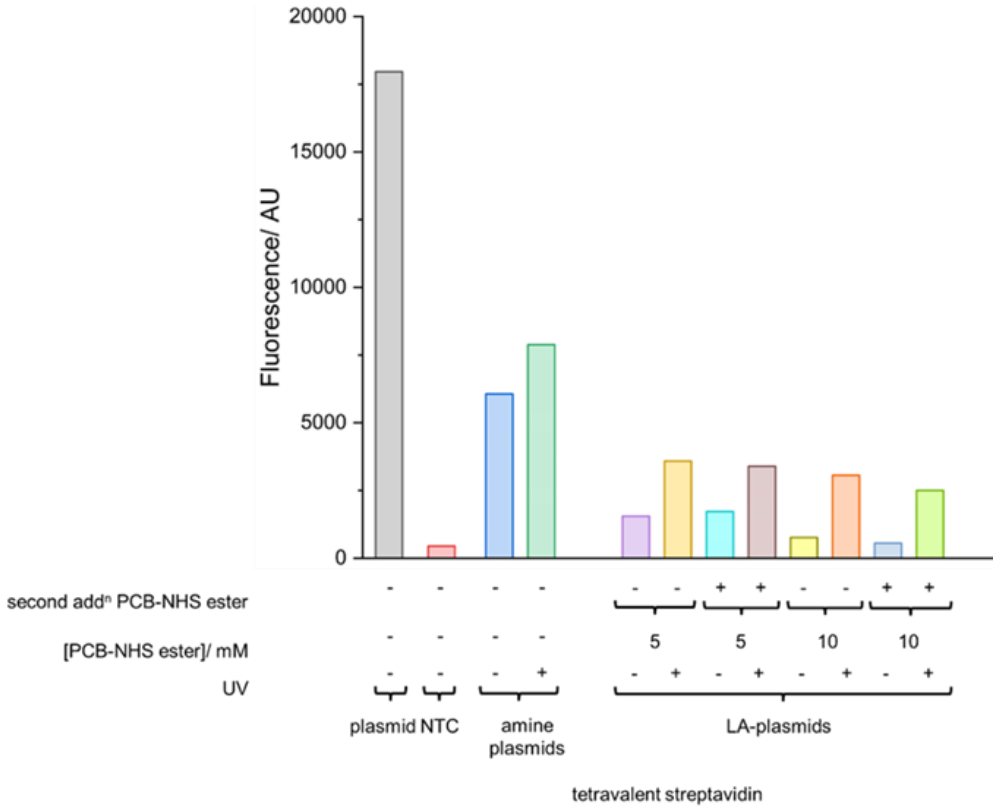


Figure 3.10. Increasing PCB-NHS ester concentration and streptavidin equivalents to assess their effects on IVTT of mV_nicks plasmid

Plasmids were extended with primers containing 7 amino-C6-dT modifications at the T7 promoter, the regular triphosphates and 1:19 ratio of 5'-ap-dUTP:dTTP, reacted with PCB-NHS ester then bound to WTS. The total reaction volumes were 90 µL and either 5 or 10 mM PCB-NHS ester was initially reacted with 3.4 µg mV_nicks plasmid and sodium bicarbonate (final concentration = 100 mM, pH 8.5). After 1.5 h, PCB-NHS ester (4.5 [5 mM] or 9.0 µL [10 mM], 50 mM) was added to the corresponding conditions, and all reactions were left to react at 20 ºC overnight. WTS equivalent was also increased. These alterations benefitted the OFF state whilst being detrimental to the ON state (n = 1).

A possible explanation for the high OFF state was the incomplete reaction between amine and PCB-NHS ester. This facile reaction should occur readily in the presence of sodium bicarbonate as the base. However, water is a nucleophilic solvent, and under basic conditions could effectively compete with the amines present on plasmid to quench PCB-NHS ester. Booth *et. al.* employed a 5 mM concentration of PCB-NHS ester and this condition had been used thus far. However, their reaction involved the

modification of only 7 hexylamine groups on a short oligo. As plasmids were much larger and contained up to ~40 amino modifications situated far apart, higher PCB-NHS ester concentration might be required to fully activate the amines. To account for the rapid hydrolysis of PCB-NHS ester, the reagent was replenished 1-2 h after the reaction began. Additionally, 3 streptavidin equivalents were also used henceforth to ensure all biotins were fully bound.

Both the OFF and ON states decreased in a dose-dependent manner (Figure 3.10). Expectantly, higher plasmid caging translated to a lower OFF state; whereas the lower ON state reflected leftover caged PCB following illumination. At this stage, >95% of LA-plasmid expression had been repressed relative to the native plasmid. To improve the ON state, it was necessary to characterise the plasmids to quantify the number of amine and PCB modifications. This knowledge would inform how LA-plasmids could be adjusted to enhance the ON state.

3.7 Quantifying Plasmid Modifications

3.7.1 HPLC and LC-MS to Quantify Amine Modifications

The number of amino modifications could be estimated through HPLC by measuring the amine nucleoside's peak area and comparing these to known standards. As the triphosphates rendered nucleotides extremely polar and difficult to separate by HPLC, these groups had to be removed.

This could be achieved through concerted actions of an enzyme mix reported by Quinlivan *et. al.* and adapted from Crain's original protocol.[203,204] The recipe involved Benzonase which digested dsDNA into short oligomers. Phosphodiesterase I catalysed the exonucleolytic reaction of these oligos in the 3'-5' direction to give

nucleoside 5'-phosphates. Finally, alkaline phosphatase hydrolysed the phosphate groups to form 2'-deoxynucleosides.

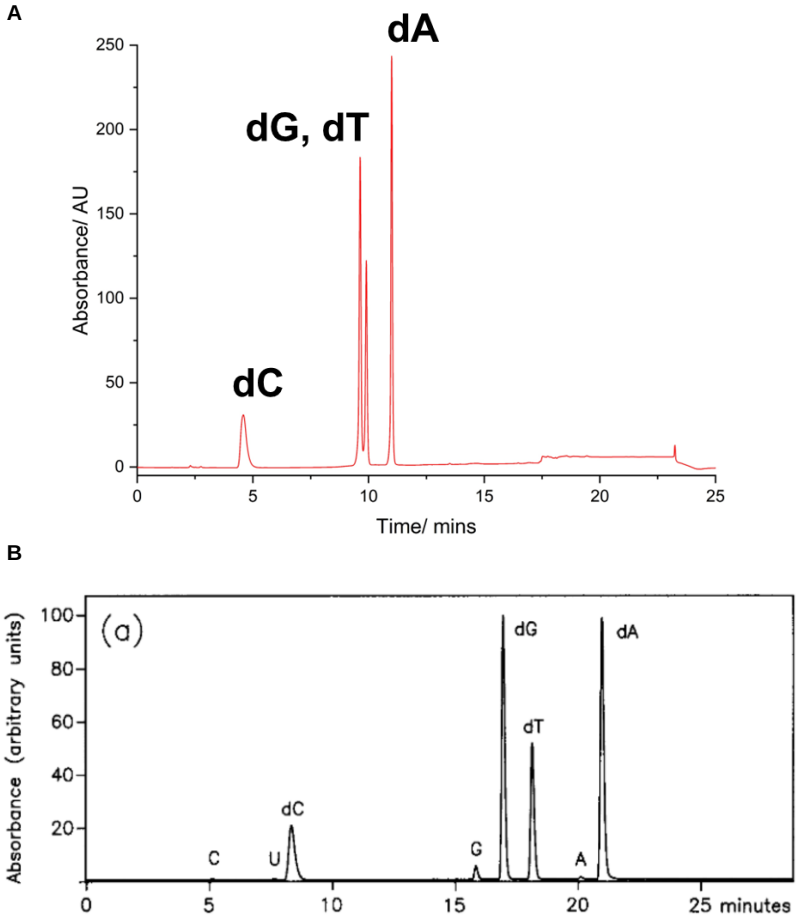


Figure 3.11. Optimised HPLC conditions to visualize the standard nucleosides

A. HPLC chromatogram of the nucleoside standards. Absorbance was measured at 260 nm. The order and signal profiles agreed with B. Crain's HPLC chromatogram - figure adapted from reference 204.[204]

Nucleoside standards of the 4 bases were commercially available and these were used to establish a HPLC run (Figure 3.11). The order and profiles of nucleoside signals agreed well with the literature.[204] Next, native plasmids were digested following Quinlivan's protocol and analysed at different time points by HPLC. Enzyme concentrations had to be doubled to enable complete digestion, which occurred within 1 h (Figure 3.12).

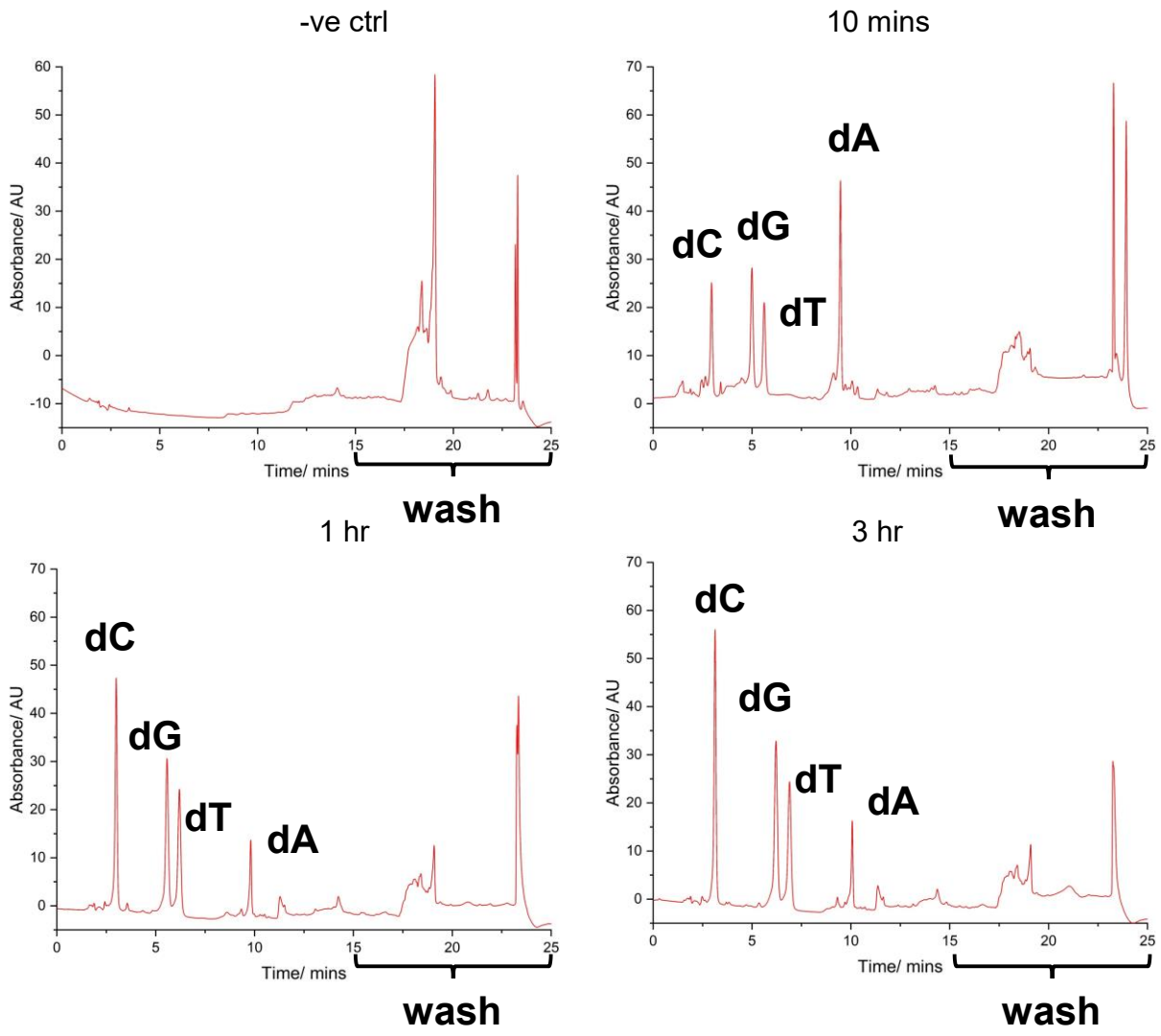


Figure 3.12. mV_nicks plasmid digestion time points

Plasmids were digested using Quinlivan's protocol with doubled enzyme concentrations. The nucleosides were analysed by HPLC (see methods Section 6.4.5); the reaction was finished by 1 h. Absorbance was measured at 280 nm.

Liquid chromatography-mass spectrometry (LC-MS) allowed reaction monitoring through sensitive mass detection. Bulk primer (1 µg) and plasmid digestion (37 µg) reactions were performed and analysed via LC-MS. However, neither the primer nor plasmid digestion reaction gave signals for their amino nucleosides, whilst the regular nucleosides (dA, dT, dG, dC) could be detected (Figure 3.13).

3.7.2 Quantifying PCB Modifications using Dye-labelled MVS

Although the number of amine triphosphates could not be ascertained, the number of streptavidin could still be determined through quantifying PCB. The strategy required dye-labelled MVS to bind to biotin on the plasmids. After the complete binding was achieved, excess MVS-dye was washed away. At this stage, the plasmid concentration had to be re-measured. This is typically done using a Nanodrop® spectrophotometer, which measures absorbance at 260 nm. This number could be used to calculate the DNA concentration via the Beer-Lambert law:

$$A = \varepsilon c l$$

Where: A = absorbance (arbitrary units)

ε = molar absorption coefficient ($dm^3\ mol^{-1}\ cm^{-1}$)

c = concentration ($mol\ dm^{-3}$)

l = path length (cm, typically assumed to be 1)

However, for PCB and LA-plasmids, Nanodrop readings would give an overestimation of the concentration. This was because these plasmids contained 2-nitrobenzyl groups[205] and streptavidin[206], which also absorbed at 260 nm. To circumvent this issue, PCB plasmid concentrations were measured with Pico488 dye. This reagent is a DNA-binding dye that is specific for dsDNA. Moreover, Pico488's UV-Vis spectra (using https://www.fpbase.org/spectra/) do not overlap with those of streptavidin[207] and 2-nitrobenzyl and could thus be used to reliably estimate the MVS-dye plasmid concentration.

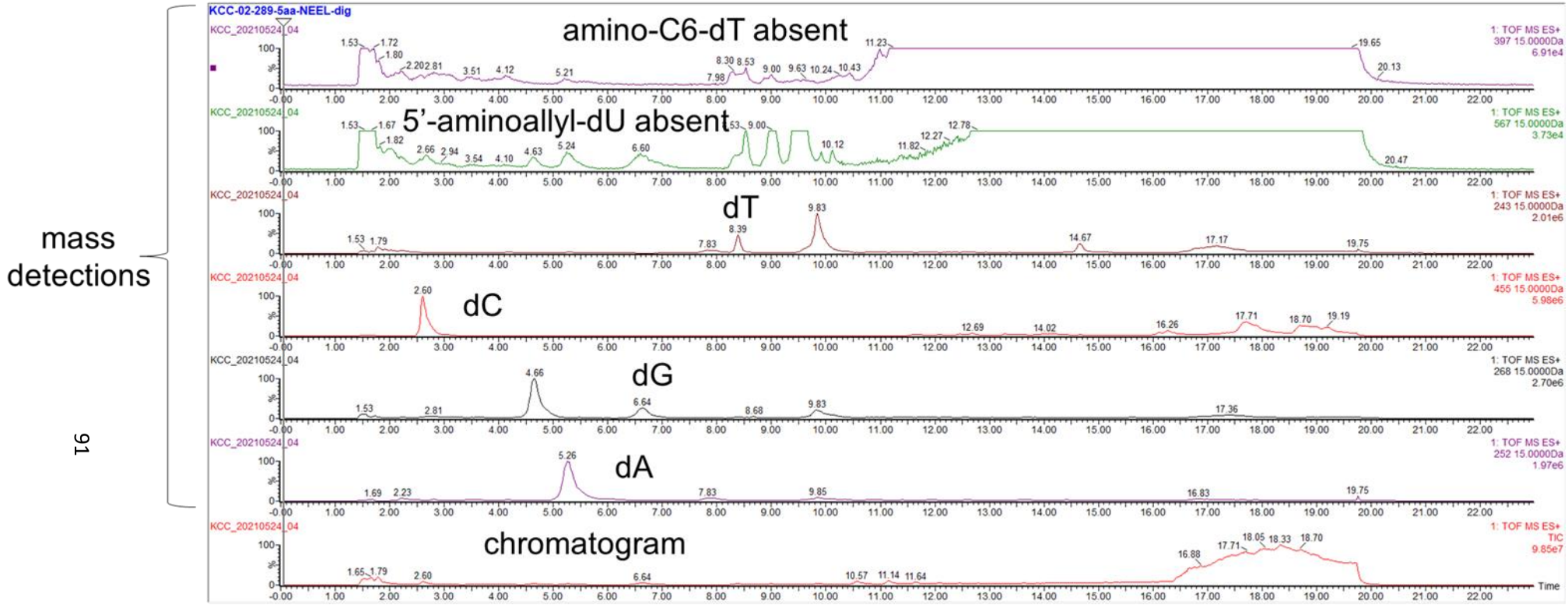


Figure 3.13. LC-MS chromatograms of digested mV_nicks amine plasmids

Digested 5% 5'-aa-dUTP plasmid was analysed by LC-MS. The nucleosides were detected by their masses. Amino-C6-dT and aminoallyl-dU were absent.

As Pico488 was used to measure the concentration of plasmid-MVS-dye, the dye conjugated to MVS must also not interfere with Pico488 absorption and emission. Atto647 was chosen as it possessed a high quantum yield, was photostable and water-soluble, and importantly, its spectra do not overlap with those of Pico488 (compared using https://www.fpbase.org/spectra/). Plasmid-MVS-dye concentration could thus be estimated by measuring their fluorescence upon binding to Pico488 and calculated from a calibration curve of known standards. Once the concentration of plasmid conjugate was known, a second calibration curve of MVS-dye was established and used to determine the streptavidin concentration, and thereby the number of streptavidin and PCB per plasmid.

Initially, using 5 equivalents of MVS-Atto647, this method estimated 10 and 50 PCBs per plasmid, for plasmids extended with the 7 amines primer in presence of 0 or 5% 5'-aa-dUTP, respectively (Table 3.1). These values were deemed acceptable compared to the theoretical estimates for these plasmids. To ascertain the reliability of these calculations, the same experiment was repeated with 8 equivalents of MVS-Atto647. If the values were reliable, an increase in MVS-dye should not affect the empirically-derived PCB estimates.

amino dUTP in extension reaction/ %	0		5	
MVS-dye equivalents	5	8	5	8
Theoretical PCB/plasmid	7	7	40	40
Derived PCB/plasmid	10	26	50	87

Table 3.1. Estimating PCB and streptavidin using MVS-Atto647

Plasmids were extended with primers containing 7 amino-C6-dT modifications at the T7 promoter, the regular triphosphates and either no 5'-aa-dUTP, or 1:19 ratio of 5'-aa-dUTP:dTTP, reacted with PCB-NHS ester then bound to MVS-dye. 20 mM PCB-NHS ester was initially reacted with 6.0 µg mV_nicks plasmid and sodium bicarbonate (final concentration = 100 mM, pH 8.5) in a 65 µL total reaction volume at 20 ºC. After 1 h, PCB-NHS ester (6.0 µL, 100 mM) was added to the reaction, and all reactions were left to react at 4 ºC overnight. PCB plasmids were bound to either 5 or 8 MVS-dye equivalents and their fluorescent was measured. The number of PCB per plasmid was calculated from a titration curve.

At 8 equivalents MVS-Atto647, these estimates increased to 26 and 87 PCBs. This suggested this method could not reliably be used to calculate PCB modifications. This was likely due to the incomplete removal of excess MVS-dye. The residual MVS-dye would artificially inflate the readings, hence increasing the estimated values when a higher MVS-dye equivalent was used. Furthermore, the purification of plasmid-streptavidin-Atto647 was inefficient and consistently gave very low concentrations. This was thought to be due to the size of these conjugates, which may inhibit their elution through the PCR column membranes. Low concentrations would further impact the validity of the estimations, e.g. concentrations were often <5.0 ng uL^{-1}, so even a minor deviation of ±0.5 between readings would significantly affect the PCB calculations.

Originally, the number of amine and PCB modifications were quantified to determine the minimal number of modifications required to ensure a tight OFF state, and high ON state after photorelease. This was necessary because when a mixture of 5'-aa-dUTP and dTTP were used in the extension reaction, T7 DNAP may have a preference for one triphosphate over the other and would incorporate them randomly across the plasmid, so the actual average numbers of amines/PCBs/streptavidin modifications per plasmid were unknown.

When only one triphosphate was used exclusively, as was the case for plasmids extended with the 7-amine oligo and in the exclusive presence of the regular triphosphates, the number and positions of the amine modifications were already pre-determined, e.g. they were dictated by the amine primer. Thus, if caging the promoter was sufficient to achieve tight repression and high photoactivation, then the presence of modifications across the plasmid would not be necessary. A comparison was made between 0% and 5% 5'-aa-dUTP LA-plasmids, both containing a caged T7 promoter

(note that a similar comparison had been performed [Figure 3.9B], but a lower PCB-NHS ester concentration had been used previously and therefore the amines were incompletely modified). When only the promoter was caged, MVS LA-plasmids could repress expression by ~88%. This represented caging due solely to steric hindrance imposed by streptavidin. When WTS was used, >95% repression was achieved (Figure 3.14A). The tighter OFF state reflected WTS' ability to cross-link plasmids, which made it more difficult for T7 RNAP to access.

For the 5% 5'-ap-dUTP LA-plasmids (Figure 3.14B), similar repression relative to 0% 5'-ap-dUTP WTS LA-plasmids was achieved with either MVS or WTS. This suggested there was no additional benefit to having streptavidin placed across the plasmid and blocking T7 RNAP from accessing its promoter was the sole factor to control plasmid expression.

Moreover, the photomodulation efficiency was the highest for 0% 5'-ap-dUTP WTS LA-plasmids. This was unsurprising, as there were only 7 uncaging events required for these plasmids, versus ~40 events for the 5% 5'-ap-dUTP LA-plasmids. The uncaging process was expected to be more complete for the former, thus giving a higher ON state. Hence, subsequent LA-plasmids were only caged at the promoter. This had the advantage of ensuring a more homogeneous LA-plasmid population, as the exact number and location of the modifications could be controlled. The fewer caging groups also meant lower toxicity from UV (as milder conditions would be required) and uncaging by-products.

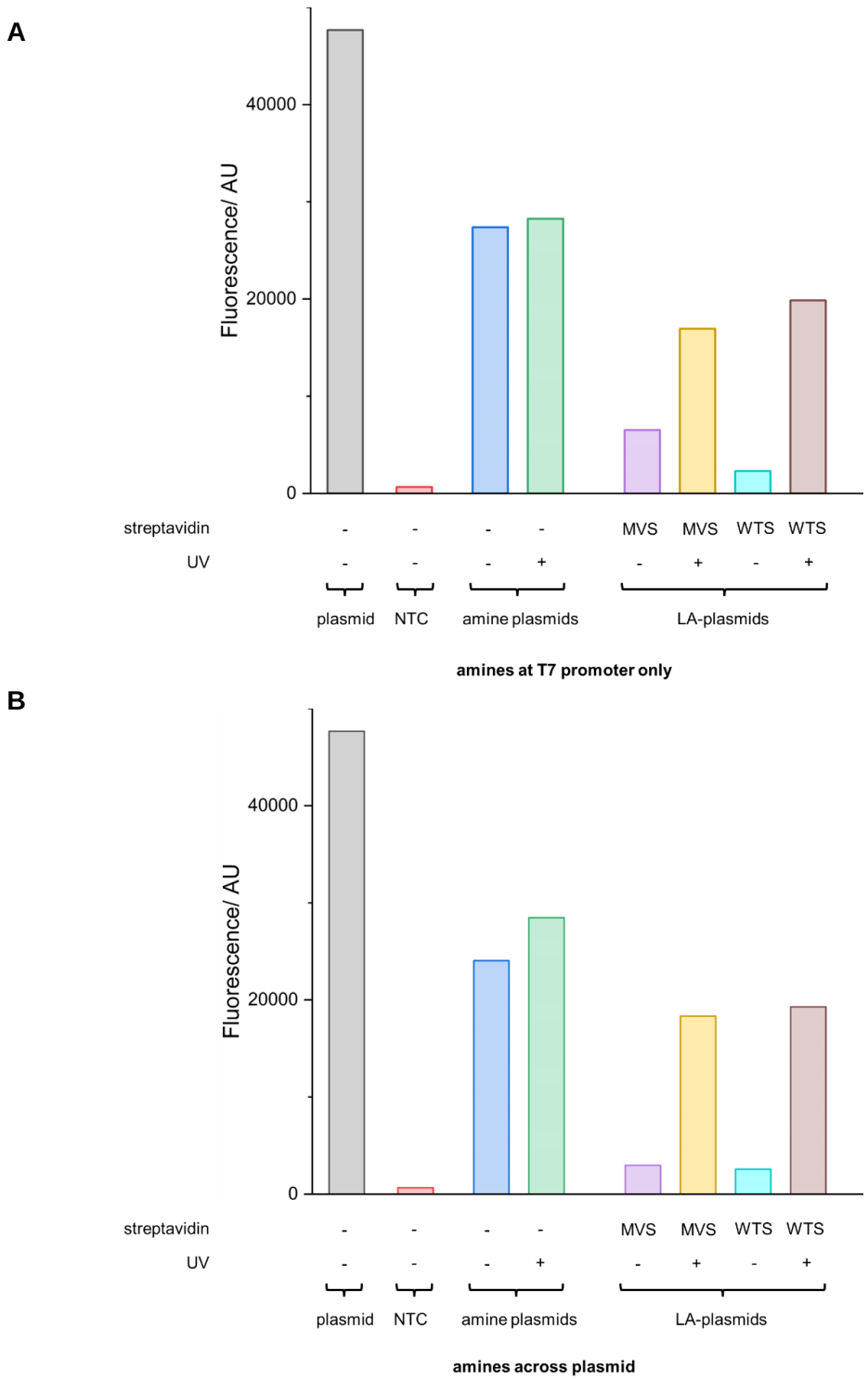


Figure 3.14. Comparing IVTT of mV_nicks plasmids with and without modifications across the plasmids

Plasmids were extended with a primer contanining 7 amino-C6-dT modifications and regular triphosphates and in the presence of A. 0 or B. 5% 5'-aa-dUTP relative to dTTP. 10 mM PCB-NHS ester was initially reacted with 5.0 µg mV_nicks plasmid and sodium bicarbonate (final concentration = 100 mM, pH 8.5) in a 175 µL total reaction volume at 20 ºC. After 1.5 h, PCB-NHS ester (17.5 µL, 50 mM) was added to the reaction, and all reactions were left to react at 20 ºC overnight, then bound to either MVS or WTS. These were then used for IVTT (*n* = 1).

3.8 LA-plasmids with Caged T7 Promoter

Up to now, an amine primer was annealed, followed by extension in presence of amino dUTPs to give amine plasmids. Amine plasmids were then reacted with PCB-NHS ester to furnish PCB plasmids. With the new strategy to solely cage the promoter, all modifications were contained within the primer. This meant that the amine primer could first be reacted with PCB-NHS ester to give the PCB oligo, which could be purified, then annealed and extended to give PCB plasmid (Figure 3.15).

The latter strategy was analogous to how Booth *et. al.* prepared their LA-DNA. The additional benefit of this method over the previous was that the PCB oligo could be HPLC purified, ensuring the removal of all quenched PCB by-products, which would not have been possible previously, as plasmids were not suitable for HPLC.

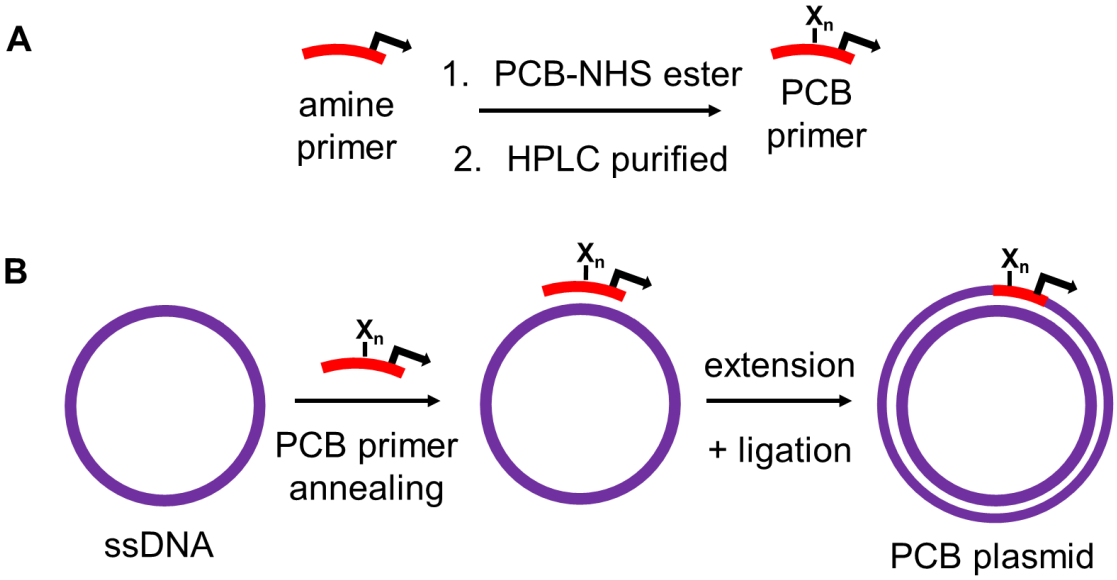


Figure 3.15. Annealing PCB oligo to the ssDNA template

A. Amine primer was reacted with PCB-NHS ester and HPLC-purified to obtain pure PCB primer. B. PCB primer was used for NEEL to directly give PCB plasmid.

The 7-PCBs primer was prepared from a 7 amino-C6-dT-containing primer, HPLC-purified, and confirmed through LC-MS and denaturing polyacrylamide gel electrophoresis (Appendix Figure A1, Table 6.7). The base-catalysed substitution of the reactive amines at the NHS-ester was trivial. The disappearance of the starting material peak indicated the reaction had gone to completion. Using this HPLC protocol, the PCB oligo often appeared between 22-25 mins. The other major peak's identity at ca. 19 mins was not verified, but most likely represented the quenched PCB compound as it absorbed at 365 nm.

LC-MS showed the 7-PCBs oligo was the major product, but an adjacent minor peak corresponding to the incompletely reacted 6-PCBs oligo was also present due to co-elution on HPLC (1% abundance relative to the 7-PCBs oligo, Appendix Figure A1C). Attachment of 7 PCBs retarded the oligo relative to the unmodified oligo on PAGE (Appendix, Figure A1D). A milder UV irradiance (~1.1 mW cm^{-2} for 3 mins) was used to uncage PCB oligo as there were fewer PCB groups, and this setting was reported to work well for LA-DNA.[208] These conditions led to the formation of multiple bands, which corresponded to the successive removal of PCB groups.

Before using PCB oligo to resynthesise plasmids, it was necessary to ascertain whether the oligo would survive annealing conditions. Specifically, the prolonged incubation at elevated temperatures may cleave the carbamate on the PCB oligo to remove both the UV-responsive and biotin functionalities.

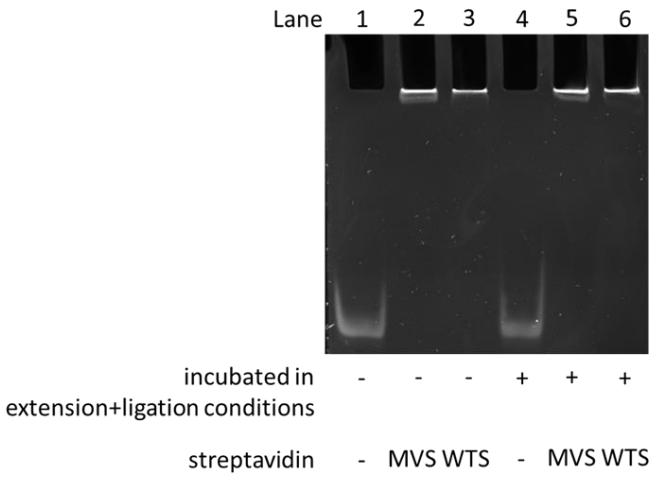


Figure 3.16. Compatibility of T7 7-PCBs oligo in extension and ligation conditions

PCB oligo was not degraded in extension and ligation conditions and could still bind streptavidin.

Figure 3.16 showed that PCB could bind streptavidin as normal, and the oligo was amenable to annealing conditions. The order of operation, e.g. performing NEEL followed by reaction with PCB-NHS ester, or performing NEEL with the PCB primer, also did not matter, as both could efficiently give LA-plasmids (Figure 3.17). Thus, the T7 DNAP proved remarkably flexible as it could incorporate both 5'-aa- and 5'-ap-dUTP during extension, and tolerated amino-C6-dT and PCB modifications on the primer. Hereafter, the primer containing amino-C6-dT modifications was used to prepare PCB oligo. LA-plasmids were then re-synthesised with the PCB oligo and extended with the regular triphosphates.

From the PAGE data (Appendix, Figure A1), the gentler UV condition could not remove the majority of PCBs. Therefore, a comparison was made between this and the previously optimised UV parameters (2.5 mW cm^{-2} for 10 mins). Whilst there was an improvement upon irradiating plasmids at 1.1 mW cm^{-2}, this was minor (Figure 3.18). The higher UV settings led to more uncaging, and hence higher expression from the T7 promoter.

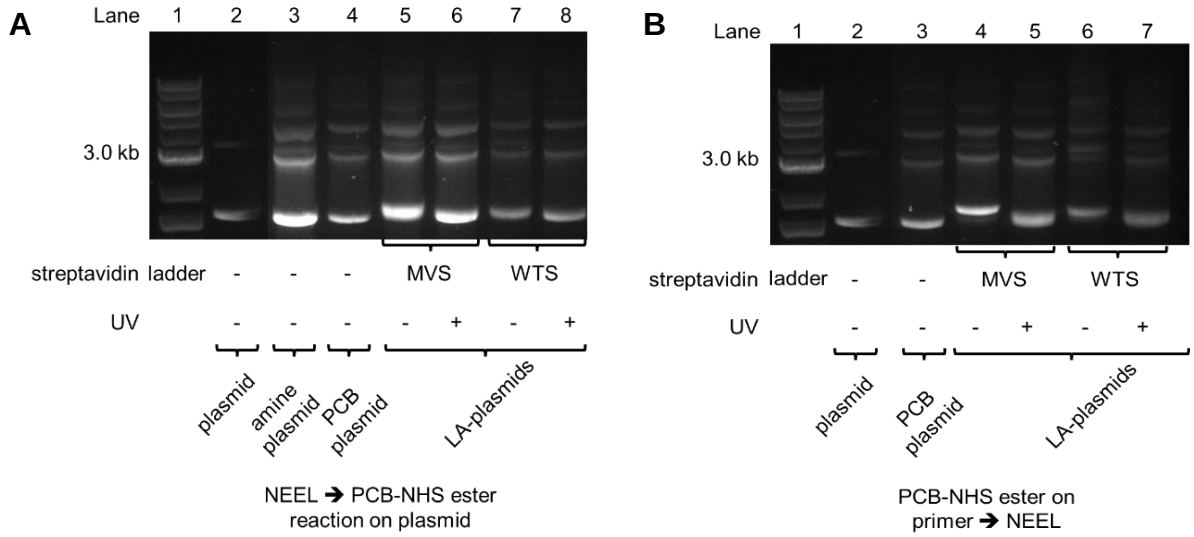


Figure 3.17. Comparing the order of operation to synthesise mV_nicks LA-plasmids

LA-plasmids could be prepared by A. performing NEEL with an amine primer followed by a reaction with PCB-NHS ester or B. performing NEEL with a PCB primer.

The optimised Kunkel protocol was used to prepare LA-plasmids, and these were re-tested in cell-free reactions. IVT showed that although reformed, amine, and PCB plasmids could be transcribed, LA-plasmids abolished transcription in the absence of the UV stimulus. After photorelease, expression was restored (Figure 3.19A). This trend was again illustrated through IVTT, whereby LA-plasmids expression was abolished by ~98% for both MVS and WTS LA-plasmids compared to the native plasmid (Figure 3.19B). The photomodulation efficiency for both plasmids was ~19-fold. The similar performances between MVS and WTS were in contrast to previous experiences when WTS LA-plasmids typically gave a lower OFF state. This was attributed to WTS' ability to crosslink plasmids, which likely did account for the tighter repression. However, the order of operation being the major difference, it was perhaps here where the explanation lay.

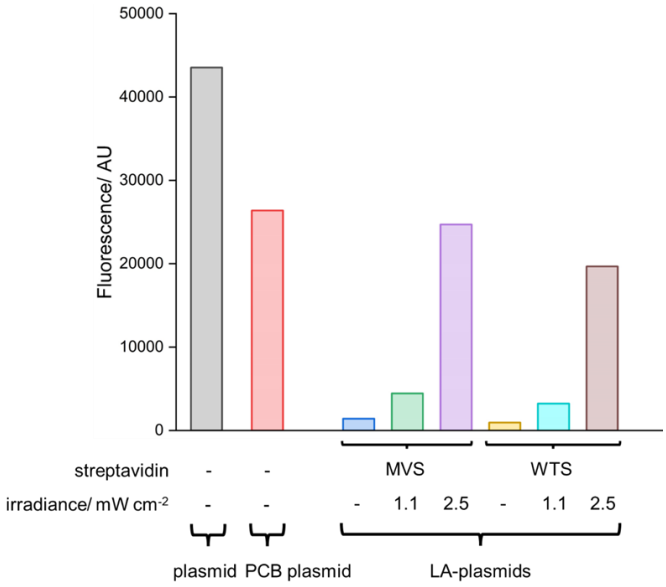


Figure 3.18 Comparison of milder and previously optimized UV conditions on IVTT performance of mV_nicks LA-plasmids

A higher irradiance and longer irradiation duration led to more uncaging and higher fluorescence (*n* = 1).

Previously, when the amine plasmid was reacted with PCB-NHS ester, it was possible that the reaction had not taken place for all the available amine sites, importantly those at the T7 promoter. This was supported by the observation that increasing PCB-NHS ester in the reaction with the amino-modified plasmid produced a better OFF state (Figure 3.10). This resulted in uneven and incomplete blocking of the T7 promoter and hence, WTS crosslinking could enhance repression. By using a HPLC purified PCB oligo, the majority of resultant LA-plasmids would be fully modified at the promoter. Here, steric hindrance of streptavidin was enough to block T7 RNAP to give a similar repression between MVS and WTS LA-plasmids.

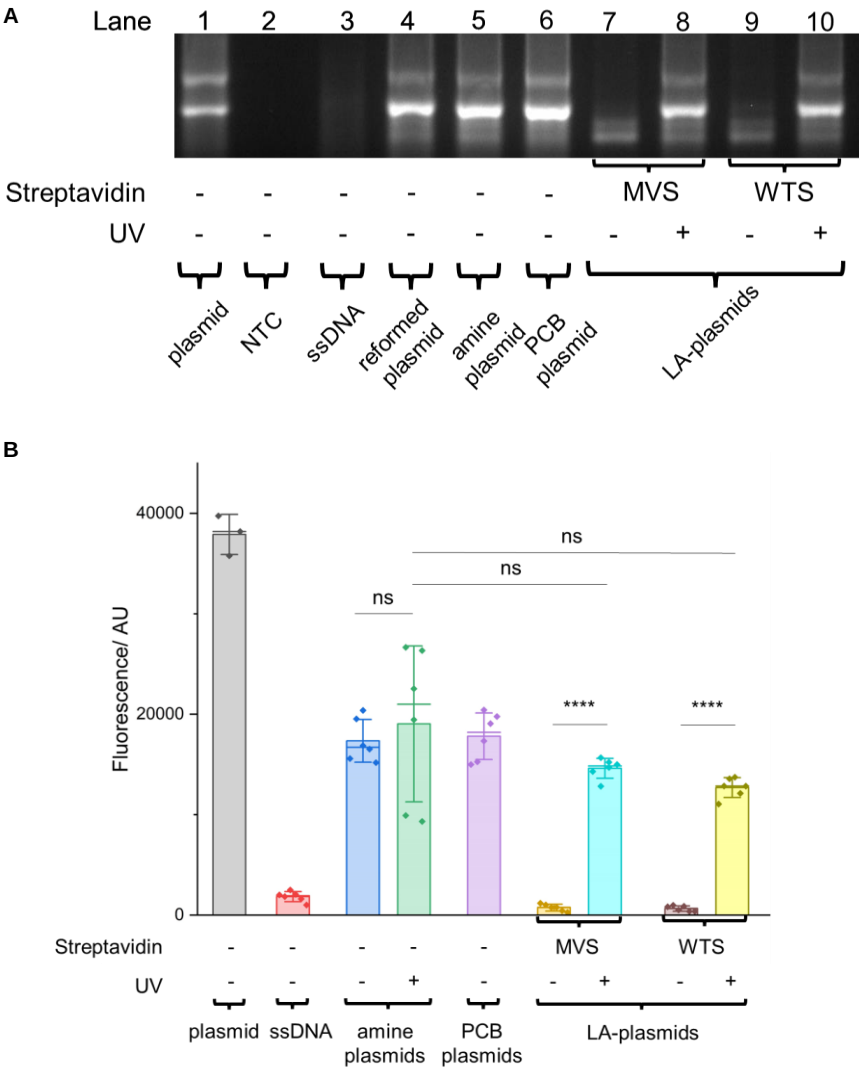


Figure 3.19. IVT and IVTT of T7-caged mV_nicks LA-plasmids

LA-plasmids prepared with the modified NEEL strategy were reassessed with A. IVT and B. IVTT. Fluorescence was background-subtracted from NTC. Native plasmid and NTC were technical triplicates (*n* = 3). All other conditions were biological duplicates of technical triplicates (*n* = 6). Error bars showed standard deviations. Two-tailed, unpaired Student's t-test was performed. ns = non-significant (P > 0.05); ****P ≤ 0.0001.

MVS and WTS LA-plasmids performed ~77% and ~67% relative to the 'amine plasmid +UV' condition. This was probably due to the incomplete removal of PCB groups, despite the higher UV settings, and also the minor production of nitric oxide radicals that interfered with the IVTT components.

Finally, these plasmids were transformed into bacteria grown on plates containing the antibiotic ampicillin. The plasmid expressed beta-lactamase that could deactivate ampicillin via lactam ring-opening. This gene was under the control of a nearby AmpR promoter, which was proximal to the T7 promoter, the presence of streptavidin on the latter may also enable the blocking of AmpR accessibility given their proximity.

Amine plasmid was tolerated well by bacteria, whilst the presence of PCB dramatically dampened colony formation (Figure 3.20). This was surprising given the amine and PCB plasmid's similar performance in cell-free reactions. A plausible explanation could be that PCB flexibility increased the overall size of the plasmid, which made it more difficult for bacterial uptake. Streptavidin further increased the bulkiness and decreased colonies even more. Here, WTS LA-plasmids produced fewer colonies compared to MVS. This could again be reasoned through WTS crosslinking, which created plasmid aggregates that could not enter cells. Additionally, once inside cells, these aggregates could more effectively sequester AmpR. Finally, photolysis via UV uncaging produced a more noticeable gain in colonies in WTS LA-plasmids. The minor difference before and after UV for MVS LA-plasmids probably meant that the colony suppression was primarily due to the inability of plasmid uptake, rather than the blocking of the beta-lactamase expression. Additionally, once plasmid replication occurs, the newly produced plasmids would not contain photocaging modifications and subsequent progeny would be able to express the ampicillin resistance gene. For WTS, the cross-linked plasmid aggregates may inhibit DNA replication, further accounting for the better repression compared to MVS LA-plasmids, and the slight increase in colony number once photolysis of the PCB group has occurred, as plasmids could now be replicated.

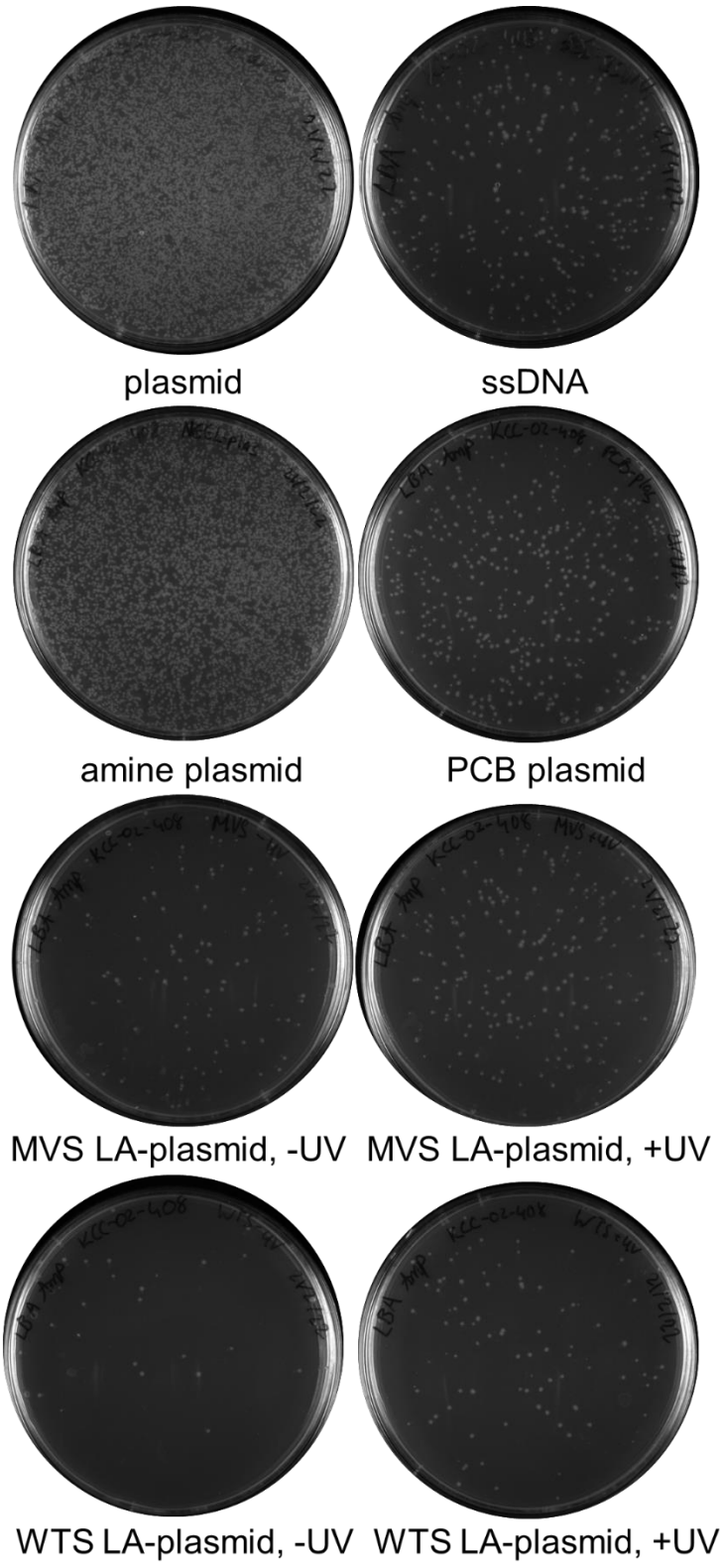


Figure 3.20. mV_nicks WTS LA-plasmids inhibited colony formation

Plasmids were transformed into bacteria to assess the effects of the T7-caged promoter on the proximal AmpR promoter.

3.9 Caging the Tac Promoter

The tac promoter is a widely used *E. coli* promoter for protein production (Section 3.1.2). It contains multiple sites that may prove amenable for photocaging. To establish whether the LA-plasmid strategy could be adapted to control expression from this ubiquitous promoter, the BspQI, and BbvCI recognition sites were cloned into a plasmid expressing mNeonGreen, a green fluorescent reporter protein (mNG), under the control of the tac promoter (referred henceforth as pTac_mNG_nicks plasmid, Tables 6.8-6.11).

Element	Consensus sequence	Plasmid sequence	Matching sequence
-35 region	TTAACA	TTGACA	5/6
-10 region	TATAAT	TATAAT	6/6

Table 3.2. Comparison between plasmid -35 and -10 regions and the corresponding consensus sequences in the tac promoter of the pTac_mNG_nicks plasmid

Plasmid sequences with underlined bases match those of the consensus sequence.

Table 3.2 showed that the -35 and -10 regions of the tac promoter contain high homology to the *E. coli* consensus sequence at these regions. All 5 thymines within these regions were important for transcriptional activity (Figure 3.1 and Section 3.1.1). Two amine primers, containing either 5 or 7 modifications, spanning the -35 and -10 regions, were chosen to cage the tac promoter (Table 6.12). Both amine primers replaced the thymines within the -35 and -10 regions with amino-C6-dT modifications. The 7 amines oligo contained 2 additional amine bases at -16 and -28, which showed a preference for thymine at these positions. These amine oligos were used to prepare PCB primers (Appendix Figures A2 and A3, Table 6.13).

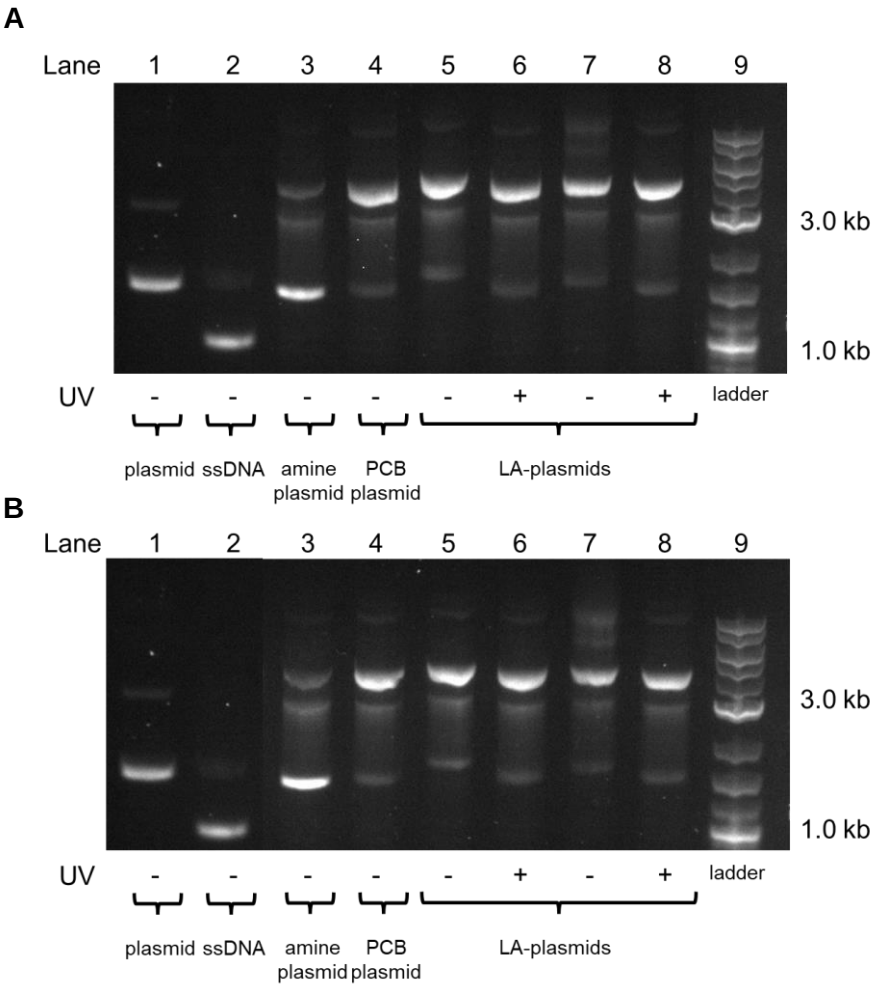


Figure 3.21. pTac_mNG_nicks LA-plasmids containing 5 and 7 modifications to cage the tac promoter

LA-plasmids were prepared from A. 5-PCBs oligo or B. 7-PCBs oligo. The proximity of the PCB groups to the 5'-end interfered with ligation.

Once again, successful PCB attachment was confirmed using LC-MS and denaturing PAGE. This time, NEEL gave the extended plasmids as the major products, indicating that ligation was inefficient (Figure 3.21). That amine plasmid could be re-ligated and not PCB plasmid suggested that the PCB groups, particularly the one closest to the 5'-end where ligation occurred, had hindered ligation. The presence of contiguous PCBs at the second and third bases could also produce an additive steric blocking effect to impede the ligase (Table 6.12).

The T7 RNAP could express from the tac promoter (Figure 3.22). However, the tac promoter appeared sensitive to base modifications, as amine plasmids completely abrogated expression from this promoter. Thus, although UV could remove the majority of PCB groups (Appendix Figure A2D), uncaging the LA-plasmids had little to no positive effect on expression. Surprisingly, PCB plasmids partially restored expression despite bearing additional PCB groups compared to amine plasmids. These trends were observed for both plasmids carrying either 5 or 7 amino modifications.

The obvious next step was to reduce the number of amino modifications at the tac promoter. Two primers were designed, each possessing 2 amino modifications, and were used to prepare PCB oligos (Tables 6.12 and 6.13, Appendix Figure A4). The first contained modifications upstream and downstream of the -35 and -10 regions (Tac-2amines1), or the modifications were incorporated in the region between the -35 and -10 elements (Tac-2amines2). These positions were chosen as they showed no preference for thymines (20-30% abundance, Figure 3.1), and hence the presence of amines was not expected to abolish expression. In addition, to avoid disruption of the ligation reaction, the amino-C6-dT base nearest to the 5'-ligation junction was positioned further away.

Using these primers, ligation was effective, and the re-ligated LA-plasmids were once again the major products (Figure 3.23). Amine primers used for NEEL should position the first modified base at least >2 bases from the 5'-ligation junction to minimise potential interference with the ligase. When the amines were positioned outside the -35 and -10 regions, expression was decreased but was not abolished (Figure 3.24A). Streptavidin helped in suppressing the OFF state, but fluorescence remained substantial. Uncaging with UV led to an increase in fluorescence.

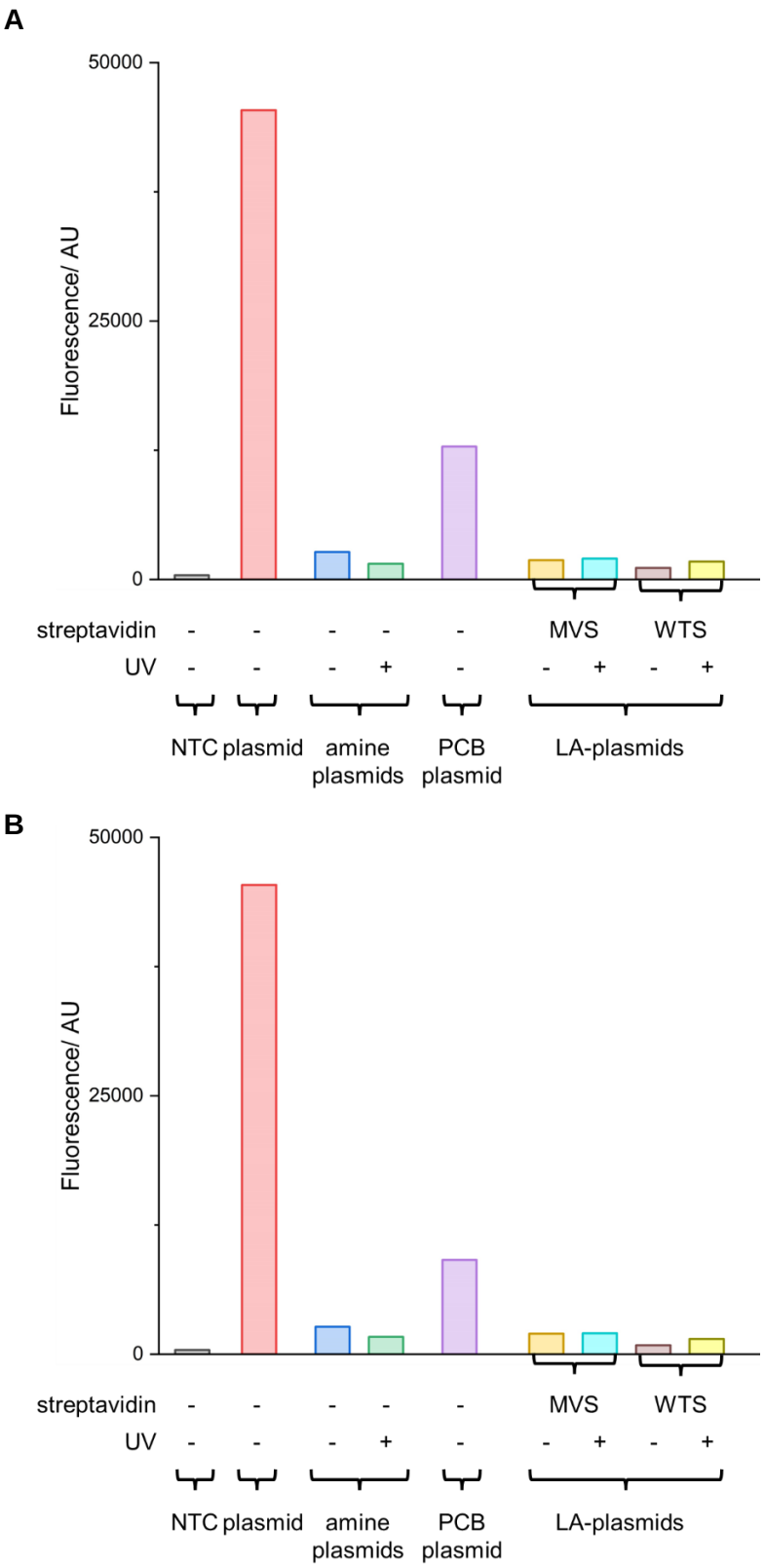


Figure 3.22. IVTT of pTac_mNG_nicks LA-plasmids containing 5 or 7 modifications to cage the tac promoter

IVTT with plasmids prepared from primers containing **A.** 5 modifications or **B.** 7 modifications (n = 1).

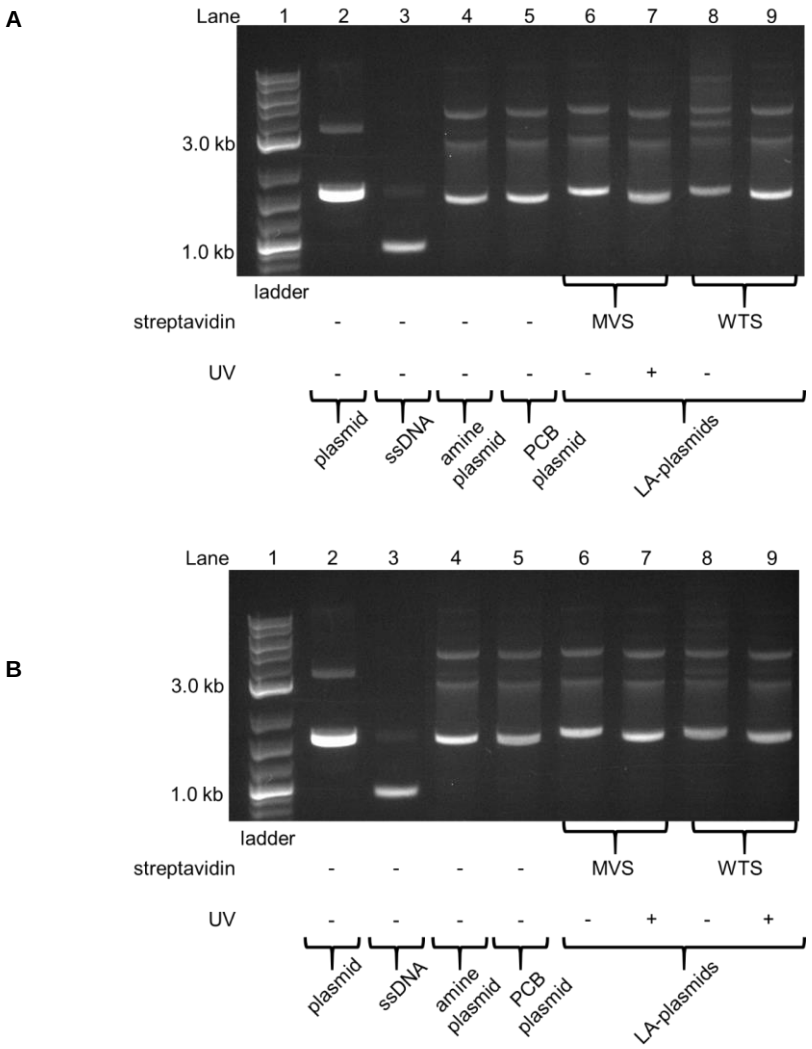


Figure 3.23. pTac_mNG_nicks LA-plasmids re-synthesised using Tac-2PCBs1 and Tac-2PCBs2

LA-plasmids were prepared from A. Tac-2PCBs1 or B. Tac-2PCBs2 oligos. By positioning the first amino-C6-dT further away from the 5'-ligation junction, ligation plasmids became the major products.

When the amines were located in between the -35 and -10 regions, expression was comparatively lowered across all conditions, and once again the presence of streptavidin further promoted repression, whilst UV marginally improved expression (Figure 3.24B). The better repression observed here was perhaps due to the positioning and proximity of the modifications in Tac-2amines2 (7 bp apart) compared to those in Tac-2amines1 (39 bp apart).

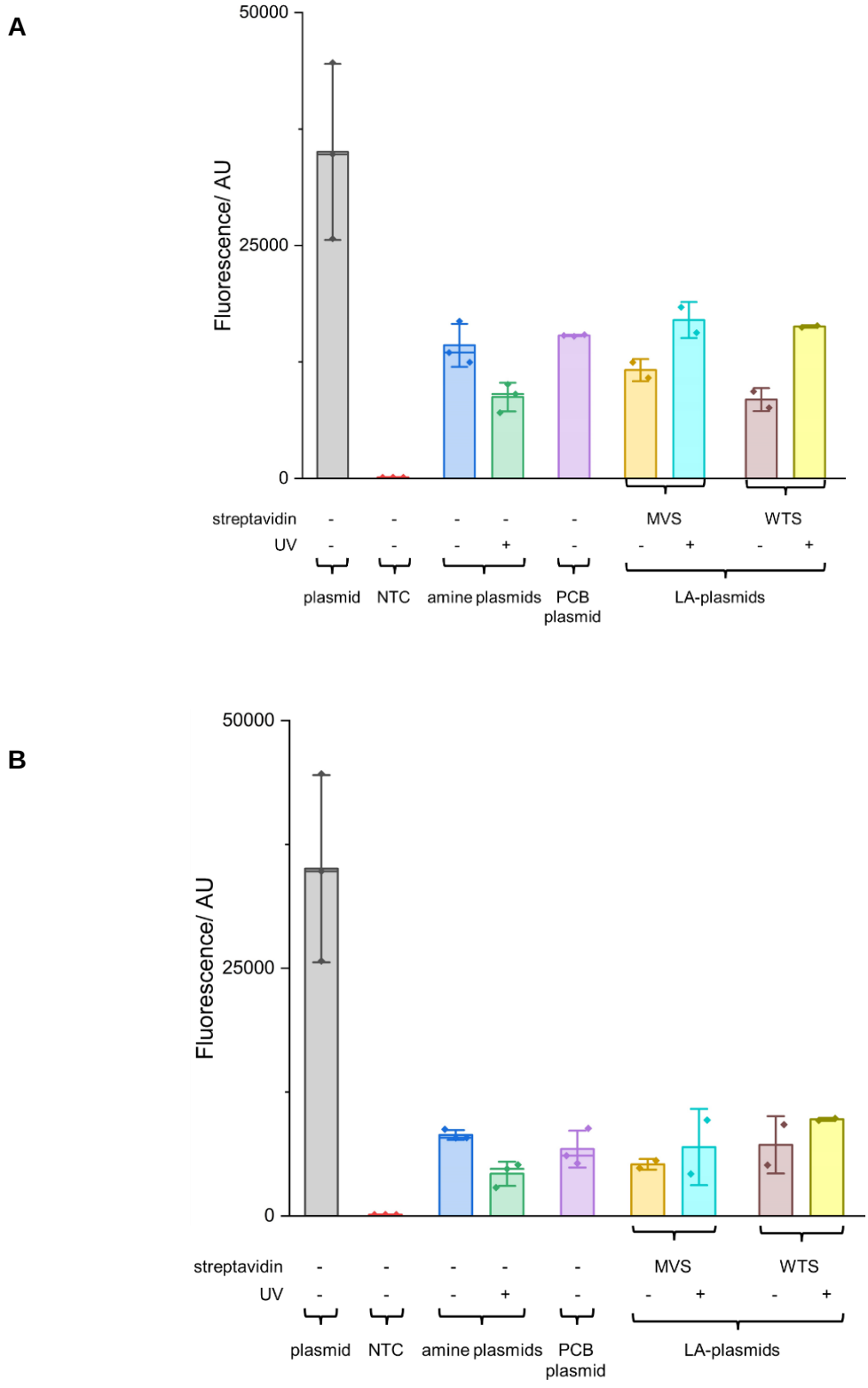


Figure 3.24. IVTT of pTac_mNG_nicks LA-plasmids resynthesised using Tac-2PCBs1 and Tac-2PCBs2 primers

IVTT with plasmids prepared from **A.** Tac-2PCBs1 or **B.** Tac-2PCBs2 oligos. Native plasmid, NTC, amine and PCB plasmids were technical triplicates (*n* = 3). LA-plasmids were technical duplicates (*n* = 2). Error bars showed standard deviations.

In the former case, these bases were likely in contact with σ factor 3, which contributed to the docking of the RNAP holoenzyme to the promoter.[187] The modifications were situated close enough to permit coordinated steric blocking of σ factor 3. For Tac-2amines1, the distal placement of the amine bases precluded this additive blocking effect, and their positions outside of the -35 and -10 regions were likely to be less important.[187] Although work with this system did not proceed beyond this point, the tac promoter proved responsive to the LA-plasmid strategy.

In particular, it has been shown that minimal leakiness could be achieved by having 5 amino modifications within the critical -35 and -10 regions. Interestingly, expression was rescued by appending the PCB groups. This suggested steric hindrance was not the only factor in determining the interaction between the transcription complex at these positions. Charges on the primary amines could also contribute to deterring complex association, which may explain the improvement observed when PCBs were attached. Although simply decreasing the number of amino modifications rescued expression, the OFF states remained leaky even when streptavidin was attached. However, this showed that with more fine-tuning in terms of the number and placement of the modifications, effective control could be achieved.

Hawley and McClure published a comprehensive study on the individual base's importance to *E. coli* RNAP recognition within the -50 to +10 regions (Figure 3.1).[188] The most highly conserved sequences across the 168 promoters studied were the -35 and -10 regions, which possessed 5 thymines that could be replaced with amino-C6-dT. This massively reduced the number of combinations required for iterative testing. As had been shown, minor changes within these regions were not accommodated. However, the effects were only studied cooperatively.

Individual bases made different contributions to RNAP affinity, and the search could begin with the less conserved bases. This may only moderately repress expression, but bulkier moieties could be installed at this point. As the -35 and -10 regions were so abundant across bacterial promoters, the ability to universally and externally control their expression would present a major advancement toward basic research, and warranted further studies. Furthermore, for inducible promoters such as the tac promoter, the additional stimulus would ensure tighter control. To obtain an output, the promoter would require uncaging by UV and the addition of IPTG, thus such a system could also act as a basic logic gate to study population dynamics.[208]

3.10 Caging the Sp6 Promoter

The Sp6 promoter is widely used for IVT to generate RNA for diverse purposes including vaccines, gene targeting (acting as guide RNA or antisense RNA), and RNA labelling.[199,209,210] Thus, it presented an attractive platform to apply our light-dependent caging. The bacteriophage-derived Sp6 RNAP is highly specific to its cognate promoter and polymerises processively across the plasmid without disassociating to give very long transcripts. Hence, the enzyme is normally used with linear DNA templates. *In vitro* transcription protocols often recommend linearising the plasmid to give blunt or 5'-overhangs downstream of the GOI.[211]

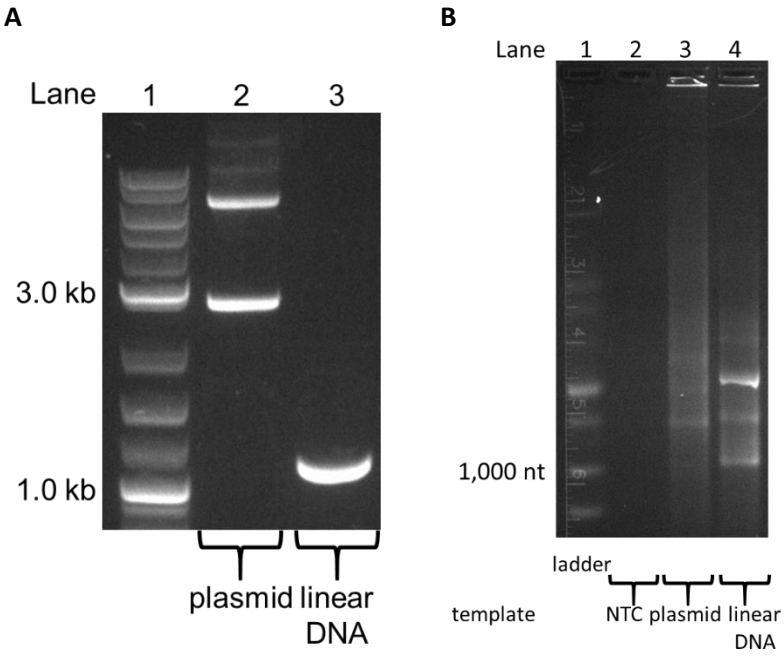


Figure 3.25 Sp6 transcription from the Sp6_mNG plasmid versus linear DNA template

A. PCR to obtain linear DNA containing the Sp6 promoter and the GOI. B. Transcription with Sp6 RNAP from either a plasmid or linear DNA template.

Initially, transcription was compared between the plasmid and a linear template obtained from PCR (Tables 6.8-6.10). These templates contained an Sp6 promoter controlling expression of mNG (henceforth referred to as Sp6_mNG plasmid). The forward primer annealed 5 bases upstream of the SP6 promoter whilst the reverse primer annealed to a region downstream of the SV40 polyadenosine signal to give a ~1.1 kb fragment (Figure 3.25A, Table 6.12). Transcription from the plasmid gave a large smear that extended to the well, and virtually no product at ~1000 nt. The smear denoted transcripts of varying lengths resulting from non-specific dissociation between RNAP and template. In contrast, when the PCR product was used as the template, 3 bands could be detected. The higher-order products were likely concatenated structures of the product band at ~1,100 nt (Figure 3.25B).

Primers containing either 5 or 7 amino-C6-dT spanning the Sp6 promoter were used to prepare PCB oligos (Appendix Figures A5 and A6, Table 6.13). These were then used as forward primers in PCR reactions to generate amine or PCB DNA. PCR from the 7-PCBs oligo always gave minimal product and so studies were not continued with this primer. Placement of the final amine, and thereby PCB, 2 bases away from the 3'-end must have prevented extension by the polymerase (Table 6.12). However, the smaller amine modification at the same position did not impede extension.

On the other hand, PCR with the 5-PCBs oligo was facile. This was because, in the 5 amines oligo, the amine nearest the 3'-end was 9 nt from the final base. Taken together with previous experience with the tac PCB oligos (Figure 3.21), locations of the 5'- and 3'-most amine bases were the determinant factors towards ligation and extension, respectively. Ideally, these should be positioned at least 5 bases from either terminus to avoid any potential interference with enzyme activity. Nevertheless, extension to generate PCB DNA was only attempted for DreamTaq DNAP, and it was possible other DNAPs could accommodate PCB modifications.

Depending on the choice of streptavidin, binding to the 5-PCBs DNA produced different band profiles. When MVS was used, the bands produced were close together. The different bands would have contained different numbers of MVS. For WTS, multiple bands were also observed, though these were further apart. The increased separation signified a more pronounced mass difference, which suggested higher-order bands were formed from the association of multiple linear fragments, afforded through inter-strand crosslinking. Nevertheless, UV illumination effectively uncaged DNA to give a single band (Figure 3.26A).

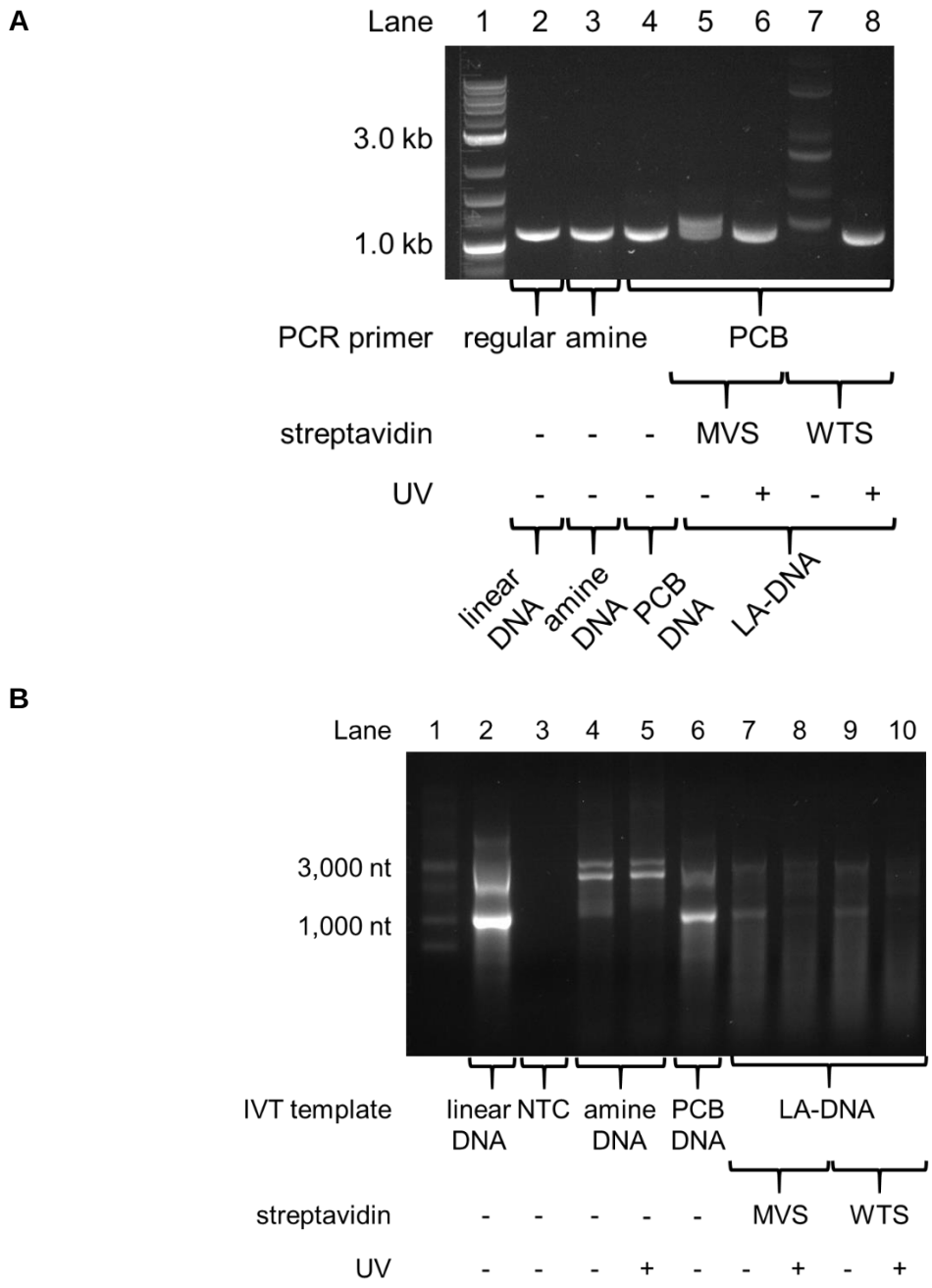


Figure 3.26. IVT of Sp6_mNG linear DNA templates
A. The regular, amine, and PCB DNA were PCR amplified using a normal, amine, or PCB oligo as the forward primer, respectively. The PCR products were then bound to streptavidin and uncaged. B. IVT was performed from the templates in (A).

The template without modifications produced the most intense product band (Figure 3.26B). Notably, PCB DNA, and to a lesser extent caged DNA, were the only other templates that produced the correct-sized transcripts. Amine DNA templates gave

anomalous bands with slower mobilities, whereas uncaging the LA-DNA abolished transcription altogether. At first sight, these observations were counterintuitive. However, analogous to the results for caging the tac promoter, charges likely contributed to the RNAP and promoter interaction. Naked primary amines carried a positive charge under reaction conditions (40 mM Tris-HCl, pH 7.9). These charges were buried after reaction with PCB-NHS ester and binding streptavidin, hence an increase in expression. PCB removal with UV re-introduced the positively-charged amines, hence transcription was suppressed upon uncaging.

This chapter fine-tuned the LA-plasmid caging strategy initially developed in Chapter 2 with the T7 promoter. It was determined that having an excess of modifications across the plasmid was detrimental to the ON state, which was attributed to UV damage and/or the production of NO radicals upon UV uncaging, and incomplete removal of the PCB groups. The number of amino modifications, PCB-NHS ester concentration, and streptavidin used to prepare LA-plasmids were optimised without much success. Significant progress was made after the discovery that simply caging the promoter was enough to achieve basal expression. In addition to lowering the number of modifications, this approach also enabled precise control over the number and positions of the modifications. LA-plasmids abolished expression in IVT and IVTT reactions. Low-dose UV irradiation removed PCB and gave a 19-fold enhancement in IVTT. When this strategy was applied to the tac and Sp6 promoters, the observation that repression occurred when the primary amine modifications were present, which could be ameliorated with their sequestration, suggested a charge-based mechanism instead of sterics could be used to cage these promoters.

4

Regulating a Mammalian Promoter and Plasmid Delivery

4.1 Introduction

4.1.1 Mammalian Transcription and the Core Promoter Elements

Compared to bacterial and phage transcription, mammalian transcription is a much more complex process involving multiple protein components and promoter elements.[212] There are 3 distinct RNA polymerases, however, RNAP II is responsible for the transcription of protein-coding genes and will be the focus of this discussion. In regards to the promoter regions, the TATA box and initiator element (Inr) are strong regulators of transcriptional activity and form the primary sites to optimise photocaging.

Restricted by the length of the amine primer, promoter elements downstream of the Inr could not be targeted. Instead, the B recognition elements immediately upstream and downstream of the TATA box (BRE^u and BRE^d, respectively) were conveniently positioned (Figure 4.1); as these elements were implicated in stabilising the transcription complex, they were also included as targeting sites.



Figure 4.1. RNAP II core promoter elements

Common RNAP II promoter elements and their relative positions. The TATA box, Inr, BRE^u, and BRE^d were chosen as target sites due to their proximity and importance in preinitiation complex assembly. Figure adapted from reference 212 under the CC BY-NC 4.0 license terms.[212]

These elements, along with their cognate protein partners will be discussed in turn. It is worth mentioning that although these core promoter elements are omnipresent amongst eukaryotic and archaea, a universal promoter element does not exist, and many promoters do not possess any of the known core motifs.[212] However, for the CMV promoter that is widely employed for mammalian transgene expression, and used in this study, full and partial consensus sequences for all of the above promoter elements were available to exert photocontrol (Table 4.1).

Element	Consensus sequence	Plasmid sequence	Matching sequence
TATA box	TATAWAWR	TATATAAG	8/8
Inr	YYANWYY	CCATTCT	7/7
BRE^u	SSRCGCC	GGAGGTC	5/7
BRE^d	RTDKKKK	CAATGCT	4/7

Table 4.1. Comparison between plasmid core promoter elements and the corresponding consensus sequences in the CMV promoter

(W) A or T; (R) A or G; (S) G or C; (D) A, G, or T; (K) G or T; (Y) C or T; (N) A, C, G, or T. Plasmid sequence with underlined bases match those of the consensus sequence.

4.1.1.1 TATA Box and Inr

Transcription initiation requires the formation of a preinitiation complex (PIC) at the promoter. The PIC is formed through the association of RNAP II and six other transcription factors (TFIIA/B/D/E/F/H), and it has been shown that *in vivo* RNAP II transcription requires all TF components.[213] The TATA-box binding protein (TBP), a subunit of TFIID, begins this process by binding to the TATA box ~30 bp upstream of the TSS. Notwithstanding its importance, the TATA box is not universal; in TATA-less promoters, the Inr can mediate basal-level transcription.[214] However, when both elements are present, they can act cooperatively to recruit TFIID,[212] and synergistically to improve transcription.[215]

The Inr is a 6-7 bp motif that encompasses the TSS, directs where transcription begins, and stabilises the TBP-TATA box interaction.[214] Its consensus sequence was determined by comparing the start site sequences of 500 genes transcribed by RNAP II.[216] Within this sequence, the adenine at +1 and thymine at +3 are the most important in terms of transcriptional strength. Replacement of thymine with any other base decreased expression by ~78-97%.[217] Presence of pyrimidines at the -2, +4, and +5 positions also contributed to TFIID binding, as their changes reduced Inr activity by about 2-fold. Mutations at other bases within the Inr consensus sequence had little effect.[217]

4.1.1.2 B Recognition Elements

After binding of TBP/TFIID to the TATA box and Inr, TFIIB associates with TBP and binds to DNA sequences immediately on either side of the TATA box.[212] TFIIB-BRE binding stabilises TBP-TATA box interaction and also positions the PIC at the

promoter.[215] Because TFIIB can only recognise its DNA targets through TBP, the absence of a TATA box precludes functional BRE sites.[212]

The presence of multiple thymines within BRE consensus sequences provides many opportunities to control this promoter (Table 4.1).[215,218] BREs' exact role concerning transcriptional activity is unclear, as these elements have both enhanced or repressed transcript production in cells, and *in vitro* assays with nuclear extracts in a promoter-dependent manner.[215,218] However, due to their importance in mediating TFIIB binding, stabilising the TBP-TATA box interaction, and orienting the PIC, blocking the holoenzyme contacts at these sites was expected to impede transcription initiation.

4.1.2 Nucleic Acid Delivery

For the transgene to be expressed *in vivo*, the plasmid must overcome many physical barriers to arrive at its site of action – the nucleus. To achieve this, a range of delivery methods have been developed over the years, which can be broadly divided into 2 categories, viral vectors, and non-viral-based methods. These carriers act as protective delivery vehicles by encapsulating the genetic cargo within their lumen. Alternatively, functionalities that can associate with nucleic acids, either through covalent linkages or supramolecular forces, may also mediate cell entry. Viral vectors are the predominant carrier of choice, comprising ~60% of gene therapy clinical trials as of 2022 (see Gene Therapy Clinical Trials Worldwide Database) owing to their superior transduction efficiency.

4.1.2.1 Problems Associated with Viral Vectors

Despite their prominence, there are several issues inherent to viral vectors that limit their use.[219] Immunogenicity against the viral capsid proteins is a major problem and

has resulted in severe complications and even death.[220] Some vectors such as adeno-associated viruses (AAVs) and retroviruses are capable of inserting their genomes into the host, which may lead to oncogenesis. One of the early setbacks for this field was the onset of T-cell leukaemia amongst 4 out of 9 patients in a clinical trial for X-linked severe combined immunodeficiency as a consequence of insertional mutagenesis.[221] The cost associated with viral vector production is notoriously expensive, making these drugs prohibitive. Glybera, which uses AAV as the vehicle to deliver lipoprotein lipase to patients deficient in this protein,[222] was the world's first million-dollar drug. It was quickly removed from the market as a result of the costly price tag. Subsequent viral-based drugs entering the market have continually pushed this price boundary, with the most recent costing $3.5 million per dose.[223,224]

Other problems are unique to the individual viral vector. 1) Widespread existing immunity for adenoviruses (35% in the U.S.[225] and 90% in the Ivory Coast[226]) hinders their efficacy. 2) Limited genetic material capacity for AAVs (~4.5 kb maximum)[227] severely restricts their potential. 3) An additional nuance for AAVs is the difference in observable outcomes between humans and small animal models. This requires researchers to conduct pre-clinical studies in non-human primates before clinical trials commence.[219]

4.1.2.2 Liposomes as Nanocarriers

Liposomes are the most common class of FDA-approved non-viral vectors.[228] Liposomes are spherical structures formed primarily of lipids. The hydrophilic heads are placed in contact with the aqueous environment whereas the hydrophobic tails are buried to give a lipid bilayer surrounding an internal aqueous core. A subclass of liposomes, known as lipid nanoparticles (LNP), are used to deliver nucleic acids. LNPs

are distinguished through their micellar sub-compartments where the genetic material resides (Figure 4.2).[228] Their lipid composition makes them highly biocompatible and affordable, thus avoiding the two major problems associated with viral vectors. In water, positively charged lipids spontaneously self-assemble to accommodate a large amount of negatively-charged nucleic acids.[228] The wide range of suitable lipids as well as the recent increased understanding of the relationship between lipid physicochemical properties and their effects on LNP circulation, uptake, and release of genetic payload have expedited the clinical translation of the first LNP-based gene therapy and vaccines.

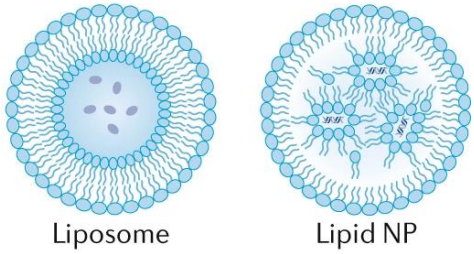


Figure 4.2. Comparison of different lipid-based nanocarriers

Liposome contains a lipid bilayer surrounding an aqueous core, whereas LNP contains micellar sub-compartments. Figure adapted from reference 228 with permission from Springer Nature.[228]

Non-modified LNP formulations are rapidly cleared *in vivo* through the reticuloendothelial system (RES).[229] Their removal is aided by the interaction between LNPs and opsonins. Opsonins are plasma proteins, such as antibodies, which can tag foreign materials for their removal by macrophages.[229] To minimise opsonisation, polyethene glycol (PEG) is included in LNP preparations to modulate its size,[228] and sterically and electrostatically block LNP interaction with opsonins, which could extend LNP half-life up to 45 h in humans.[230]

In addition, the size and charge of LNP also play a role in their bioavailability. Opsonisation decreases with decreasing LNP size.[231] Overall, largely charged LNPs

could be removed in minutes by the liver, or within an hour by the spleen[229,231], and are prone to activating the immune system.[228] However, LNPs with a diameter <10 nm also risk being filtered by the kidneys.[228] Addition of cholesterol also extends LNP blood circulation durations and membrane fusion.[232]

4.1.2.3 LNP-based Gene Therapy and Vaccines

Patisiran is the first LNP formulation for gene therapy. The drug uses LNP as a nanocarrier to deliver siRNA targetting the gene encoding transthyretin (TTR) to treat hereditary disease transthyretin-mediated amyloidosis (hATTR). Amyloidosis is a group of diseases characterised by the accumulation of misfolded proteins that result in amyloid fibrils. In hATTR, the TTR gene is mutated, and the aberrant TTR proteins form amyloid deposits in the heart and nerves; its victims suffer from polyneuropathies that can become fatal. Globally, hATTR affects up to 50,000 people.[233]

To encapsulate the nucleic acids, the lipids must be positively charged. As previously mentioned, such charges could prove cytotoxic by activating the immune system.[234] The use of an ionisable lipid with pKa ~6.4 enabled the association of lipids and siRNA at pH 4 whilst remaining neutral at physiological pH. Minor deviations from this pKa led to a rapid loss in activity.[234] The need for a precise pKa reflected the optimal charge to avoid systemic clearance and a positive charge to facilitate endosomal escape.[65,235] After a search of over 300 ionisable lipids, MC3 possessed the ideal pKa and was used in the final formulation.[234]

PEGylation also helped to compact the LNP to give an optimal 100-nm diameter. However, as PEG chains could interfere with target cell recognition, PEG-lipids with short acyl chains were chosen. In the plasma, these chains were shed to reveal naked LNPs that could interact with their targets.[234] These LNPs accumulated in the liver

through binding to apolipoprotein E, which targeted the particles to receptors on the surface of hepatocytes.[236] Patisiran repressed TTR levels by >80% compared to the placebo group and showed no safety concerns in humans.

The onset of COVID-19 accelerated the development of two LNP-based vaccines and saw their entering the market at unprecedented speed. In many ways, their LNP compositions and production resemble those of Patisiran. Notably, both employ an ionisable lipid that is formulated at pH 4 to give a positive charge for RNA complexation. At physiological pH, the LNPs become neutral; these charge alterations were desired for analogous reasons to Patisiran's LNP. Other components such as cholesterol and PEG were also included for stability, prolonged systemic lifetime, and to control the LNP size (~100 nm in diameter).[237] The rapid successes of these LNP vectors are a testament to the mechanistic understanding of LNP activity *in vivo* and signified the arrival of more LNP-based technology to come.

4.1.2.4 Outstanding Challenges for LNP-based Therapeutics

Notwithstanding their recent successes, components necessary in LNP production have been shown to activate the immune system. Furthermore, it's been reported that repeated injections of PEGylated LNPs have resulted in a shorter LNP lifetime on subsequent injections in animal models.[238,239] This is problematic when a repeated dosage of the drug is required.

Multiple injections of empty PEGylated LNPs in rats are known to trigger the production of anti-PEG immunoglobulin.[240] Worryingly, conventional liposome followed by PEGylated liposome injections have also raised anti-PEG antibodies.[241,242] Presumably, immunity was mounted when these patients had been exposed to PEG

through means other than PEGylated LNPs, suggesting that even the first dose of PEGylated liposomes might not be tolerated.

Some liposomal systems could activate the complement system, which forms part of the innate immune system. Once triggered, hypersensitivity reactions including anaphylaxis, rash, swollen eyes and extremities, and even cardiopulmonary distress could ensue.[243] These reactions have been reported for 2-45% of patients receiving experimental and clinically approved liposomal drugs.[243] Likewise, ionisable lipids have also been linked to heightened levels of inflammatory cytokines and reactive oxygen species.[244,245] Liver and lung injuries were detected in rodents following LNP administration, which was proposed to be derived from the LNP formulation and raised level of inflammatory factors.[245–247]

LNPs are also known to have the highest accumulation in the liver and spleen.[248] Thus, while Patisiran proved to be effective against TTR produced in the liver, it remains to be seen whether LNPs could target extrahepatic tissues. Finally, it has been shown that the lipid mixing method used to prepare LNPs could only encapsulate up to 6 kbp plasmid cargo.[249,250] This size dependency is severely restrictive and limits the use of LNPs to the delivery of small nucleic acids, and precludes their adoption for plasmids.

4.1.2.5 Transferrin-mediated Gene Delivery

Transferrin (Tf) is an iron-binding protein.[251] It supplies cells with iron by associating with the transferrin receptor (TfR). The Tf-TfR complex then enters cells via clathrin-coated pits endocytosis.[251] As the pH within the endosome becomes more acidic (~5.5-5.0), ferric Tf releases the iron ions, which can freely exit the endosome. Apo-Tf is recycled back to the extracellular environment, leaving TfR free for further binding with ferric Tf.[251] TfR is often overexpressed in cancer cells, making Tf an attractive

means to selectively deliver drugs or nucleic acids to tumour targets.[251–253] A therapeutic cargo covalently linked to Tf could repeatedly enter cells and bypass the endosomal degradation via the Tf-recycling pathway.[254] Moreover, Tf-mediated endocytosis of naked plasmids occurred within 1 min, which further enhanced DNA survival by avoiding circulating DNase.[254]

Although Tf has been conjugated to small molecule drugs and liposomes, efficient liposomal delivery requires proprietary lipid preparation, in addition to the previously mentioned problems associated with LNPs, thus their wide adoption may not be feasible. Niisu *et. al.* achieved naked plasmid delivery with Tf by exploiting the potent binding between streptavidin and biotin. The plasmids and Tf were linked through non-covalent interactions mediated by streptavidin and biotin (Figure 4.3). To biotinylate transferrin and the plasmid, they were reacted with commercially available NHS-SS-biotin and photoprobe-SS-biotin, respectively. Using this facile approach, the authors showed that Tf-plasmids were transfected exclusively into cells overexpressing TfR. Transferrin-mediated transfection outperformed standard lipofection and retroviral transfection, but not as well as adenoviral transfection, but the last is compromised by its safety profile as mentioned in Section 4.1.2.1. Significantly, the construct also preferentially accumulated in the mice's tumour for at least 1 week.[254]

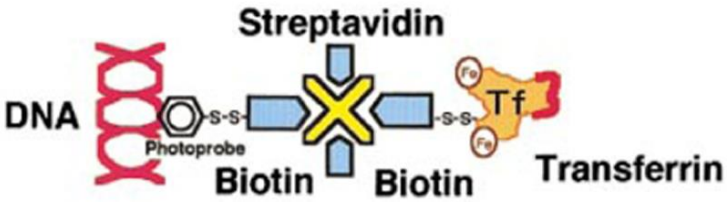


Figure 4.3. Niisu *et. al.*'s approach to attaching Tf to naked plasmids

Transferrin was bound to naked plasmids through biotin and streptavidin interaction. Figure adapted from reference 254 with permission from Wiley Online Library.[254]

Ganciclovir is an antiviral prodrug. It is activated after repeated phosphorylation events and resembles deoxyguanosine triphosphate (dGTP). As such, it can compete with dGTP for incorporation into DNA, thereby inhibiting DNA synthesis.[255] Co-administration of the construct bearing a plasmid encoding the suicide gene herpes simplex virus thymidine kinase (HSV-TK), which phosphorylates ganciclovir, alongside this prodrug into cancerous mice greatly boosted their survival relative to controls without transferrin.[254] This confirmed transferrin was necessary to deliver the plasmid into cells.

4.1.2.6 Albumin-mediated Gene Delivery

Albumin is the most abundant plasma protein in humans.[256] It acts as an antioxidant and plays a crucial role in transporting important substances such as cholesterol, lipids, amino acids, etc. for cell growth and survival.[257] Caveolin-mediated endocytosis is generally accepted as its mode of entry into cells; however, studies have reported other possible mechanisms for its cell uptake, and the actual mode depends on the cell type and exact conditions.[257] Whichever means adopted for entry, once inside the cells, albumin ends up in the early endosomes. It is then either recycled back to the cell surface or processed via lysosomal degradation.[257]

Caveolae are a dimer of caveolin proteins. They form plasma membrane-derived pouches used for the internalisation of cargo and are present in many cell types. Albumin binds to its receptors gp60 and SPARC, which triggers a signalling cascade that ultimately results in the phosphorylation of the caveolae components and invaginations.[258,259] Caveolin, gp60, and SPARC are all upregulated in cancer cells, making albumin an opportunistic carrier to deliver cancer therapeutics.[259,260] Similar to Tf, the use of albumin is advantageous since it is biocompatible and poses minimal

risk of immunogenicity, has a long blood circulation,[261], and can be metabolised *in vivo* to give harmless degradation products.[262]

The abundant presence of positively charged amino acids in albumins enables effortless electrostatic compaction of plasmids without the need for elaborate modifications.[262] Chen and colleagues employed albumin to deliver plasmids by cationisation of human serum albumin (CHSA) by amide coupling with ethylene diamine. CHSA could readily associate with plasmids and a nuclear localisation signal to give a construct that could rapidly enter cells. Their *in vitro* and *in vivo* transduction in two different cell lines and mice was well-tolerated. Their *in vivo* activity in these hosts was also higher compared to plasmids transfected with lipofectamine. Dye-labelled plasmids revealed that cellular uptake occurred within the first half hour, and fluorescence gradually increased up to 4 h.[262]

Co-treatment with various cell entry inhibitors informed albumin's mode of entry in these particular cell lines. Conditions resulting in the disruption of metabolic activity or endocytosis resulted in major decreases in cell uptake, confirming that an energy-dependent endocytosis process was at play.[262] Whilst both the HepG2 and A549 cell lines used clathrin-dependent endocytosis, only A549 employed macropinocytosis, confirming that the mechanism for albumin-mediated entry is cell line-dependent.[262] Nevertheless, CHSA possessed more positive charges on its surface than albumin, so the actual mechanism for cell uptake may differ from that of conventional albumin.

4.2 Chapter Overview

This chapter expands the utility of our caging strategy by applying it to control the CMV promoter expressing mNG. A nickase and exonuclease were confirmed to act non-specifically. Changing to the less reactive λ Exonuclease enabled the mNG plasmid

to be resynthesised. Caging only the TATA box was not enough to repress expression. Additional steric blocking of surrounding promoter elements was needed to regulate promoter activity. Gene expression was efficiently repressed and conditionally activated through UV irradiation. Finally, plasmid self-delivery was attempted through the non-covalent linkage of transferrin or albumin to the plasmids. Construct aggregation prevented cell entry, and strategies to overcome this issue are suggested.

4.3 Nickase promiscuity revisited

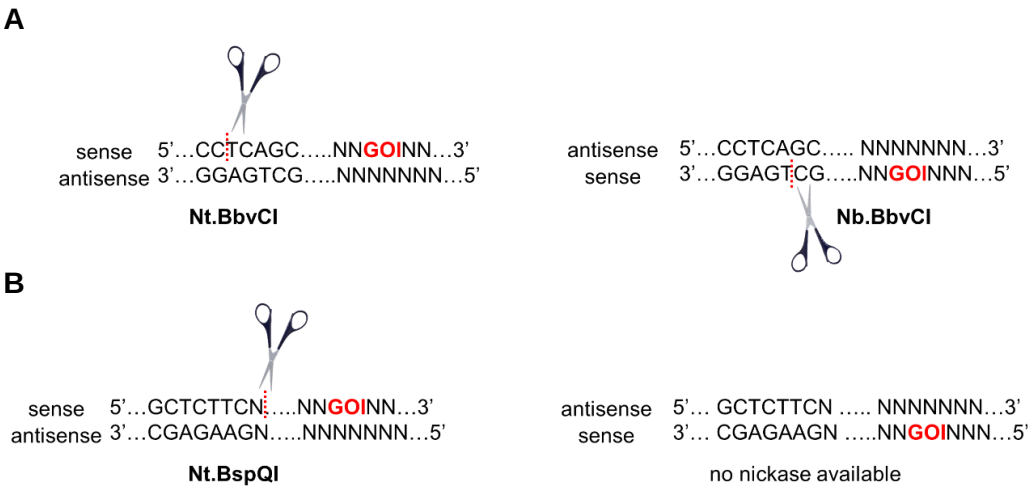


Figure 4.4. Schematic illustrating directionality of nickase recognition site

A. When nickase for either strand exists, the choice of nickase simply depends on which strand is the sense strand. B. When only one nickase exists, which could only cut the antisense strand (right), to apply the LA-plasmid caging strategy, the recognition site had to be inverted.

The ability to control plasmid expression within mammalian cells would greatly extend the scope of the LA-plasmid approach beyond *in vitro* applications. The CMV promoter is one of the most widely used promoters for mammalian expression. It has been well-studied and its regulatory elements are also conveniently placed close to one another, which made it a suitable choice to apply our caging strategy. A mammalian codon-optimised plasmid encoding mNG under the control of the CMV promoter (referred to

as CMV_mNG plasmid) was chosen as it already contained a recognition site for the nickase Nt.BspQI.

For many nickases, two variants exist that cut either the ‘top’ (t) or ‘bottom’ (b) strand, where the ‘t’ and ‘b’ designations are dictated by the directionality of the recognition site. In these cases, one could simply choose the nickase that would nick the sense strand (Figure 4.4A). However, for some recognition sites such as that for BspQI, only one variant existed, e.g. Nt.BspQI. Suppose the recognition site for this nickase was located on the antisense strand (Figure 4.4B, right), to apply the LA-plasmid caging, the recognition site would have to be inverted, which would necessitate a cloning step. The ability to cage either strand of the plasmid would greatly enhance the flexibility of this approach and obviate the need to clone should the nickase recognition site be located on the antisense strand.

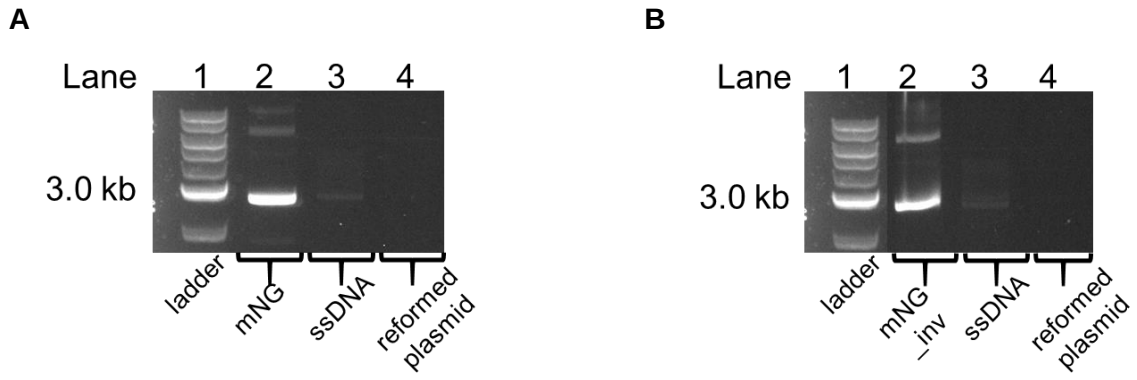


Figure 4.5. Initial NEEL attempts for the CMV_mNG and mNG_inv plasmids

Nicking CMV_mNG and mNG_inv with Nt.BspQI followed by NEEL did not give the reformed plasmids A. CMV_mNG B. mNG_inv.

To investigate whether the antisense strand could also be caged, the BspQI recognition sequence was inverted to give mNG_inv (Tables 6.14-6.17). If both strands proved equally amenable for caging, then the strategy could be used to

prepare LA-plasmids regardless of the directionality of the nickase's recognition site. CMV_mNG and mNG_inv were both nicked with Nt.BspQI and resynthesised using NEEL. In both instances, NEEL failed to give the reformed plasmids (Figure 4.5).

Previously, when NEEL could not resynthesise plasmids, the problem was attributed to nickase promiscuity (Section 2.5). The promiscuous nickase in that instance was Nt.BbvCI. A comparison between Nt.BbvCI and Nt.BspQI had shown the latter to cut specifically. Thus, it was unexpected that NEEL with Nt.BspQI did not work this time. The very faint ssDNA bands in Figure 4.5 suggested the problem lay upstream of the extension and ligation reactions, which could either be the nickase or exonuclease step.

To verify if the nickase was the issue, different mammalian expression vectors were treated to a range of nickases with different recognition sequences. In Appendix Figure A7, the plasmids either had a single recognition site or 2 recognition sites that were located on the same strand. For each of these scenarios, after the exonuclease reaction, there should be a band for the ssDNA. As previously demonstrated for CMV_mNG and mNG_inv, their reactions with Nt.BspQI followed by exonuclease treatment gave little to no ssDNA. The mV_N1 plasmid encoded the mVenus fluorescent reporter and a multiple cloning site (MCS) immediately upstream of mV so a GOI could be cloned to create a fusion protein with mV as the reporter. After exonuclease, no ssDNA band was detected when either Nt.BspQI or Nb.BsrDI was employed as the nickase. The ChR2-EYFP plasmid expressed Channelrhodopsin-2 fused to the enhanced yellow fluorescent protein. ChR2 is a light-controlled ion channel that features in photoreceptors and coordinates green algae response to light.[263] Once again, when it was nicked with Nb.BsmI and reacted with an

exonuclease, ssDNA was not observed. These reactions indicated that these nickases all nicked unspecifically.

In Appendix Figure A8, the plasmids did not have a recognition site for the nickases being studied. Therefore, there should be no change in band positionings after reactions with nickase and exonuclease. All plasmids treated with Nb.BbvCI remained uncut, indicating this nickase did not have aberrant nickase activity. On the other hand, Nt.BbvCI did not cut either CMV_mNG or mNG_inv, but did exhibit low activity against mV_N1, and nicked ChR2-EYFP completely.

It became clear that nickases may not be regarded as wholly specific. Previous work with the mC and mV plasmids (Section 2.5) had shown that Nt.BspQI could cut specifically but according to the results shown here, this was not the case (Appendix Figure A7A). Indeed, except Nb.BbvCI, all nickases tested with the nick-exonuclease reactions thus far had caused unexpected degradation of the ssDNA template. Therefore, before applying NEEL, the nickase should be assessed for its ability to nick specifically regardless of the number and directionality of its cognate recognition site.

As CMV_mNG and mNG_inv were compatible with both Nt.- and Nb.BbvCI, the BbvCI recognition sequence was cloned into this plasmid to give mNG_BbvCI (this was the same plasmid as the Sp6_mNG plasmid used in Section 3.10, Tables 6.14-6.17). This time, minimal ssDNA was observed, and NEEL only gave faint extended and ligated bands (Figure 4.6A). To determine whether the problem was due to the manufacturer, the reactions were re-tested using different enzyme manufacturers. Nt.Bpu10I and Nb.Bpu10I have the same recognition sequences as Nt.- and Nb.BbvCI, respectively, but were supplied by Thermo Fisher instead of New England Biolabs (NEB). Indeed, neither of these enzymes could cut CMV_mNG and mNG_inv, which lacked their

recognition sequence (Appendix Figure A9). When NEEL was re-attempted on mNG_BbvCI with these nickases, the desired plasmids were still not formed (Figure 4.6B).

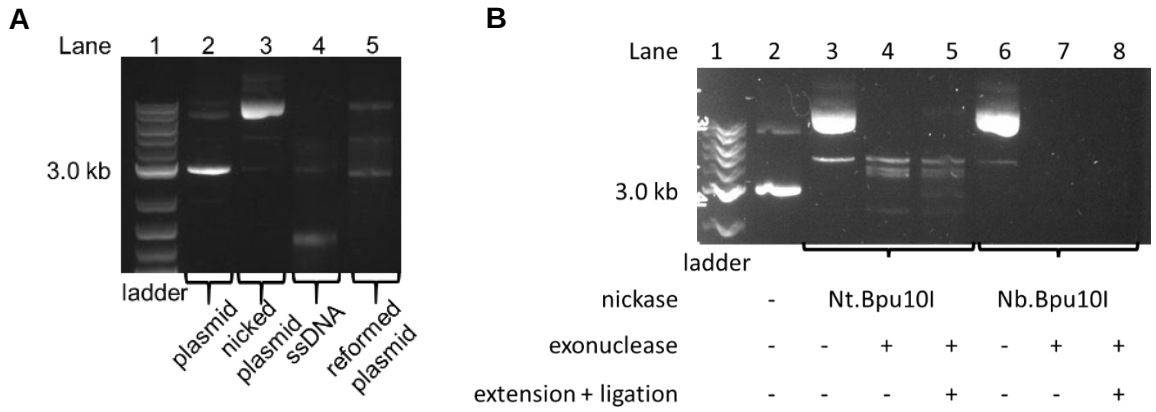


Figure 4.6. NEEL with the mNG_BbvCI plasmid

The BbvCI recognition site was cloned into CMV_mNG to give mNG_BbvCI. A. When this plasmid was nicked with Nt.BbvCI or B. Nt- and Nb.Bpu10I and resynthesised, minimal extended and ligated bands were observed.

As mNG_BbvCI was larger than mV (4.8 vs 2.9 kb), the polymerase may not have fully extended through the plasmid. To rule out problems associated with the annealing, extension, and ligation steps, NEEL was repeated with higher DNAP and ligase concentrations, or different primers (Figure 4.7) but none of these changes could give the reformed bands. Considering the only difference between mNG_BbvCI versus CMV_mNG and mNG_inv plasmids was the BbvCI recognition sequence (5'-CCTCAGC-3') and the latter plasmids were not nicked, two possibilities arose. Either the nickase was cutting both strands of mNG_BbvCI at the BbvCI recognition site, which was unlikely, or the problem was in the exonuclease step.

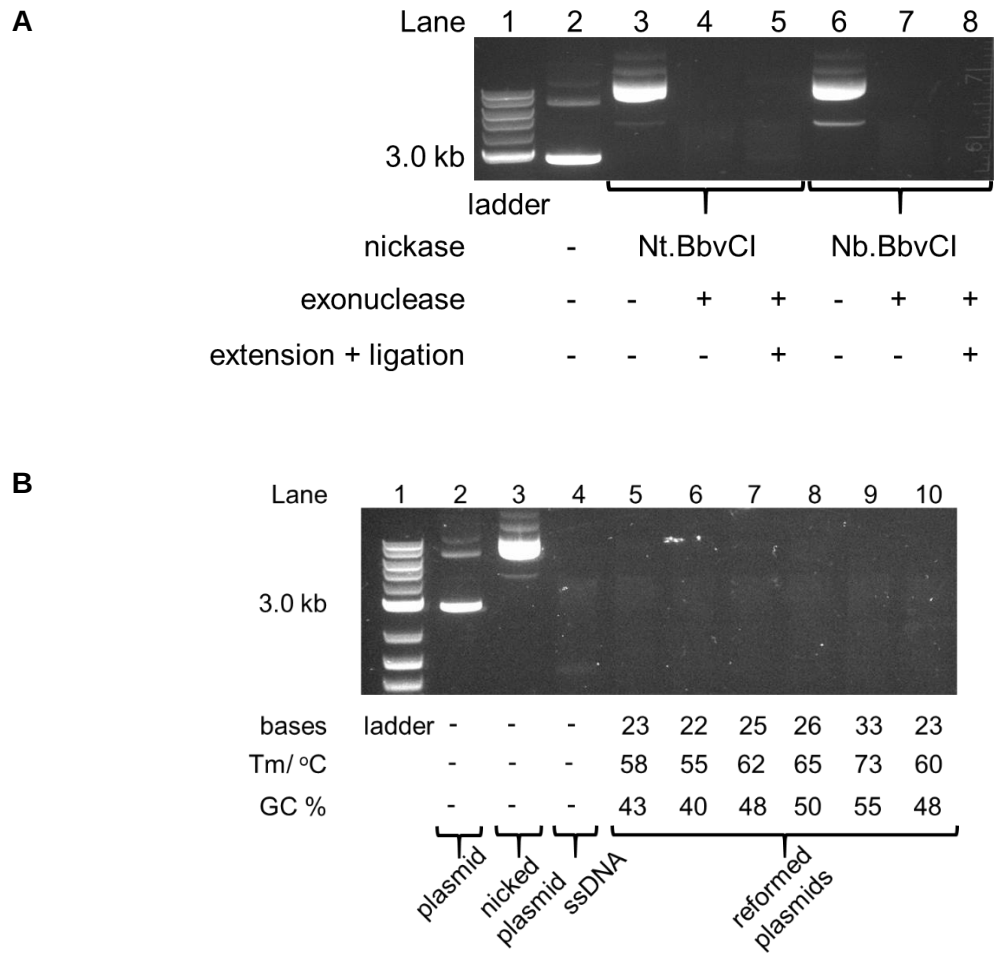


Figure 4.7. NEEL reattempted for mNG_BbvCI plasmid with higher enzyme concentrations or different primers

A. DNAP and ligase concentrations were doubled in extension and ligation reactions. B. Different primers with varying lengths, melting temperatures, GC contents, and annealing positions were used to resynthesise plasmids.

4.4 Lambda Exonuclease

Thus far, Exonuclease III had been used to generate the ssDNA template. This enzyme has 3'-5' exonuclease activity and could accept most linear DNA or nicked plasmid as substrate. Another suitable exonuclease to generate ssDNA from a nicked plasmid is λ Exonuclease. It cleaves in the 5'-3' direction with a similar substrate profile

compared to Exonuclease III but has a strong preference for 5′-phosphates over unphosphorylated substrates.

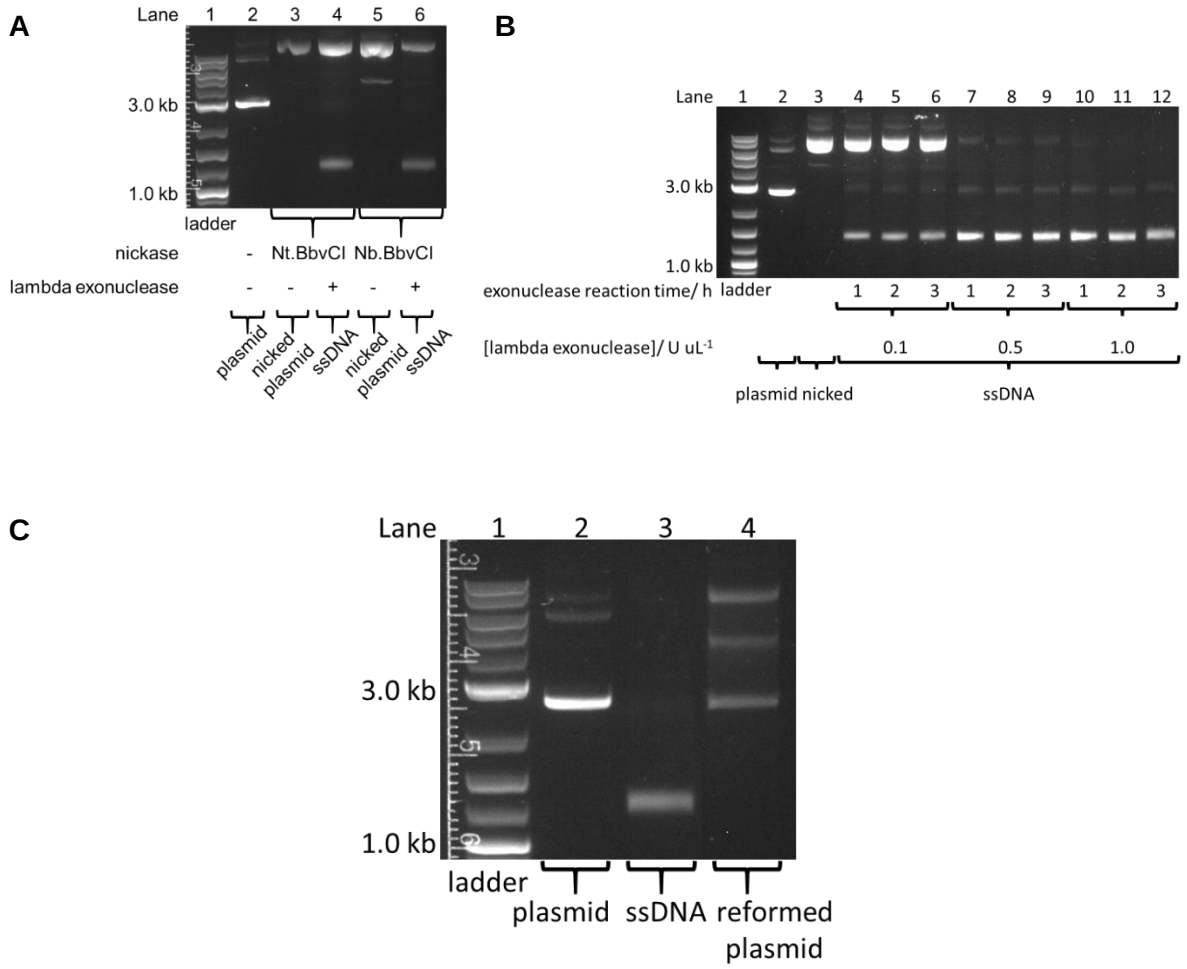


Figure 4.8. mNG_BbvCI plasmid resynthesis with λ Exonuclease

A. Using the manufacturer's conditions, λ Exonuclease could generate appreciable quantities of ssDNA despite the reaction not having gone to completion. B. Enzyme concentration and reaction time were varied C. ssDNA obtained after reaction with λ Exonuclease could give NEEL products.

After nicking mNG_BbvCI with either BbvCI nickase, the nicked strand was removed with λ Exonuclease according to the manufacturer's protocol (Figure 4.8A). Despite not having gone to completion, the ssDNA produced was much more intense relative to any attempt thus far for this plasmid. After optimisation, an enzyme concentration of 1.0 U μL^{-1} and a 3 h reaction time could convert all nicked plasmid into ssDNA

(Figure 4.8B). Note the faint band co-migrating with supercoiled plasmid was concatemeric ssDNA and not leftover native plasmid, as there was no detectable supercoiled plasmid after the nickase reaction. When this ssDNA was used as the template, NEEL products were obtained (Figure 4.8C).

Although the different exonuclease behaviours were not studied further, some possible explanations may be inferred from these experiments. Random nicks could be introduced into either strand of the plasmid through mechanical shearing (pipetting, vortexing) or heating. Nicked plasmids would have undergone these processes by the time they were reacted with exonuclease. As these nicks would likely occur across the phosphate-sugar backbone, the non-discriminatory activity of Exonuclease III could very well cause the degradation of both nicked strands. λ Exonuclease's preferential activity against substrates with 5'-phosphates would leave 5'-unphosphorylated substrates intact. However, Appendix Figure A9 showed that the majority of unreacted plasmids were not nicked and remained supercoiled, so this was unlikely responsible for Exonuclease III's aberrant activity.

That nickase would cut both strands was also not possible, as in this case, both exonucleases would degrade the whole plasmid, but this was only observed for Exonuclease III. Thus, Exonuclease III was likely removing the ssDNA after the initial nicked strand had been degraded. This activity did not occur when dsDNA was not nicked. Furthermore, this behaviour was substrate dependent, as it could still furnish mV ssDNA in Chapter 3.

Exonuclease III also possesses 3′-phosphatase, RNase H, and apurinic/apyrimidinic (AP) endonuclease activities.[264] The last role is involved in the DNA base excision repair pathway to correct mismatched or damaged nucleotides. After DNA glycosylase

removes the nitrogenous base, Exonuclease III can nick the AP site to give 5'-deoxyribose 5-phosphates.[264] However, such AP endonuclease/nickase activity is not relevant here, as there is no base mismatch and DNA glycosylase involved. To the best of my knowledge, this peculiar activity denoted by the removal of a covalently-closed ssDNA by Exonuclease III has not been reported in the literature.

Concerning nickase's promiscuity, the only nickase that could be said to nick unspecifically was Nt.BbvCI, as the plasmids which it had nicked did not contain its target site (Appendix Figure A11B and Figure 2.8B). The other examples did contain nicking sites for the nickase tested, and ssDNA degradation might be due to Exonuclease III as described above. λ Exonuclease appeared to have none of this aberrant activity. It was used to resynthesise the mV and tac plasmids in Chapter 3, and the mNG_BbvCI plasmid for the rest of this Chapter, amongst other vector constructs without issue. Using this optimised plasmid resynthesis strategy, 2 nmol of the amine primer typically gave 20-25% of the HPLC-purified PCB primer, whilst 75 µg of the native plasmid could give an overall ~35% yield of the PCB plasmid.

4.5 Caging the TATA Box

Deiters *et. al.* have previously controlled plasmid expression from the CMV promoter by caging the TATA box with three 6-nitropiperonyloxymethyl (NPOM) groups (Section 1.5.2, Figure 1.9A).[173] However, compared to their caging strategy, the LA-plasmid approach has several advantages. To install the NPOM functionality, a double-nickase reaction was performed. The nicked fragment was then removed, and the oligo containing a similar sequence with 3 NPOMs was attached and re-ligated.[173] This necessitated cloning of the nickase sites at positions adjacent to where the modifications would be, which could affect the promoter's regulatory role. The NPOM

caging group was also situated on the WCF face, which disrupted hybridisation. To obtain the oligo containing NPOMs, they relied on modified phosphoramidites, which may not be accessible to many research groups.

The LA-plasmid strategy as presented here could bypass the above issues. The nickase site could be situated anywhere on the sense strand of the plasmid. Meanwhile, the caging groups were positioned on the non-WCF face, and so did not disrupt hybridisation, and the amine oligos were commercially available. The same 3 thymines that were caged by Deiters *et. al.* were used as the starting point to demonstrate our approach (Table 6.19). Thus, a primer containing 3 amino-C6-dT modifications annealing at the TATA box was reacted with PCB-NHS ester to obtain a 3-PCBs oligo (Appendix Figure A10, Table 6.20).

Before assessing mNG LA-plasmid, transfection conditions were established. Lipofectamine^{TM} 3000 (L3000) is a cationic lipid transfection reagent that is commonly used to deliver nucleic acids into eukaryotic cells. It acts by forming a complex with negatively charged plasmids. This neutralises the overall charge and enables the complex to traverse the negatively charged plasma membrane.

L3000 equivalents and timing of analysis (days post-transfection, dpt) were assessed. 100 ng/well of the native mNG_BbvCI plasmid was transfected into human embryonic kidney 293T cells (HEK293T) and analysed qualitatively via epifluorescence microscopy (Figure 4.9). Increasing the amount of L3000 was beneficial to transfection; similarly, analysis 2 days after transfection revealed much higher fluorescence compared to 1 dpt, with minimal improvement in expression thereafter. Subsequently, analysis was carried out at 2 dpt to minimise experimental time, and 2

equivalents of L3000 were used to avoid potential cell toxicity conferred by the transfection reagent.

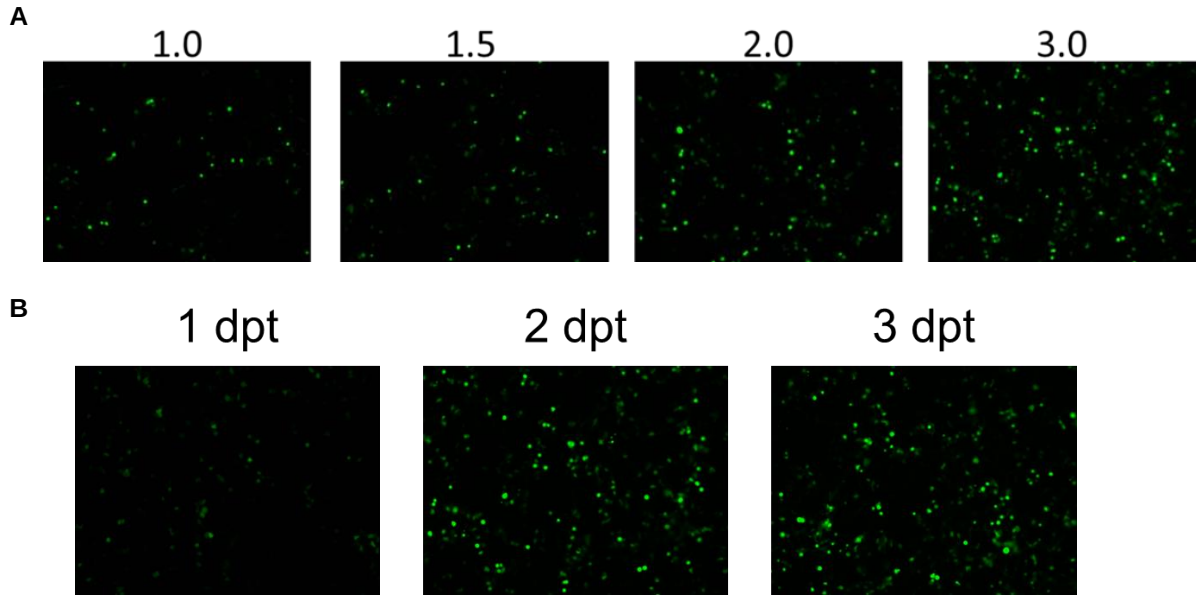


Figure 4.9. Establishing transfection conditions

100 ng of the mNG_BbvCI plasmid was transfected into HEK293T cells to determine transfection conditions. A. L3000 equivalents (L3000 volumes in µL per 100 ng plasmid) were varied and showed increasing transfection efficiency with higher L3000 amounts. B. The time of analysis was assessed, and each well was transfected with a L3000:DNA ratio of 3:1; a 2 dpt was chosen for subsequent experiments. Each image was representative of technical duplicates (n = 2).

Amine plasmids containing 3 amino-C6-dT either on the sense or antisense strand were transfected into HEK293T cells. As amine plasmids represented the best ON state possible, ideally, they should exhibit fluorescence close to the native plasmid transfection. As this plasmid does not contain origins of replication that would enable plasmid replication in mammalian cells, the plasmids do not get replicated and will become increasingly diluted with successive rounds of cell division. However, the experimental timeframe between transfection and analysis is only 2 days and is commonly used in literature to assess photocaging,[173] hence it was expected this dilution effect is not significant and the majority of cells should exhibit fluorescence.

Cells transfected with plasmids where the modifications were on the antisense strand performed much worse compared to the condition where the amines were on the sense strand (Figure 4.10). Subsequently, only the sense strand was caged.

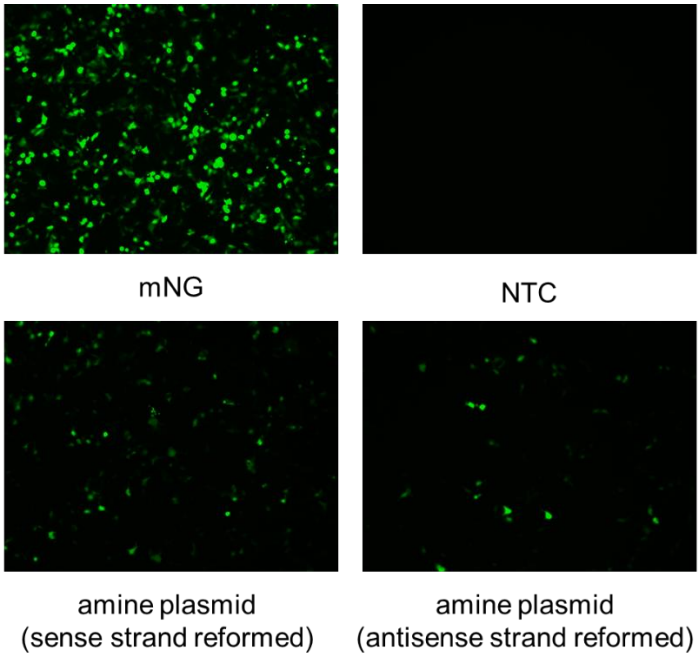


Figure 4.10. Comparison between mNG_BbvCI plasmid expression where either the sense or antisense strand was resynthesised

Amine plasmids with modifications on either the sense or antisense strand were transfected into HEK293T cells and their fluorescence was compared to the native plasmid. Each image was representative of technical duplicates (n = 2).

With only 3 streptavidins, LA-plasmid bands produced the same migration as the native plasmid on the gel, although faint smears could be observed for WTS LA-plasmids due to plasmid crosslinking (Figure 4.11A). As the number of streptavidin was reduced, the UV irradiance was also lowered to 1.1 mW cm^{-2}. MVS LA-plasmids still showed high expression (Figure 4.11B). WTS LA-plasmids achieved better repression relative to MVS, consistent with observations thus far for the two streptavidin. The mild UV conditions employed did not increase expression, however.

This UV intensity was still not enough to uncage 3 PCBs and was reminiscent of previous experience with uncaging the T7 promoter.

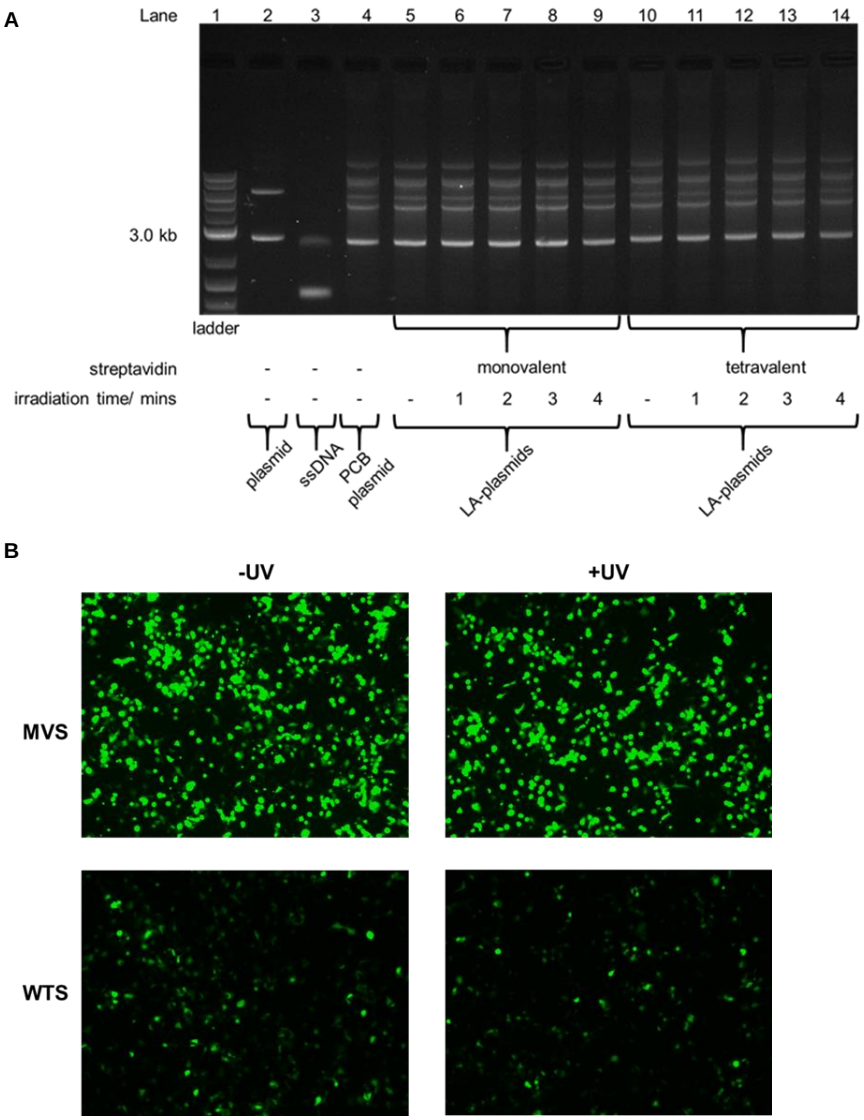


Figure 4.11. Transfection of mNG_BbvCI LA-plasmids containing 3 modifications

A. LA-plasmids prepared from a 3-PCBs oligo were irradiated at 1.1 mW cm^{-2} for different durations. B. 100 ng of LA-plasmids from (A) were transfected into HEK293T cells. +UV conditions shown used an irradiance of 1.1 mW cm^{-2} for 4 mins. Each image was representative of technical duplicates (n = 2).

4.6 Caging the Initiator and B Recognition Elements

Unlike Deiters' experience with caging the TATA box, three caging groups could not efficiently control expression, despite the much bulkier streptavidin compared to

NPOMs. This was likely due to the placement of the modifications on the non-WCF versus WCF face.

Mammalian transcription involves the assembly of the transcriptional PIC, which in turn depends on the association between various transcription factors and RNAP II with the TATA box and other nearby regulatory elements (Section 4.1.1, Table 4.1). The proximity of these elements within the CMV promoter was fortuitous, as this enabled their simultaneous caging to attain a better OFF state.

Two primers containing either 8 or 9 amino-C6-dT caging the upstream and downstream BRE, TATA box, and Inr elements were purchased (Table 6.19). For the 8 amines-modified oligo, HPLC revealed multiple peaks (Appendix Figure A11). After the reaction with PCB-NHS ester (Table 6.20), MS deconvolution of the peak at 8.80 mins did not yield a mass for the product. Denaturing PAGE showed 2 bands for the starting material and product. The 9 amines-modified oligo also exhibited 2 peaks in the LC run, which converged to a single peak after the reaction with PCB-NHS ester (Appendix Figure A12, Table 6.20). MS confirmed this to be the correct 9-PCBs oligo. Denaturing PAGE of the amines and PCBs oligo showed multiple bands for the starting material, but these again converged to a single band after PCB incorporation.

The additional signals in the starting material chromatograms probably represented incomplete deprotected products from solid-phase DNA synthesis. The mass differences between the correct amine oligos and the anomalous peaks were 69 for both the 8 and 9 amines oligos which did not correspond to any common MS adducts. As the DNA synthesis was proprietary, the identity of the additional functional group could not be deduced. Nevertheless, these PCB oligos successfully generated the respective LA-plasmids (Figure 4.12).

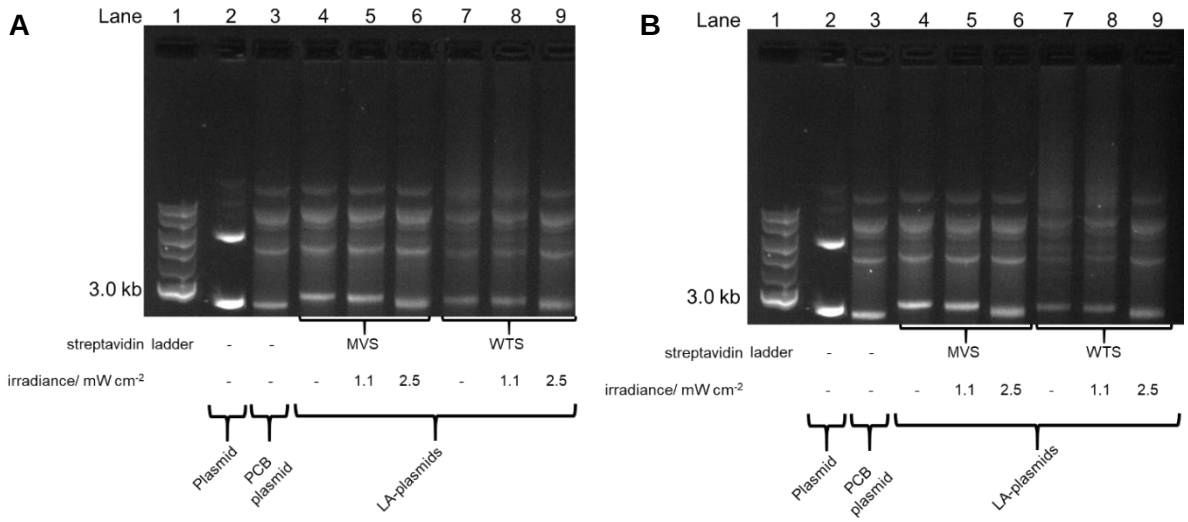


Figure 4.12. mNG_BbvCI LA-plasmids containing 8 or 9 modifications

LA-plasmids prepared from either A. 8-PCBs or B. 9-PCBs oligos were irradiated at 1.1 (4 mins) or 2.5 mW cm^{-2} (10 mins).

As the previous UV settings did not seem to fully cleave PCBs (Figure 4.11), these conditions were compared against the originally optimised UV conditions. Increasing UV intensity and duration showed a stepwise band shift pattern (Figure 4.12). LA-plasmids containing 8 modifications were leakier than plasmids with 9 modifications (Figure 4.13, -UV conditions), and consequently, the difference between ON/OFF states was more obvious for the latter. This was unsurprising, as the 9 amines-modified oligo contained modifications at the Inr that were not present in the 8 amines oligo. The additional modifications were found at the crucial +3 and +5 positions, which were strong regulators of transcriptional activity.[217] Thus, studies were continued with the primer containing 9 modifications.

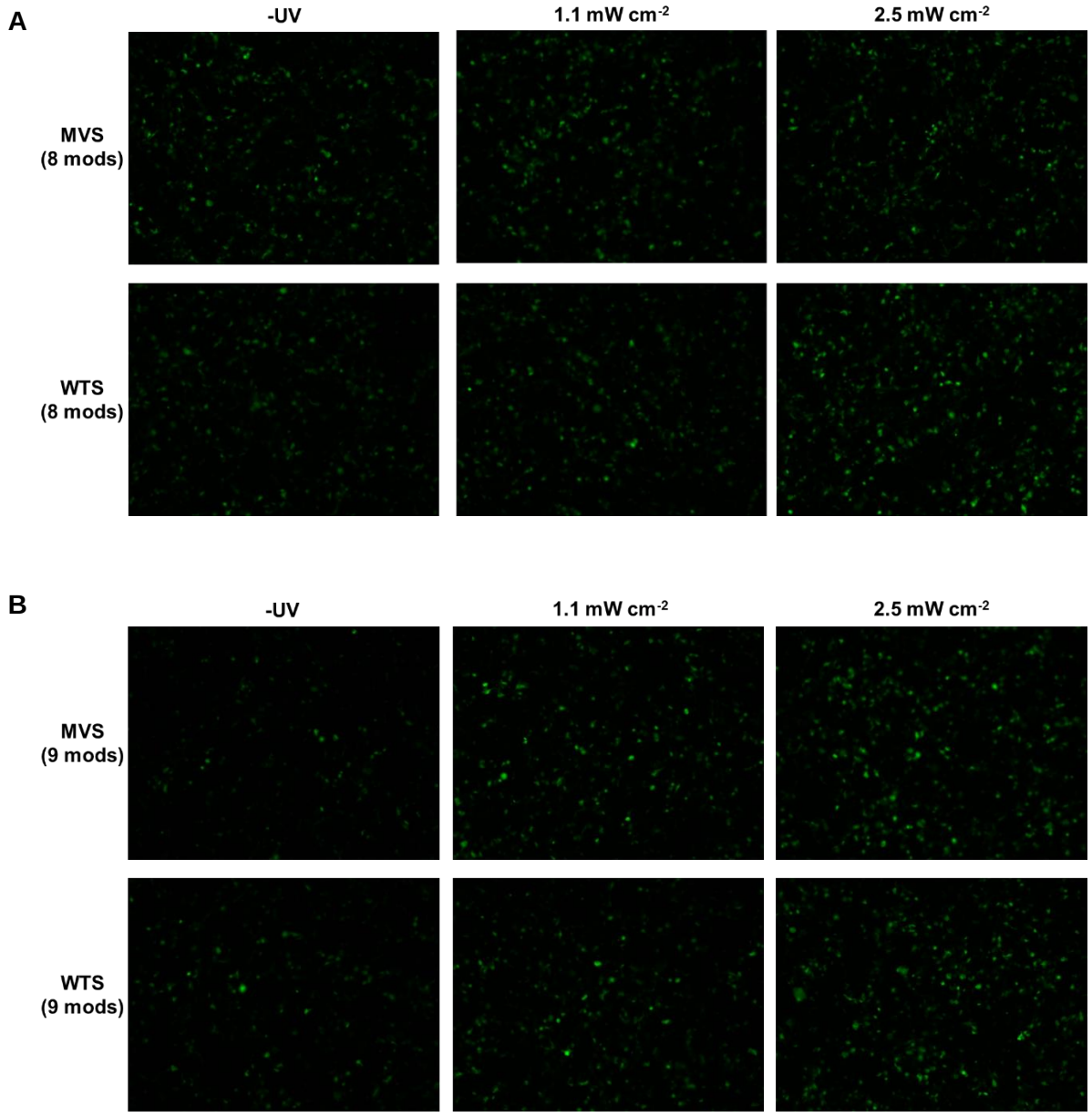


Figure 4.13 Uncaging mNG_BbvCI LA-plasmids containing 8 or 9 modifications

100 ng of LA-plasmids containing A. 8 or B. 9 modifications were transfected into HEK293T cells and uncaged with the indicated UV conditions. Each image was representative of technical duplicates (*n* = 2).

Although the number of amino modifications for the CMV and T7 promoters was similar (9 and 7 amino-C6-dT, respectively), the same UV conditions may not be transferrable as the reaction conditions differed. For T7, cell-free reactions were performed in 10-µL reaction volumes. For CMV, transfected cells were grown in

approximately 190-200 µL culture media. The higher reaction volumes in the latter would partially absorb the UV before it could reach the cells adhered at the bottom, thus lowering the actual amount of UV absorbed by PCB.

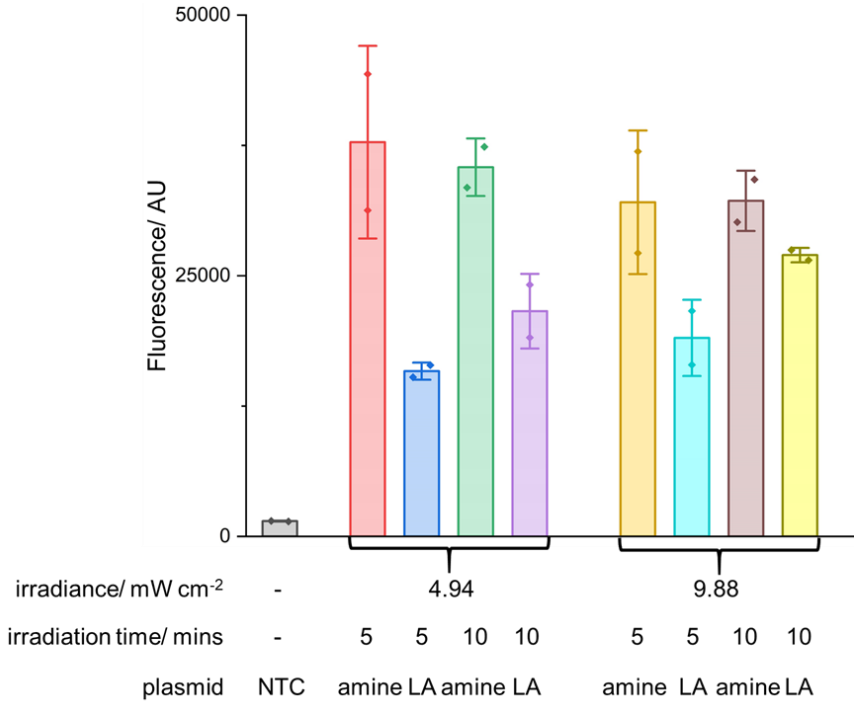


Figure 4.14 Optimising UV irradiation conditions with mNG_BbvCI amine and LA-plasmids

100 ng of amine or MVS LA-plasmids was transfected into HEK293T cells and irradiated with varying UV intensities and durations. Fluorescence was measured on the plate reader. Error bars showed standard deviations for technical duplicates (*n* = 2).

UV conditions were re-optimised with amine plasmids and MVS LA-plasmids (Figure 4.14). Higher irradiance did result in higher expression from LA-plasmids, but this also damaged the expression from amine plasmids. Cell viability assay showed that there was minimal cell damage at the conditions that gave the highest ON state: 9.88 mW cm^{-2} for 10 mins (Appendix Figure A13). These UV conditions were still milder or comparable to those used in the literature.[173,265] Given this condition would yield the

best photomodulation efficiency and was not damaging to cells, it was chosen for the next round of transfection.

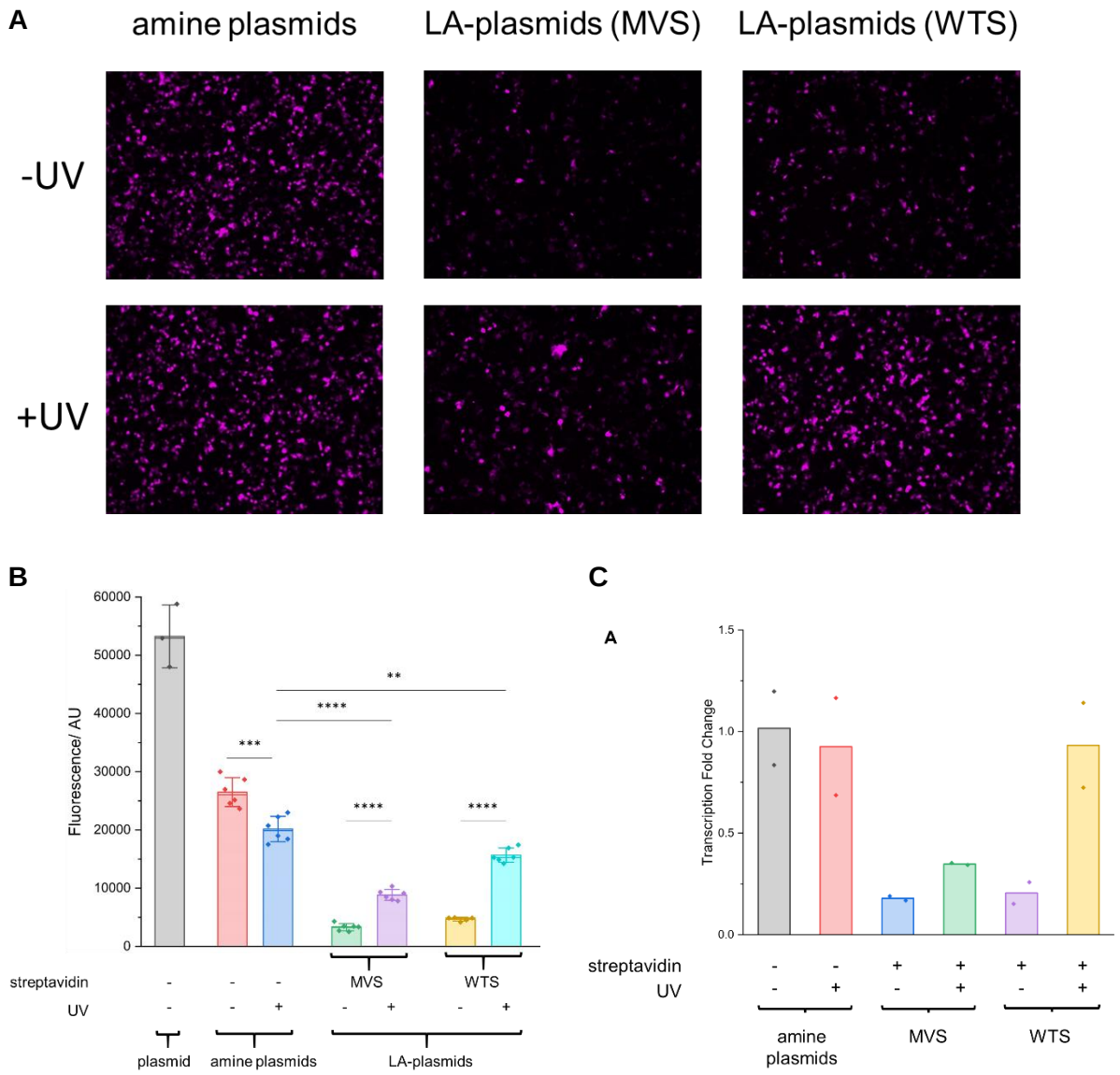


Figure 4.15. mNG_BbvCI LA-plasmids were uncaged with optimised UV settings and expression of mNG was assessed

100 ng of LA-plasmids were transfected into HEK293T cells and analysed by A. microscopy; images were representative of the respective conditions. B. fluorescent plate reader; conditions were background-subtracted before plotting and analysis. Error bars showed standard deviations. Two-tailed, unpaired Student's t-test was performed. ns = non-significant ($P > 0.05$); **$P \leq 0.01$; ***$0.0001 \leq P \leq 0.001$ and ****$P \leq 0.0001$. C. RT-qPCR; ΔΔCt method was used to calculate the transcription fold change of the expression of mNG relative to the 'amine plasmid -UV' control. For (A-C), repeats comprised of technical triplicates for the 'native plasmid' and 'NTC' conditions ($n = 3$), and biological duplicates of technical triplicates for all other conditions ($n = 6$) except for the 'WTS -UV' condition ($n = 5$ due to sample loss).

PCB primer was uncaged with the higher UV parameters and the fully-cleaved oligo formed the major band (Appendix Figure A14). This was in contrast to the multiple bands observed in Appendix Figure A12D when milder UV settings were used. Cells transfected with LA-plasmids were analysed for transcriptional activity with reverse transcription quantitative real-time PCR (RT-qPCR), transcription was repressed by up to 96% relative to the native plasmid. When the newly optimised UV conditions were applied to uncage LA-plasmids, an increase in transcription for both MVS and WTS plasmids was observed. However, WTS LA-plasmids produced considerably higher transcripts and was comparable to the 'amine plasmid +UV' condition (Figure 4.15C).

This trend manifested in protein expression, as demonstrated by both microscopy and fluorescent plate reader (Figure 4.15A and 4.15B). Fluorescence was reduced by ~94 and 91% for MVS and WTS, respectively. After photorelease, expression increased by 2.7- and 3.3-fold. Altogether, the LA-plasmid caging strategy could be used to effectively control expression from the ubiquitous T7 and CMV promoters. In both cases, the expression could be restored to levels similar to the 'amine plasmid +UV' standard by irradiation with mild UV conditions.

4.7 Self-delivery with Biotinylated Transferrin and BSA

Plasmid transfection is notoriously challenging due to its large size and high negative charge. This often acts as the main barrier to translating nucleic acid technologies into gene therapy. Hence, transfection reagents such as L3000 are often used in basic research. However, different transfection reagent formulations could exhibit low transfection efficiency, reagent toxicity, and batch-to-batch variation. Moreover, such cationic reagents are not cell-specific. For these reasons, they are not compatible *in*

vivo, severely limiting their transferability. Therefore, a means for gene delivery that could overcome the issues associated with toxicity and inefficient transfection was desired.

To expand the utility of LA-plasmids, it was envisaged that self-delivery could be incorporated into LA-plasmids by leveraging the multivalency of WTS. Conceptually, biotinylated delivery agents could bind to WTS in a manner that would leave one or more subunits vacant so that WTS could subsequently bind to LA-plasmids. The final construct, light-activated self-delivered plasmids (LA-SD-plasmids) could spontaneously enter cells once added to the medium. Within cells, the construct should remain dormant until activated by the UV stimulus (Figure 4.16).

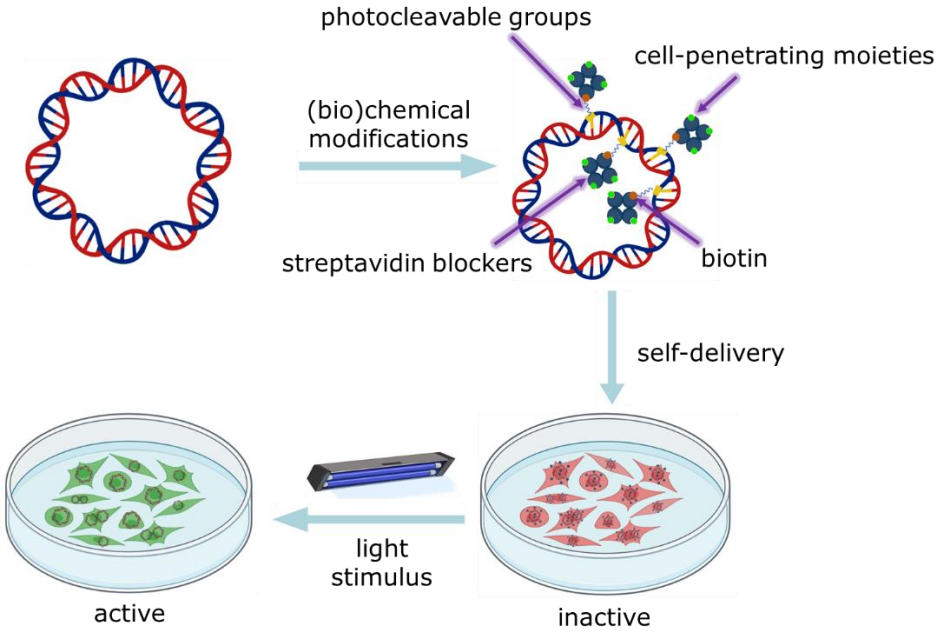


Figure 4.16. Overview of LA-SD plasmid

Streptavidin could bind to cell-penetrating moieties such as biotinylated transferrin and albumin. Upon binding plasmids, this plasmid construct could deliver itself into cells whereby it remained dormant until UV uncaging.

This technology would present a massive advancement within the field of molecular, and developmental biology where plasmid transfection is commonplace. It would also

simplify and reduce the cost of experimental protocols as the plasmid construct could simply be added to cells without transfection complex preparation, which was costly and time-consuming.

Due to the strong, quick, and facile binding between streptavidin and biotin, a range of biotinylated reagents were commercially available. Two such reagents that were amenable for self-delivery purposes included biotinylated transferrin (Bt-Tf) and bovine serum albumin (Bt-BSA, Section 4.1). To visually assess transfection, WTS-Texas Red (TxRed) was chosen to easily detect the construct once it has entered cells.

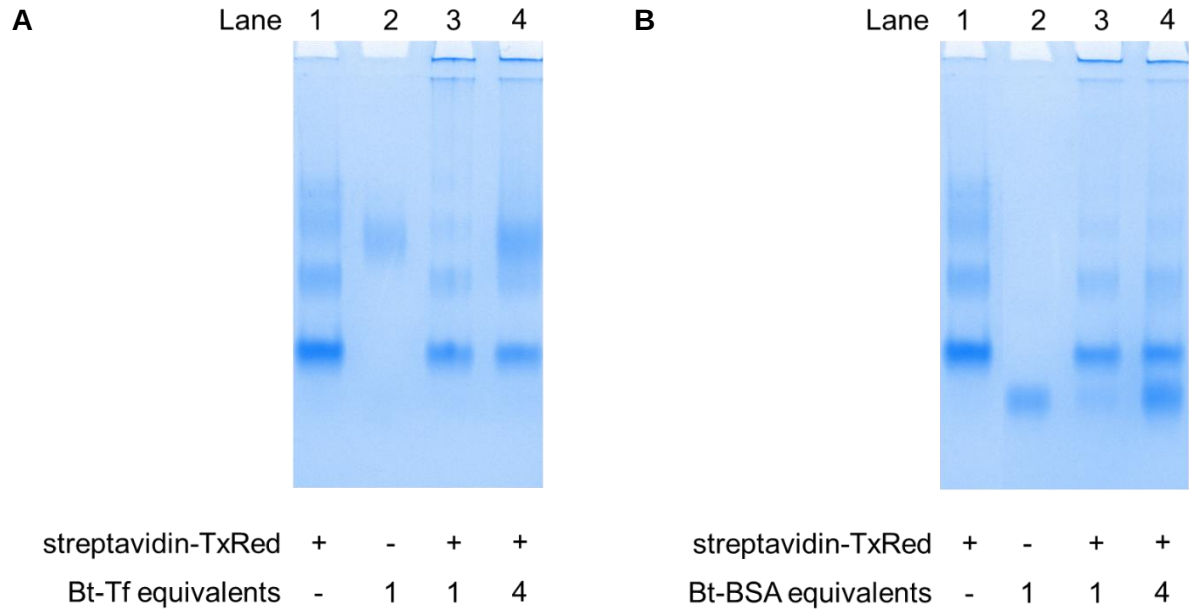


Figure 4.17. Binding Bt-Tf and Bt-BSA to WTS-TxRed

WTS-TxRed was bound to different equivalents of A. Bt-Tf and B. Bt-BSA. Although binding occurred, denoted by the presence of bands at the wells, the reaction was incomplete as WTS-TxRed bands were still present even at molar excess Bt-Tf/BSA.

Although binding was detected at the molar equivalent Bt-Tf/BSA, complete binding was still not observed at 4 equivalents of the biotinylated reagents, as WTS-TxRed bands remained (Figure 4.17). This was unexpected, considering biotinylated

reagents were often labelled with >1 biotin, as the degree of biotinylation could not be controlled, and varied from batch to batch. After verifying with their respective manufacturer, the degree of biotinylation for Bt-Tf and Bt-BSA were 6 Bt-BSA and 12, respectively. This meant that complete binding should have occurred at less than 1 equivalent of Bt-Tf/BSA.

Labelling of WTS with TxRed could affect streptavidin-biotin interaction by altering WTS quaternary structure or occupying important residues involved in biotin binding. To determine if the problem was due to WTS-TxRed, the experiment was repeated with unlabelled WTS. This time, full binding was achieved with 1 equivalent of the biotinylated reagent (Figure 4.18).

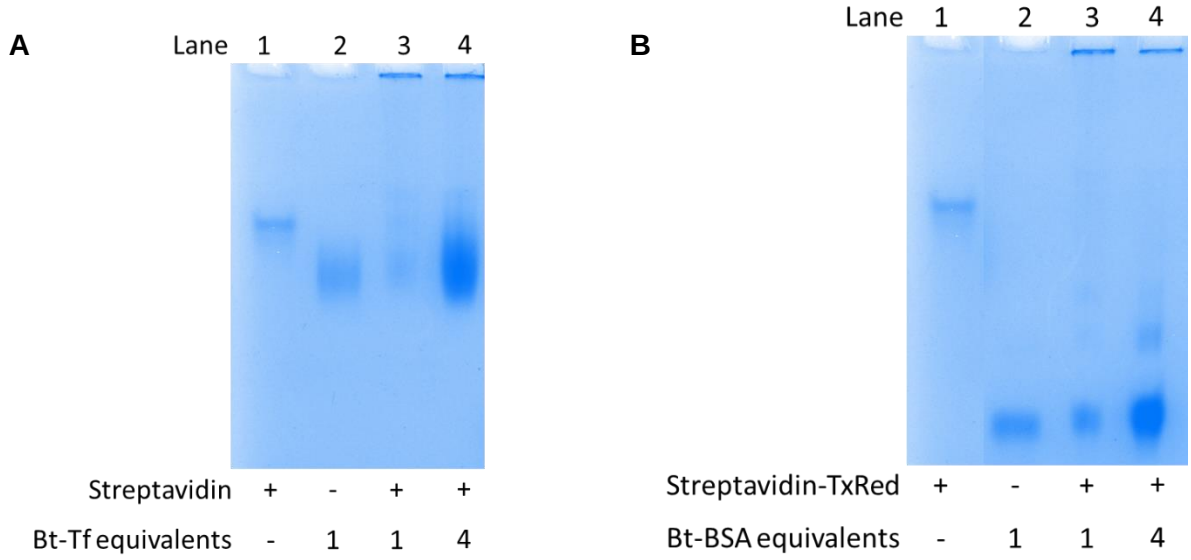


Figure 4.18. Binding Bt-Tf and Bt-BSA to WTS

WTS was bound to different equivalents of A. Bt-Tf and B. Bt-BSA. Binding was complete at 1 equivalent of the biotinylated reagent, denoted by the disappearance of the WTS band.

An alternative method was required to visualise the construct within cells. To this end, Bt-Cy5 (1 biotin per Cy5) could be used to label WTS before binding to Bt-Tf/BSA. To ensure 1:1 binding, Bt-Cy5 was incubated with WTS at high dilution. The reaction was

then lyophilised to dryness, resuspended, and bound to Bt-Tf/BSA. Once again, PAGE showed that complete binding could be affected at the molar equivalence of the biotinylated reagent (Figure 4.19).

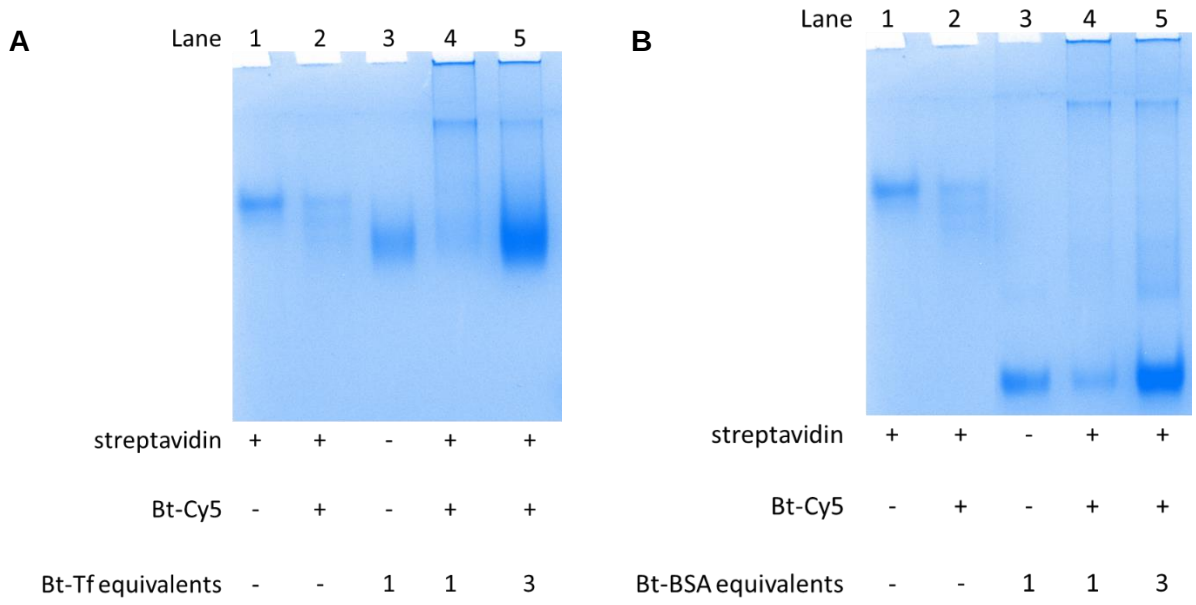


Figure 4.19. Binding Bt-Tf and Bt-BSA to WTS-Cy5

WTS was bound to 1 equivalent of Bt-Cy5 at high dilution to maximise 1:1 labelling. Subsequently, different equivalents of A. Bt-Tf and B. Bt-BSA were bound to WTS-Cy5.

HeLa cells expressed high levels of the TfR and caveolin-1 (CAV1) proteins. These proteins were responsible for the binding and endocytosis of Tf and BSA, respectively (Sections 4.1.2.5 and 4.1.2.6). Moreover, this cell line stably expressed green fluorescent protein (GFP), which enabled the overlay of the cells (GFP channel) and construct (TxRed channel) to determine whether the construct had been internalised.

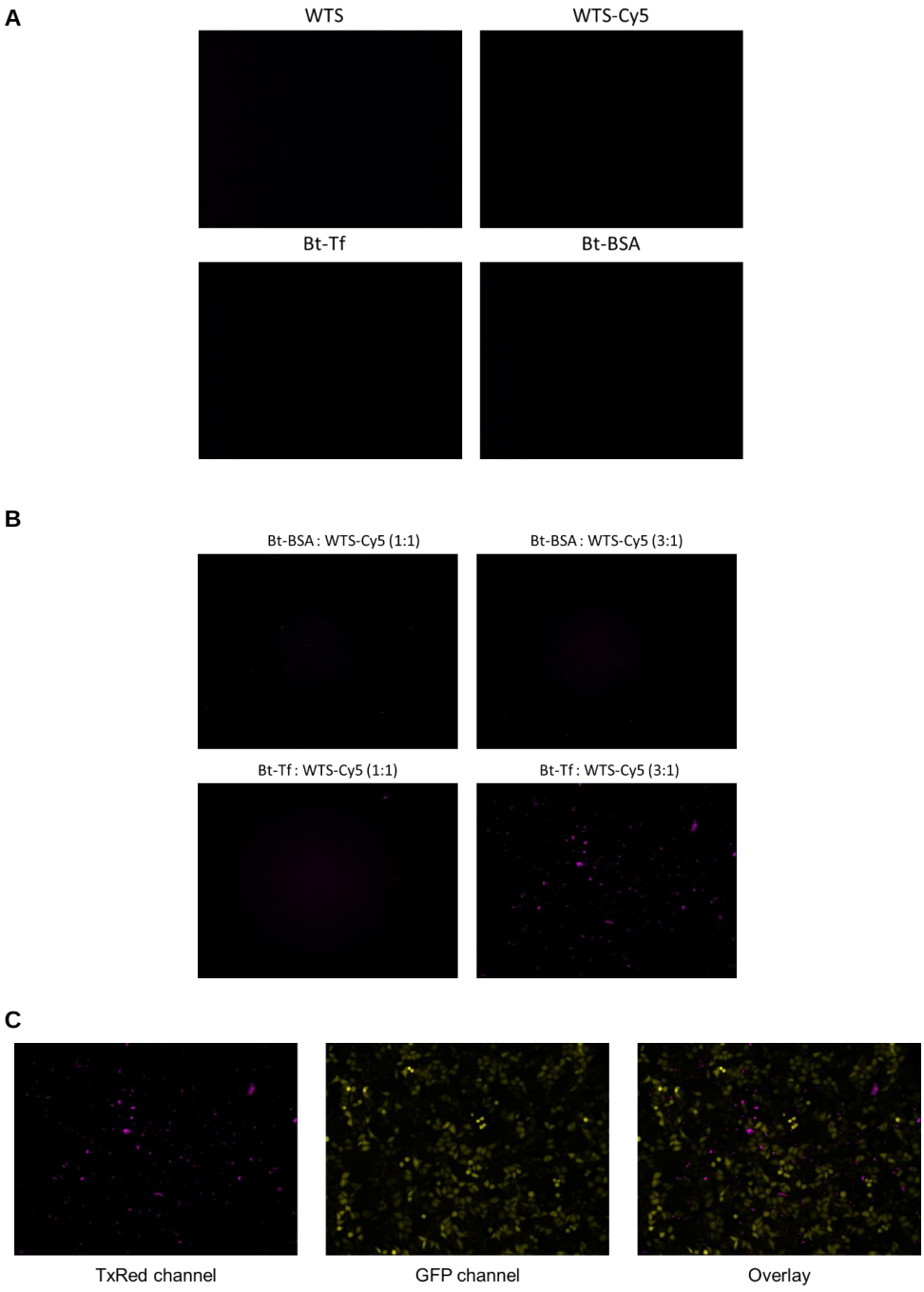


Figure 4.20. Transfection of WTS-Cy5-Tf/BSA constructs

Constructs (1.0 µg based on amount of streptavidin) were added to HeLa-GFP cells and visualised after 7 h. A. None of the controls induced fluorescence by themselves. B. Bt-Tf and WTS-Cy5 constructs could fluoresce at 3 : 1 Bt-Tf eqv. C. Overlay of the TxRed (magenta) and GFP (yellow) channels showed that fluorescence in (B) was outside of cells. Images were representative of technical duplicates (*n* = 2).

The addition of WTS, WTS-Cy5, and Bt-Tf/BSA to HeLa-GFP showed that none of the components by themselves could lead to fluorescence (Figure 4.20A). When the

constructs were assembled as before, only Bt-Tf showed Cy5 fluorescence at 3 equivalents Bt-Tf. However, the overlay of the TxRed and GFP channels showed that the majority of these constructs remained outside cells (Figure 4.20B and 4.20C). WTS crosslinking likely caused aggregation, which could have prevented cell entry.

To minimise aggregation, the Bt-Tf/BSA equivalents were lowered further and directly added to the WTS and Bt-Cy5 reaction at high volume. The ratio of Bt-Cy5 was also increased in case 1 equivalent was not fluorescent enough to be detectable (Appendix Figure A15). As complete binding was achieved with as low as 0.5 equivalent of the Bt-Tf/BSA reagent, constructs formed with >0.5 equivalent of the Bt-Tf/BSA were not assessed for transfection. Once more, these constructs were added directly to cells. Under these conditions, Tf produced the highest fluorescence at 1 equivalent Bt-Cy5 and 0.5 equivalent Bt-Tf. However, the channel overlay still suggested the constructs were outside the cells (Figure 4.21).

This chapter expanded the scope of LA-plasmid by adapting the strategy to cage the CMV promoter. Initially, NEEL could not be used to resynthesise the plasmid. After testing various nickase and plasmid combinations, Nt.BbvCI was definitively shown to possess unspecific nicking activity (Appendix Figure A8B). However, the use of a more specific nickase, Nb.BbvCI, still did not afford the reformed plasmids. By switching to λ Exonuclease, LA-plasmids were successfully generated. To confirm the generality of this exonuclease, it was used to prepare LA-plasmids from Chapter 3 without issue. LA-plasmids with 3 modifications at the TATA box could not regulate expression from the CMV promoter. The availability of other nearby core promoter elements enabled their simultaneous caging. With 9 modifications at the TATA box, Inr, and BREs, protein expression was reduced by 94%. UV uncaging led to a 3.3-fold increase in fluorescence. Finally, self-delivery with biotinylated transferrin and BSA was attempted, but due to the presence of multiple biotins on these reagents, it was not possible to prevent construct aggregation. A strategy to label proteins with a single biotin is

discussed in Chapter 5. The possibility to simultaneously regulate gene expression and facilely deliver plasmids would greatly advance our plasmid caging approach. If successfully implemented, this technology could be adopted by the wider research community for precise control and delivery of plasmids.

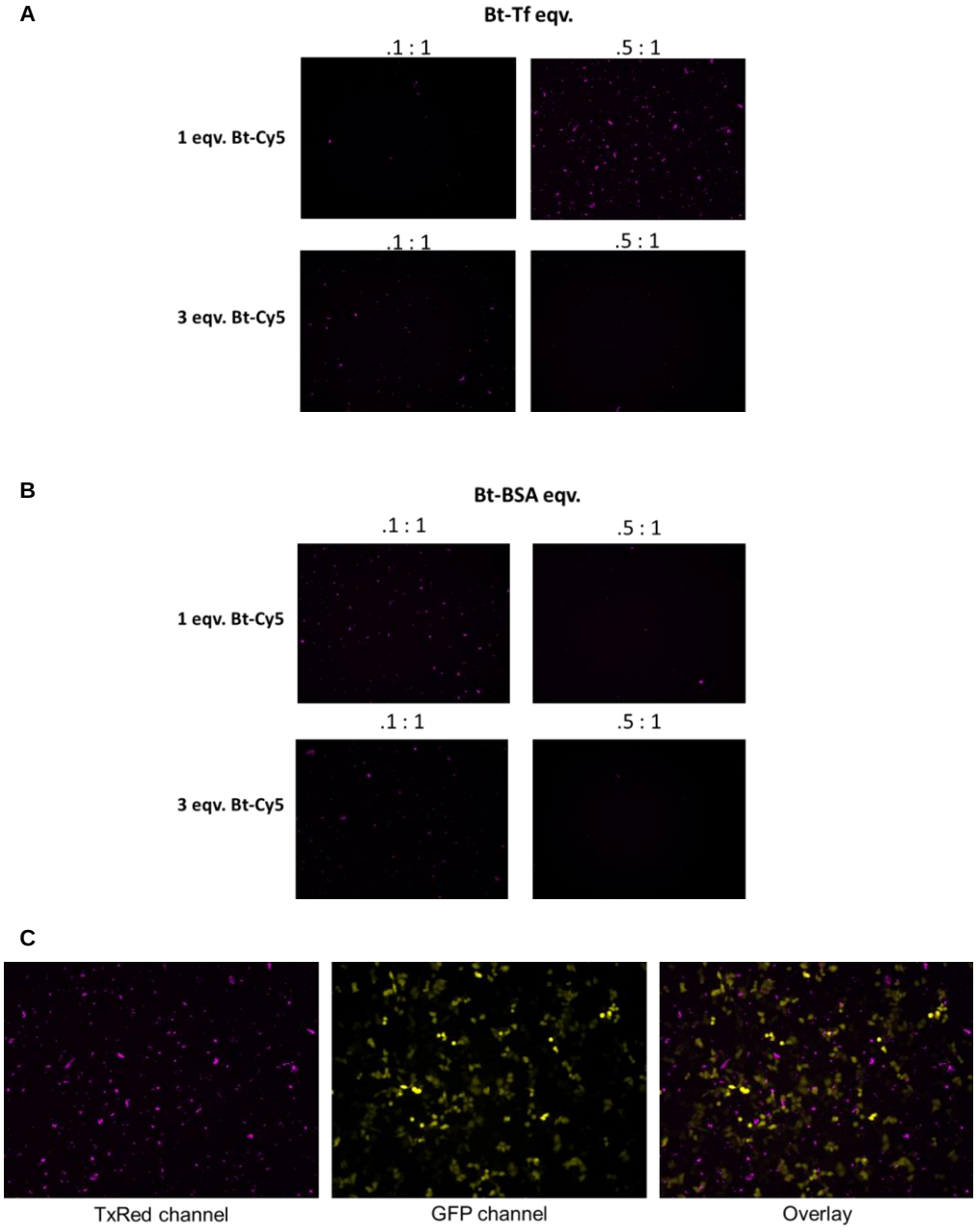


Figure 4.21 Transfection of WTS-Cy5-Tf/BSA constructs using conditions to minimise aggregation

Constructs (1.0 µg, based on amount of streptavidin) from Appendix Figure A15 were added to HeLa-GFP cells and visualised after 4 h. A. Transferrin constructs B. BSA constructs. C. Overlay of the transferrin construct formed from binding 0.5 eqv. Bt-Tf to WTS bound to 1 eqv. Bt-Cy5 (n = 1).

5

Conclusion and Future Directions

Light-controlled nucleic acid expression has garnered much attention in the past 2 decades. Although extensive work has been done to control linear DNA and RNA, plasmids have received comparatively less consideration. This is despite their widespread utility across biology and biochemistry laboratories, and the advantages they could offer over their linear counterparts. Hitherto, plasmid caging has been leaky, and/or required the use of niche phosphoramidites or strict placement of specific DNA sequences which severely limited their utility. The dearth of literature in this regard, in addition to the shortcomings associated with the existing platforms to control plasmid expression, stimulated our efforts to address these outstanding challenges. The work outlined here aimed to engineer a general approach that would enable effective

plasmid caging, with the hope that it might prove useful to the research community to control plasmid expression for applications in cell-free systems and mammalian cells.

The LA-plasmid technology took inspiration from the LA-DNA platform developed by Booth *et. al.*[177] that relied on photocleavable biotin (PCB) and bulky streptavidin to inhibit expression until a UV stimulus was applied. Initially, a method was required to install these modifications onto the plasmid. Chapter 2 delineated the optimisation process relating to NEEL. Breakthrough came with the realisation that the Kunkel method for mutagenesis could be redeployed to resynthesise the plasmid. Specifically, the high flexibility of the T7 DNAP in accommodating both the amine primer and amino dUTPs was crucial in reforming the plasmids. When these amine plasmids were reacted with PCB-NHS ester, the resultant plasmids could bind to streptavidin to give LA-plasmids. To minimise leakiness, any leftover native plasmids needed to be removed. Attempts at isolating the ssDNA were unsuccessful due to low yield or salt contamination. By optimising DpnI conditions, unreacted plasmids were effectively destroyed, whilst leaving NEEL plasmids mostly intact.

Chapter 3 fine-tuned the number of modifications on the plasmid necessary to control expression from a T7 promoter in cell-free conditions. Significantly, it was determined that only the promoter required caging, and having modifications across the plasmid did not benefit the photomodulation efficiency, and even negatively impacted the ON state and toxicity. This simplified our strategy and meant that the exact number and positionings of the modifications were pre-determined. T7 LA-plasmids showed tight control in cell-free expression reactions and photoactivation restored expression to levels comparable to the amine plasmid control. Attempts to demonstrate caging for the tac and Sp6 promoters revealed that the amine modifications alone were enough to abolish expression. For both promoters, the expression could be partially recovered

when PCBs were attached, and this was ascribed to the sequestration of charges on the amines.

To illustrate the scope of this technology, the strategy was expanded to control expression from the ubiquitous CMV promoter in mammalian cells. With 9 modifications, transcription (up to 82%) and translation (up to 88%) were both repressed. Although a higher UV dosage was required, these UV conditions were similar to those in the literature and did not impact cell viability. It's also important to highlight that in the development of NEEL, at least two commercial nucleases were found to act non-specifically: the nickase Nt.BbvCI and Exonuclease III. λ Exonuclease is recommended for NEEL as it has worked with all plasmid sequences attempted. Regarding the nickase, an initial nick-exonuclease test should be performed to determine if the nickase is compatible. Finally, initial attempts at incorporating self-delivery to LA-plasmids by using biotinylated transferrin and albumin were unsuccessful, which were likely due to aggregation.

For LA-plasmids to become a truly useful tool within research, and its translation to gene therapy realised, further development is needed. Several improvements to expand the utility of LA-plasmids are outlined below that, if accomplished, would greatly boost LA-plasmids' attractiveness.

The reliance on nickase's recognition sequence meant that this approach is not truly sequence agnostic. Moreover, the short nickase recognition site meant that the strategy could only work for small-sized plasmids. For example, the possibility that a 7-base recognition site was present in a given plasmid is $1/4^7$ or 1/16384. However, plasmids >30 kb in length (16384 x 2) would, by chance, possess 2 of the same

recognition sequences. Should these be found on both plasmid strands, exonuclease would degrade both strands.

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Cas9 endonuclease relies on a single guide RNA (sgRNA) to direct the activity of Cas9. Its nickase variant, D10A or H840A Cas9, would be suitable for our needs. The programmable nature of the sgRNA and the high target specificity of Cas9 would overcome the nickase promiscuity, and plasmid-size limitation, and make this strategy largely sequence-independent. Although the reliance on the protospacer adjacent motif (PAM) to direct the ribonucleoprotein complex to the nickase site still presented a sequence dependency, in reality, the short length (typically 3-4 nt) of most PAM sequences meant that these sites could be found on virtually any plasmid.

One advantage of streptavidin is the ability to incorporate extra functionalities owing to its multivalency. An outstanding challenge in gene therapy is their delivery into cells. Thus, the ability to simultaneously exert control and deliver plasmids is desirable. Our efforts in this regard showed that aggregation precluded plasmid entering cells. To minimise aggregation, singly-biotinylated transferrin and BSA could be used.[266]

In nature, BirA acts as a biotin ligase and can transfer a single biotin to its partner, the biotin carboxy carrier protein (BCCP). This specific interaction is possible through the recognition of a minimal peptide sequence on BCCP. The natural 13-residue sequence, or the 15-residue variant, AviTag™, could be fused to the N- or C-terminus of transferrin or BSA. The recombinant plasmid could be expressed in heterologous hosts including bacteria, yeast, and mammalian cells. Alternatively, AviTag-labelled Tf/BSA could be biotinylated *in vitro* with BirA.[266] Aggregation could be further reduced by using divalent streptavidin.[267]

The modularity of PCB meant that alternative stimulus-responsive groups could be employed to suit the researcher's needs. Although the UV dosage employed had been minimised and could be tolerated by cells, UV is cytotoxic and could trigger aberrant mutations within the genome that ultimately lead to cancer. Furthermore, UV penetration is limited to a few micrometres and so would be completely blocked by the skin.[170] Thus, the UV-reactive 2-nitrobenzyl could be switched to coumarin-based groups that respond to longer wavelengths.[145] Stimuli other than light, such as temperature,[268] could also be used, which would prove more relevant to effect changes within the body.

Although the T7 LA-plasmids exhibited a minimal OFF state, our current strategy is promoter-dependent, as demonstrated with the tac and Sp6 promoters. Although the CMV promoter proved amenable to exert photocaging, an even tighter control would be required, especially within gene therapy which may not tolerate any leakiness. Furthermore, the LA-plasmid strategy requires initial optimisation to determine the ideal number and placement of photocaging groups for promoters other than the T7 promoter. Additionally, an inherent advantage of using light as the stimulus is its ease to exert spatial control. However, due to the leakiness still observed with the CMV promoter, spatial patterning could not be achieved. These problems could be alleviated by dual-control expression where 2 stimuli would be required to turn on the expression. This has been alluded to with the tac promoter, which by itself could be regulated with an exogenous inducer. However, dual-control could be generalised by having an extra stimulus-responsive functional group at the promoter, or placing multiple stimulus-responsive groups on different promoters. Ideally, these 2 groups' reactivities would not exhibit crosstalk. Within our group, this has already been demonstrated with blue and UV light to create a simple AND gate. Both stimuli were

required to reconstitute expression and enabled spatial activation of gene expression.[269] Taken together, these alterations would greatly expedite LA-plasmids' adoption in research to benefit studies where controlled and precise plasmid expression is required.

6

Methods

6.1 Materials

All enzymes were purchased from NEB unless stated otherwise. Chemicals and solvents were purchased from Merck/Sigma unless stated otherwise. QuickBlue Protein Stain was purchased from Stratech Scientific Ltd. ATP, SuperScript IV VILO Master Mix with ezDNase, Fast SYBR Green Master Mix, Lipofectamine 3000, Fluorobrite DMEM, Opti-MEM reduced serum medium, Genejet PCR Purification Kit, Nt.- and Nb.Bpu10I, DreamTaq PCR MM, and biotinylated BSA were purchased from Thermo Fischer Scientific. 5'-aa-dUTP and 5'-ap-dUTP were purchased from Jena Bioscience. Regular primers were purchased from Integrated DNA Technologies. Amine primers were purchased from ATD-Bio or Merck/Sigma. PCB-NHS ester was

purchased from Click Chemistry Tools or Sigma/Merck and stored in dry dimethylformamide (DMF) at -80 ºC. Trypsin-EDTA, FBS, Amicon 3K columns, Streptavidin-TxRed, Streptavidin-Atto647, biotinylated transferrin, 3X GelRed, and TAE Buffer were purchased from Merck/Sigma. Proteus Mini Clarification Spin Column (GEN-MSF500) was purchased from Protein Ark. TEMED (161-0801) was purchased from Bio-Rad. Gel extraction, maxiprep, miniprep, and RNeasy Mini kits were purchased from Qiagen. CellTiter-Glo Luminescent Cell Viability Assay was purchased from Promega. Pico488 dsDNA Quantification Reagent was purchased from Lumiprobe. Biotin-Cy5 was purchased from Click Chemistry Tools. HEK293T cell line was a gift from Dr. Haiping Tang (Prof. Dame Carol Robinson's group). The HeLa-GFP cell line was purchased from Cell Biolabs. Seakem® LE Agarose was purchased from Lonza. mV-N1 was a gift from M. Davidson. mC and mV plasmids (containing the T7 promoter) were available and previously prepared by Booth *et. al*.[177] CMV_mNG plasmid was purchased from Addgene (plasmid #128144). ChR2-EYFP plasmid was purchased from Addgene (plasmid #20939).

6.2 General Methods

The following methods were used throughout each chapter and will be referred to in the chapters as appropriate.

6.2.1. Homologous Recombination

100 ng vector was added to a 1.5 mL Eppendorf tube. A 3:1 or 5:1 vector:insert ratio was used and calculated using the NEB Ligation calculator (https://nebiocalculator.neb.com/#!/ligation). The following was performed by flame. XL-10 Gold was homogenised by gently blowing bubbles at the bottom of the tube. 50 uL bacteria were added to the tube containing the vector and insert. The tube was

incubated on ice for 30 mins, heat-shocked at 42 ºC for 30 s, and then again on ice for 2 mins. SOC media (500 µL) was added and the tube was incubated at 37 ºC with shaking (225 rpm) for 1 h. The solution (200 µL) was plated on LB agar plates containing ampicillin (100 µg/mL) using glass beads at 37 ºC for 14-16 h.

Colonies were picked using pipette tips and inoculated in 50-mL Falcon tubes with LB (5 mL) containing ampicillin (100 µg/mL) and incubated at 37 ºC with shaking (225 rpm) for 14-16 h. Glycerol stocks were prepared from the overnight cultures (300 uL) and glycerol (300 uL, 50% v/v, filtered) and stored at -80 ºC. The remainder of the overnight cultures were miniprepped according to the manufacturer's protocols and eluted in 50 µL Elution Buffer. Concentrations were measured using Nanodrop ND-1000 Spectrophotometer. Sequencing of the insert was performed by Source BioScience. Whole plasmid sequencing was performed by Plasmidsaurus.

6.2.2 PCR

Primers were resuspended to give 100 mM in Tris (10 mM, pH 8.0), these were further diluted to give 10 mM working solutions. PCR reactions were performed in 0.2 mL PCR tubes and contained the plasmid template (1 ng), the FOR and REV primers (1 µL, 10 mM), Phusion MM with HF Buffer or DreamTaq PCR MM (10 µL, 2X) and made up to 20 µL with MilliQ water. Reactions were cycled on a Peqlab peqSTAR 96X Universal thermal cycler with the following conditions: initial denaturation at 98 ºC for 30 s, 30 cycles of: [denaturation at 98 ºC for 10 s, annealing at varying temperatures, see the relevant tables below, extension at 72 ºC for 30 s per kb], a final extension at 72 ºC for 10 mins, and held at 4 ºC. DpnI (0.5 µL, 20 U/µL) was added directly to PCR reactions and incubated at 37 ºC for 30 mins and deactivated at 80 ºC for 20 mins. Reactions were analysed on an agarose gel. If >1 band formed, the desired band was

gel-extracted using QIAquick Gel Extraction Kit according to the manufacturer's protocols.

6.2.3 Agarose Gel Electrophoresis

100-150 ng DNA was mixed with 6X Purple Loading Dye and made up to 12 uL with MilliQ water. The reactions were loaded on a 1.0% agarose gel pre-stained with 1X GelRed. Gels were run in 1X TAE Buffer at 130 V for 55 mins. Imaging was done using a BioRad Molecular Imager® Geldoc™ XR+ gel imager and analysed using ImageLab 6.1.

6.2.4 Ethanol Precipitation

Sodium acetate (10% v/v, 3 M, pH 5.2) and ethanol (3 volumes, 99.8%) were added to nucleic acids. The tubes were vortexed and kept at -80 ºC for 1-2 h or -20 ºC overnight. Tubes were then centrifuged at 15,000 rpm at 4 ºC for 30 mins. The supernatant was carefully removed and ethanol (1 mL, 70%) was added, gently mixed by pipetting, and the tubes were centrifuged again at the above settings for 10 mins. The supernatant was removed and the tubes were dried under vacuum using a speed vac for >10 mins. Dry DNA pellets were re-suspended in MilliQ water or Tris (10 mM, pH 8.0).

6.2.5. Nicking Reactions

Bulk nicking reactions were performed in 1.5 mL Eppendorf tubes and consisted of the plasmid (75 µg), (A) Nt.BspQI – for the T7 mV plasmid or (B) Nt.BbvCI – for the CMV mNG/tac mNG plasmid (62.6 µL, 10 U/µL), (A) NEBuffer 3.1 or (B) CutSmart Buffer (62.6 µL, 10X), and made up to 625 µL with MilliQ water. The reactions were incubated on the 1.5 mL Thermomixer C (Eppendorf) at (A) 50 ºC or (B) 37 ºC for 2 h, and both

were heat-deactivated at 80 ºC for 25 mins, and stored at 4 ºC. The reactions were purified using Genejet PCR columns according to the manufacturer's instructions and eluted in 50 µL Elution Buffer. The concentration was measured using the NanoDrop.

6.2.6. Lambda Exonuclease Reactions

The total reaction volumes varied and depended on the amount of nicked plasmid used. Bulk exonuclease reactions were performed in 1.5 mL Eppendorf tubes consisting of the nicked plasmid (final concentration = 120.0 ng/µL), λ Exonuclease (final concentration = 1.0 U/µL, 5 U/µL), λ Exonuclease Reaction Buffer (final concentration = 1X, 10X), and made up to the desired volume with MilliQ water. The reactions were incubated on the Thermomixer C at 37 ºC for 3 h, heat-deactivated at 75 ºC for 20 mins, and stored at 4 ºC. The reactions were purified using Genejet PCR columns according to the manufacturer's instructions and eluted in 50 µL Elution Buffer. The concentration was measured using the NanoDrop.

6.2.7. Phosphorylation Reactions

The reactions were performed in 0.2 mL PCR tubes and consisted of the regular, amine, or PCB primer (60 pmol), T4 PNK (2 µL, 10 U/µL) in T4 PNK Buffer (4.0 µL, 10X), ATP (0.4 µL, 100 mM) and made up to 40 µL with MilliQ water. The reactions were incubated on the thermal cycler at 37 ºC for 1 h, heat-deactivated at 80 ºC for 20 mins, and stored at 4 ºC.

6.2.8. Annealing Reactions

The total reaction volumes varied and depended on the amount of ssDNA used. The molar mass of the ssDNA could be calculated using the equation:

$$\mathrm{Mr}\,(\mathrm{ssDNA}) = \mathrm{length}\,(\mathrm{ssDNA\ in\ bases}) \times 330\ \mathrm{g/base}$$

And the molar amount of ssDNA was given by:

$$\mathrm{n\ (ssDNA)} = \frac{\mathrm{m(ssDNA)}}{\mathrm{Mr\ (ssDNA)}}$$

The annealing reactions were performed in 1.5 mL Eppendorf tubes and consisted of: ssDNA template (final concentration = 12.84 ng/uL), the phosphorylated regular, amine, or PCB primer (4 molar equivalents relative to the template), magnesium chloride (final concentration = 10 mM, 500 mM), Tris (pH 7.4, final concentration = 50 mM, 1 M), and made up to desired volume with MilliQ water. The lids were parafilmed to minimise volume loss. The reactions were incubated on the thermomixer at 90 ºC for 2 mins, then cooled to 25 ºC at a cooling rate of -1 ºC/min.

6.2.9. Extension and Ligation Reactions with PCB Oligos

Note that T4 DNA Ligase's (NEB) concentration was given in Cohesive End Units/mL (CEU/mL). The following procedure used Weiss Units. 1 Weiss Unit = 67 CEU.

The total reaction volumes varied and depended on the amount of ssDNA used. The extension and ligation reactions were performed in 1.5 mL Eppendorf tubes and consisted of: the annealed mix (final concentration = 11.44 ng/µL), each of the regular dNTP (final concentration = 0.8 mM, 100 mM), DTT (final concentration = 5 mM, 1 M), T7 DNAP (final concentration = 0.05 U/µL, 10 U/µL), T4 DNA Ligase (0.05 Weiss U/µL, 5.97 Weiss U/µL), and made up to the desired volumes with MilliQ water. The reactions were incubated on the thermomixer at 20 ºC for 3 h and kept at 4 ºC.

Extension and ligation reactions in the presence of amino dUTPs: reactions were performed similarly to the above, except ratios of dTTP to 5'-aa-dUTP or 5'-ap-dUTP were used according to the percentages indicated in the main text.

6.2.10. DpnI Reactions

Note that after extension, the DNA concentration doubled. To minimise degradation of the reformed hemi-methylated plasmids, the DNA and enzyme concentrations were kept low, with a short reaction time. The reactions were performed in 1.5 mL Eppendorf tubes and consisted of: ligated plasmid (final concentration = 6.7 ng/µL), magnesium chloride (final concentration = 10 mM, 500 mM), Tris (final concentration = 50 mM, 1 M, pH 7.4), DpnI (final concentration = 0.06 U/µL, 20 U/µL), and made up to the desired volume with MilliQ water. The reactions were incubated on the thermomixer at 37 ºC for 5 mins. EDTA (final concentration = 20 mM, 500 mM, pH 8.0) was added to stop the reaction. The reactions were lyophilised on the Christ Alpha 2-4 LDplus to dryness.

The reactions were resuspended in MilliQ water (volumes depending on reaction scale) and purified using PCR columns. The reactions were eluted in Elution Buffer (50 µL, twice). Ethanol precipitation was performed as described in Section 6.2.4. The reactions were resuspended in Tris (10 mM, pH 8.0, volumes dependent on plasmid amounts), and 260/230 ratios were determined by NanoDrop, whilst concentrations were determined by Pico488.

6.2.11. Measuring Concentrations with Pico488

The concentrations of the native, amine, and PCB plasmids were measured with Pico488 according to the manufacturer's instructions before cell-free reactions and transfection. Reactions were prepared in 1.5 mL Eppendorf tubes. Briefly, standards with known concentration were prepared from a 100 ng/µL dsDNA stock solution and Tris-EDTA buffer (10 mM, pH 7.4, 1 mM EDTA), to give 100 µL with a concentration range from 0 – 2000 pg/µL. The samples contained the plasmids (1 µL) and were made up to 100 µL with the same Tris buffer. A 1X working solution of the Pico488

dye was prepared from the 200X stock by diluting in Tris buffer immediately before use. 100 µL of the working dye solution was added to each of the standard and sample. The solutions were vortexed and incubated under tin foil at room temperature for 5 mins. The solutions were transferred to a clear lid 96-well flat bottom with a low evaporation lid and centrifuged at 1,250 rpm for 1 min to allow solutions to settle and remove any bubbles. The fluorescence was measured on a Tecan Infinite M1000 Pro fluorescence plate reader. Plates were shaken at 654 rpm for 10 s and fluorescence was measured with settings: λ_{ex} = 503 nm, λ_{em} = 525 nm, bandwidth = 5 nm.

The calibration curve was plotted with the standard concentrations (x-axis) and fluorescence (y-axis) using Excel. The line of best fit was obtained and the equation was used to calculate the dsDNA concentrations.

6.2.12. Binding Streptavidin

The total reaction volumes varied and depended on the amount of dsDNA formed. The molar amounts of dsDNA were determined according to Section 6.2.8, except the molar mass for a single base pair was approximated to be 660 g. This number could then be used to calculate the molar amounts of PCB:

$$\mathrm{n\ (PCB)} = \mathrm{n\ (plasmid)} \times \mathrm{n\ (PCB\ per\ plasmid)}$$

A mass value of 52.8 kDa for streptavidin was used throughout this thesis to calculate the molar equivalents of streptavidin required. Reactions were performed in 1.5 mL Protein LoBind Tubes and consisted of: PCB plasmids (final concentration = 50.0 ng/µL), either MVS or WTS (at 3 molar equivalents relative to PCB sites), Tris (final concentration = 20 mM, 1 M, pH 7.4), and made up to the desired volume with MilliQ water. The reactions were incubated on the thermomixer at 26 ºC overnight.

6.2.13. PCB-NHS Ester Reaction with Amine Oligo

Before the reactions, 1 M sodium bicarbonate was prepared fresh and used within 1 week. The reactions consisted of the amine oligos (20 µL, 100 µM), sodium bicarbonate (20 µL, 1 M), PCB-NHS ester (100 µL, 10 mM), and made up to 200 µL with MilliQ water. The reaction was incubated at 20 ºC on the 1.5 mL thermomixer with gentle shaking (600 rpm) overnight.

The reactions were stopped with Tris (1 volume, 10 mM, pH 8.0). The reactions were purified with an Amicon 3 K 0.5 mL column by centrifuging at 14,000 g for 15 mins at room temperature. The eluate was decanted, and the concentrate was made up to 500 µL with MilliQ water. This process was repeated twice more (3 centrifugations in total). The columns were then inverted and inserted into the collection tubes. These tubes were centrifuged at 1,000 g for 2 mins. To maximise the recovery, another elution step was performed with MilliQ water (220 µL). The final volume was ~300 µL.

6.2.14. HPLC Purification of PCB Oligo

Lines B/C/D were: MilliQ water, acetonitrile, and triethylammonium bicarbonate buffer (TEAB, 100 mM, pH 8.5). Line D was kept constant at 10%. The flow rate was 1.5 mL/min. Solvent percentages refer to line C (organic phase). Separations were UV-monitored at 260, 280, 310, and 360 nm.

The PCB oligo (100 µL, 6.67 µM) was injected into an Agilent Polaris 5 C18-A (150 x 4.6 mm) analytical column heated to 50 ˚C and pre-equilibrated at 5% for ~20 mins. A 5-30% gradient over 27 mins was used to separate the oligos. The organic phase was increased to 90% at 27.01 mins and maintained until 30 mins to wash the column. The organic phase was returned to 5% at 30.01 mins and kept until 33 mins to re-

equilibrate the column. Three injections were made and the fractions were collected, pooled (Tables 6.7, 6.13, and 6.20), and lyophilised to dryness.

Reactions were resuspended in MilliQ water (1 mL) and lyophilised to dryness. Reactions were again resuspended in MilliQ water (500 µL) and purified using Amicon 3 K 0.5 mL columns similar to before (Section 6.2.13), except 420 µL was added in the second elution step. The final volume was made up to 1 mL with MilliQ water. The reactions were again lyophilised to dryness. The reactions were finally resuspended in Tris (10 mM, pH 8.0), and concentrations were measured by NanoDrop. The 260/230 was typically >1.7.

6.2.15. Denaturing PAGE

The 5 mL, 16% polyacrylamide gel solution was assembled in a 15-mL Falcon tube and consisted of Tris-Borate-EDTA (TBE, 0.5 mL, 10X), polyacrylamide (2.0 mL, 40%), GelRed (1.0 mL, 3X), urea (2.10 g). The solution was vortexed vigorously and sonicated and degassed for 10 mins. The tube was vortexed again to ensure all urea was dissolved. Ammonium persulphate (APS, 40 µL, 10% w/v) and tetramethylethylenediamine (TEMED, 2.5 µL) were added. The solution was inverted several times to mix and quickly loaded into the assembled glass plate. The solution was allowed to set for >0.5 h.

Amine and PCB oligos samples consisted of the nucleic acids (200 ng), and a homemade loading dye (5 µL, 2X), the solutions were made up to 10 µL with MilliQ water. The 1 mL 2X loading dye consisted of formamide (95% v/v), sodium dodecyl sulphate (0.02% w/v), xylene-cyanol (0.01% w/v), bromophenol blue (0.02% w/v), EDTA (1mM, 1 M, pH 8.0). The solutions were loaded onto the gel and run in 1X TBE Buffer at 200 V for 1 h. The gel was soaked in GelRed for ~5 mins and imaged.

6.2.16. LC-MS

Line A was: 8.6 mM Et_3N, 200 mM HFIP in 5% MeOH/H_2O (v/v). Line B was: 20% buffer A in MeOH. Solvent percentages refer to line B.

Samples (10 µL, 1 µM) were prepared and injected onto a Waters Acquity UPLC BEH C18 column (130 Å, 1.7 µm, 2.1 mm x 50 mm). Samples were analysed over 8 mins with a 0-70% gradient. Masses were recorded on a Waters Xevo G2 QTOF ESI-UPLC-MS system. Analysis was carried out using MassLynx v4.1.

6.2.17. Maxiprep

After plasmids had been sequenced with Sanger sequencing or whole plasmid sequencing, maxiprep was performed to obtain microgram-scale DNA. LB (100 mL) containing ampicillin (100 µg/mL) was inoculated from the desired glycerol stock and incubated at 37 ºC with gentle shaking (225 rpm) overnight. Optical density at 600 nm (OD_{600}) was measured and the cultures were harvested when OD_{600} had reached 1.0 x 10^9 cells/mL. Maxiprep was performed according to the manufacturer's protocol. Before use, RNase A and LyseBlue were added to Buffer P1. Briefly, cultures were transferred to 50-mL Falcon tubes and centrifuged at 6,000 g for 15 mins at 4 ºC. The supernatant was decanted. Buffer P1 (10 mL) was added and the pellet was pipetted to thoroughly lyse the cells. The solutions were transferred to Nalgene™ round-bottom polypropylene centrifugation tubes (Thermo Scientific). Buffer P2 (10 mL) was added and the solutions were mixed by inverting several times. The solutions were incubated at r.t. for 5 mins. Chilled Buffer P3 (10 mL) was added and the solutions were mixed by inverting several times. The solutions were precipitated on ice for 20 mins. The solutions were centrifuged at 20,000 g for 30 mins at 4 ºC. The supernatant was transferred to new polypropylene tubes and re-centrifuged at 20,000 g for 15 mins at

4 ºC. Qiagen-500 tip columns were equilibrated with Buffer QBT (10 mL). The supernatant was applied to the equilibrated column, whilst avoiding particulate carry-over. The columns were washed twice with Buffer QC (30 mL each time). DNA was eluted with Buffer QF (15 mL) into clean polypropylene tubes. DNA was precipitated by adding isopropanol (10.5 mL), mixed by inverting several times, and centrifuged at 15,000 g for 30 mins at 4 ºC. The supernatant was decanted carefully to avoid disturbing the pellets. The pellets were washed with 70% ethanol (5 mL) and centrifuged at 15,000 g for 10 mins. The supernatant was carefully decanted to avoid disturbing the pellets. The remaining ethanol was dried with a gentle stream of nitrogen. The pellets were resuspended in Tris (2 x 500 µL, 10 mM, pH 8.0). Concentrations were measured on the NanoDrop.

6.2.18. UV Irradiation

Cell-free reactions were irradiated immediately before incubations on the thermal cycler. Transfected wells were irradiated 4 h after transfection. The lamp (365 nm, 700 mA, M365-L2, Thorlabs), equipped with a collimation adapter (COP5-A, Thorlabs) was collimated at a distance of ~22.5 cm (from the base of the lamp to the bottom of the tube/plate). The UV intensity was measured with a photodiode power sensor (200-1100 nm, 50 mW, Ø = 9.5 mm, S120VC, Thorlabs) equipped with a power meter console (PM100A, Thorlabs). The sensor measured UV power in mW. These values were then converted into irradiance units by dividing the power by the area of the aperture in cm^2. PCR tubes were placed on a PCR rack, at the centre of the irradiation zone, and 4 tubes were irradiated at a time. Likewise, the irradiation zone also covered 4 wells so that technical triplicates were irradiated simultaneously. To uncage the T7 promoter, an irradiance of 2.47 mW cm^{-2} was applied for 10 mins. To uncage the CMV promoter, an irradiance of 9.88 mW cm^{-2} was applied for 10 mins. To ensure

consistency, reactions, and wells not requiring UV were also kept at r.t. away from UV, whilst UV was being applied to '+UV' conditions. The plates were re-incubated at 37 ºC, 5% CO_2 for 2 days.

6.3. Chapter 2 Methods

6.3.1. Gel extraction of ssDNA

Gel extraction was performed according to the manufacturer's protocol with minor emendations to increase yield. ssDNA was run on an agarose gel (750 ng per lane, ~15 µg in total, Section 6.2.3). The ssDNA bands were excised using a G-Pick™. Gel slabs were combined into 4 tubes and weighed. Buffer QG (1.5 volumes) was added to each tube, and the gels were incubated on the thermomixer to dissolve at 50 ºC with gentle shaking (600 rpm) for 15 mins, with vortexing every 5 mins. The samples were applied to spin columns and centrifuged for 1 min at 2,000 g (the same settings were used throughout unless stated otherwise). The flow-through was re-applied and centrifuged to maximise recovery. Buffer QG (500 µL) was applied to the columns and centrifuged. Buffer PE (2 x 750 µL) was applied and left to incubate on membranes for 2 mins at r.t., followed by centrifugation. The columns were centrifuged at max speed for 1 min to remove residual Buffer PE. Columns were transferred to 1.5 mL Eppendorf tubes. Buffer EB was pre-warmed at 50 ºC and 30 µL was applied to each column. The columns were incubated at 50 ºC for 5 min, then centrifuged at max speed for 1 min. This was repeated twice more to maximise recovery. The concentrations were measured on the NanoDrop, but yield and purity were low.

6.3.2 Electroelution of ssDNA Directly into Wells

ssDNA was run on an agarose gel (750 ng per lane, ~4.5 µg in total, Section 6.2.3) and the bands were visualised on a UV Transilluminator (Benchtop 3 UV™, UVP). The gel immediately in front of the ssDNA bands was removed using a G-Pick to create a continuous well. The running buffer was removed so that there was just enough buffer for electrophoresis, whilst leaving the gel unsubmerged to avoid ssDNA flowing out of the well. Tris (10 mM, pH 8.0) was added to the well. The bands were allowed to run momentarily and the gel was visualised frequently to monitor band migration. If bands had surpassed the well, the gel was run in the reverse direction. This process was repeated until the bands had run into the well. Tris containing the ssDNA was recovered and lyophilised to dryness. The ssDNA was resuspended in MilliQ water (400 µL) and ethanol-precipitated (Section 6.2.4). However, the recovered material suffered from low yield and salt contamination.

6.3.3. G-capsule to Purify ssDNA

Before use, the G-Capsule™ was washed with MilliQ water. ssDNA was run on an agarose gel (500 ng per lane, Section 6.2.3). The bands were excised using a G-Pick. The G-Trap™ was filled with MilliQ water, and the G-Pick and G-Trap were assembled. The gel piece was pushed with a plunger to position it close to the membrane. The G-Capsule was placed in the gel tank and the ssDNA was electroeluted into the G-Trap. The current was reversed momentarily. Buffer was removed from G-Pick and a puncture was made in the G-Trap's collection port and the buffer containing ssDNA was removed. The solution was ethanol-precipitated and the concentration was determined using NanoDrop. Once again, the yield and purity were

low using this method. Furthermore, this method was time-consuming as one band had to be removed at a time.

6.3.4. Dialysis Membrane to Purify ssDNA

ssDNA was run on an agarose gel (750 ng per lane, 4.5 µg total, Section 6.2.3). The bands were excised using a scalpel. A cellulose dialysis membrane was clipped on one side, and filled with MilliQ water. The gel slab containing the ssDNA was added to the tubing and clipped on the other side. The bands were electroeluted into the surrounding solution. The clip was carefully removed on one side, and the solution was recovered. The ssDNA was then ethanol-precipitated but the yield and purity were low. This method suffered from inconsistency, as the bands sometimes did not elute out of the gels.

6.3.5. Crush and Soak to Purify ssDNA

ssDNA was run on an agarose gel (750 ng per lane, Section 6.2.3), except the gel and running buffer were prepared with ammonium acetate instead of TAE Buffer. The bands were excised using a G-Pick. The gel slabs were transferred to a 1.5 mL Eppendorf tube and mashed with a pestle. Ammonium acetate (200 µL, 300 mM) was added and the tube was incubated on the thermomixer at 37 ºC with gentle shaking (600 rpm) overnight. The solution was transferred to a Proteus Mini Clarification Spin Column and centrifuged at 14,000 g for 2 mins to remove residual agarose. This step was repeated with either another Proteus Mini Clarification Spin Column, or the solution was purified with a PCR column. The recovered solution was lyophilised to remove ammonium acetate. The materials were resuspended in 50 µL MilliQ water the concentrations were measured using NanoDrop. The purity and yield were poor.

Significantly, for all ssDNA purification methods from an agarose gel attempted, the ssDNA recovered was not able to give the NEEL plasmids.

6.3.6. PCB-NHS ester Reaction with Amine Plasmids

Reactions were performed in 1.5 mL Eppendorf tubes and consisted of amine plasmids (5 µg), sodium bicarbonate (5 µL, 1 M), PCB-NHS ester (5 µL, 100 mM), and made up to 50 µL with MilliQ water. The reactions were incubated on the thermomixer at 20 ºC. After 2 h, PCB-NHS ester (2.5 µL, 100 mM) was added, and the reactions were vortexed and left on the thermomixer overnight.

6.3.7. Gyrase Reactions

The reactions were carried out in PCR tubes and consisted of plasmid (100 ng), Gyrase Reaction Buffer (4.0 µL, 5X), and Gyrase (0.5 µL), and made up to 20 µL with MilliQ water. The reactions were incubated on a thermal cycler at 37 ºC for 1 h and heat-deactivated at 70 ºC for 30 mins.

6.3.8. Summary of Primers, PCR, and Homologous Recombination Conditions, and HPLC Details in Chapter 2

Name	Sequence
C_C_FOR	CTGAAGCTCATCTGCACC
G_T_FOR	TCCAGGAGCGCACCATCTTC
T_C_FOR	CAAGGACGACGGCAACTAC
G_C_FOR	CATCGAGCTGAAGGGCATC
C_T_FOR	TTCAAGGAGGACGGCAACATCC
FOR_plasAmp/G_G_FOR	GGAATTCTCGAGTAAGGTTAACC
T_A_FOR	ATATCCGGAAGCTTGGCACTGG
G_A_FOR	AGCGGTATCAGCTCACTC
C_A_FOR	AAGTCAGAGGTGGCGAAACCC
A_A_FOR	AGCGTGGCGCTTTCTCATAGC
T_G_phos	GTAGGTCGTTCGCTCCAAGC
A_G_FOR	GTAAACTTGGTCTGACAGTTACC
A_T_FOR	TCTCAGCGATCTGTCTATTTCG
T_T_FOR	TCATCCATAGTTGCCTGAC
C_G_phos	GAGTTGCTCTTGCCCGGCGTC
A_C_phos	CACGGAAATGTTGAATACTC

Table 6.1. Primers used to assess dinucleotide identities at the ligation junction

The dinucleotide identities could be identified in the primer name and followed the format: X_Y_FOR or X_Y_phos where X was the nucleotide at the 3'-end and Y was the nucleotide at the 5'-end. FOR_plasAmp/G_G_FOR was used for NEEL until the plasmid was caged at the T7 promoter.

Name	Sequence
mV nick T7T FRW	TACAAGTAATGACCTCAGCGGATCCGCTCTTCCGGGAATTCTCGAGTAAGG
DHFR REV T7S GFP	ACAGCTCCTCGCCCTTGCTCACCATATGTATACCTCCTTCTTAAAGTTAAAC
GFP FRW T7S	CGGCATGGACGAGCTGTACAAGTAATGAGGATCCCGGGAATTCTCGAGTA
mV nick T7T REV	AATTCCCGGAAGAGCGGATCCGCTGAGGTCATTACTTGTACAGCTCG

Table 6.2. Chapter 2 PCR Primer Sequences

PCR	DNA template	For primer	Rev Primer
mV_nicks_BB	mVenus_CT	mV nick T7T FRW	DHFR REV T7S GFP
mV_nicks_insert	mVenus_CT	GFP FRW T7S	mV nick T7T REV

Table 6.3. Chapter 2 PCR reaction compositions

PCR	Tm/ ºC	Extension time/ mins
mV_nicks_BB	55	1.5
mV_nicks_insert	55	1.5

Table 6.4. Chapter 2 PCR reaction conditions

Plasmid	Backbone	Insert
mV_nicks	mV_nicks_BB	mV_nicks_insert

Table 6.5. Chapter 2 homologous recombination reactions

Name	Sequence
T7_7amines	GAAATXAAXACGACXCACXAXAGGGXCXAG

Table 6.6. Primers used to cage the T7 promoter

Underlined sequence represented the T7 recognition sequence. X represented amino-C6-dT modifications.

Promoter	Number of modifications	Collection period/ mins	Increments/ mins	Fractions pooled
T7	7	24.0-25.5	0.15	5-7

Table 6.7. Purification of T7 PCB oligo

6.4. Chapter 3 Methods

6.4.1. IVT with T7 RNAP

Templates were prepared according to Sections 6.2.4-6.2.12. Reactions were performed in PCR tubes and consisted of plasmid templates (200 ng), RNAPol Reaction Buffer (2 µL, 10X), regular NTPs (0.1 µL, 100 mM), RNase inhibitor (0.50 µL, 40 U/µL), T7 RNAP (2 µL, 50 U/µL), and made up to 20 µL with MilliQ water. The +UV conditions were irradiated with UV (Section 6.2.18) at 2.47 mW cm^{-2} for 10 mins, whilst other tubes were kept on the bench at r.t. The reactions were incubated on the thermal cycler at 37 ºC for 1 h. DNase I (1 µL) was added to each reaction, and incubated on the thermal cycler at 37 ºC for 10 mins. EDTA (1.10 µL, 100 mM) was added to each reaction, and the reactions were heat-deactivated on the thermal cycler at 75 ºC for

10 mins. 15 µL of the reactions were mixed with RNA Loading Dye (15.0 µL, 2X) and denatured on the thermal cycler at 90 ºC for 2 mins, kept on ice for 2 mins and analysed on an agarose gel (Section 6.2.3). Before electrophoresis, the gel equipment was cleaned with RNaseZap to remove RNase contamination and stored under cling film.

6.4.2. IVTT of the T7 and Tac Promoters

Templates were prepared according to Sections 6.2.4-6.2.12. Reactions were performed in PCR tubes and consisted of plasmid templates (100 ng), NEBexpress® S30 Synthesis Extract (2.4 µL), Protein Synthesis Buffer (5.0 µL, 2X), GamS Nuclease Inhibitor (0.2 µL, 1.5 µg/µL), Murine RNase Inhibitor (0.2 µL, 40 U/µL), T7 RNAP (0.2 µL, 450 U/µL), and made up to 10 µL with MilliQ water. The +UV conditions were irradiated with UV at 2.47 mW cm^{-2} for 10 mins (Section 6.2.18), whilst other tubes were kept on the bench at r.t. The reactions were incubated on the thermal cycler at 37 ºC for 3 h, followed by protein maturation at 4 ºC for 30 mins. The reactions were made up to 40 µL with MilliQ water and the solutions were transferred to a 384-well black, flat-bottom, optiplate. The plates were centrifuged at 1,250 rpm for 1 min to allow solutions to settle and remove bubbles. The fluorescence was measured on the plate reader. Plates were shaken at 654 rpm for 10 s and fluorescence was measured with settings: λ_{ex} = 515 nm, λ_{em} = 527 nm, bandwidth = 5 nm. Data were plotted on Origin.

6.4.3. IVT with Sp6 RNAP

PCR reactions were carried out according to Section 6.2.2 except DreamTaq PCR MM was used as the polymerase, and the reaction was scaled to 50 µL total reaction volume. The PCR reactions were purified with a Genejet PCR purification kit according to the manufacturer's protocol and eluted with Tris (30 µL, 10 mM, pH 8.0). The eluates

were ethanol-precipitated according to Section 6.2.4. The concentration was measured using Pico488 according to Section 6.2.11. Streptavidin binding was performed according to Section 6.2.12.

Reactions were performed in PCR tubes and consisted of DNA templates (500 ng), RNAPol Reaction Buffer (1 µL, 10X), regular NTPs (0.05 µL, 100 mM), RNase inhibitor (0.25 µL, 40 U/µL), DTT (0.05, 1 M), Sp6 RNAP (1 µL, 20 U/µL), and made up to 10 µL with MilliQ water. The +UV conditions were irradiated with UV (Section 6.2.18) at 2.47 mW cm^{-2} for 10 mins, whilst other tubes were kept on the bench at r.t. The reactions were incubated on the thermal cycler at 37 ºC for 2 h. DNase I (1 µL, 2 U/µL) was added to each reaction, and incubated on the thermal cycler at 37 ºC for 15 mins. EDTA (0.578 µL, 100 mM) was added to each reaction, and the reactions were heat-deactivated on the thermal cycler at 75 ºC for 10 mins. 7 µL of the reactions were mixed with RNA Loading Dye (7 µL, 2X) and denatured on the thermal cycler at 90 ºC for 2 mins, kept on ice for 2 mins and analysed on an agarose gel (Section 6.2.3). Before electrophoresis, the gel equipment was cleaned with RNaseZap to remove RNase contamination and stored under cling film.

6.4.4. Plasmid and amino-C6-dT Primer Digestion

The plasmids were digested according to literature[203] with minor emendations. Before digestion, digest mix was prepared and consisted of alkaline phosphatase (10 µL, 20 U/µL), Benzonase (1 µL, 250 U/µL), phosphodiesterase I (15 µL, 20 mU/µL), and the solution was made up to 2.5 mL with Tris (20 mM, pH 7.9). The solution was mixed by inverting and pipetting, and aliquots (50 µL) were stored at -20 ºC. Plasmid digestion reactions were performed in PCR tubes and consisted of plasmid (10 µg), the digest mix (500 µL), and made up to 1000 µL with MilliQ water. The reactions were incubated

on the thermal cycler for 3 h. The amino-C6-dT primer was digested analogously but the reaction was scaled down by a factor of 10.

The Amicon 3 K columns were initially washed with MilliQ water (500 µL) and centrifuged at 4 ºC for 20 mins to remove trace contaminants. The digestion reactions were made up to 450 µL with MilliQ water and applied to the Amicon columns. The reactions were eluted into a new collection tube by centrifugation at 4 ºC for 20 mins. The column was topped up to 500 µL with MilliQ water and re-centrifuged at the above settings. The eluates were combined and lyophilised to dryness. The reactions were resuspended in MilliQ water (50 µL). PVDF columns were washed with MilliQ water (500 µL), and centrifuged at 14,000 g for 2 mins at 4 ºC. The reactions were applied to the columns and centrifuged at 14,000 g for 2 mins at 4 ºC. 10 µL was injected for LC-MS analysis.

6.4.5. HPLC to visualise digested plasmid

Lines B/C/D were: MilliQ water, acetonitrile, and ammonium acetate buffer (100 mM, pH 7.5). Line D was kept constant at 1%. The flow rate was 1.0 mL/min. Solvent percentages refer to line C (organic phase). Separations were UV-monitored at 260 nm.

Plasmid (1 µg) was digested according to Section 6.4.4. After purification, the reactions were resuspended in MilliQ water (150 µL) and 45 µL was injected onto an Agilent Polaris 5 C18-A, (150*4.6 mm) column pre-equilibrated with 1.5% C for ~5 mins. A 1.5-35% gradient over 10 mins was used to separate the nucleosides. The organic phase was increased to 95% at 15.1 mins and maintained until 20.5 mins to wash the column. The organic phase was returned to 1.5% at 20.6 mins and kept until 25 mins to re-equilibrate the column.

6.4.6. Binding MVS-Atto647 to PCB plasmids

PCB plasmids were prepared according to Sections 6.2.4-6.2.11 and 6.3.6. Reactions were performed in 1.5 mL Protein LoBind Tubes and consisted of: PCB plasmids (final concentration = 8.33 ng/µL), MVS-Atto647 (at 5 or 8 molar equivalents relative to estimated PCB sites), Tris (final concentration = 20 mM, 1 M, pH 7.4), and made up to 60 µL with MilliQ water. The reactions were incubated on the thermomixer at 20 ºC for 1 h. Unbound MVS-Atto647 was removed with a Genejet PCR column (washed twice) and eluted with 40 µL Tris (10 mM, pH 7.4). The concentrations were measured using Pico488 (Section 6.2.11).

The below was performed in the dark. The calibration curve of MVS-Atto647 was established with a 2X serial dilution from ~17.5 – 0 nM (8 concentration standards, technical triplicates). Plasmids were made up to 130 µL with MilliQ water. 40 µL was used for each measurement (n = 3); the solutions were transferred to a 384-well black, flat-bottom, optiplate. The plates were centrifuged at 1,250 rpm for 1 min to allow solutions to settle and remove bubbles. The fluorescence was measured on the plate reader. Plates were shaken at 654 rpm for 10 s and fluorescence was measured with settings: λ_{ex} = 650 nm, λ_{em} = 668 nm, bandwidth = 5 nm. Graphs and calculations were made in Excel.

6.4.7. Summary of Primers, PCR, and Homologous Recombination Conditions, and HPLC Details in Chapter 3

Name	Sequence
pTac_nicks_ins_FOR1	GCTCTTCCCTCAGCCGGTAATACGGTTATCCAC
pTac_nicks_ins_REV1	GCCATAACCATGAGTGATAACAC
pTac_nicks_ins_FOR2	GTGTTATCACTCATGGTTATGGC
pTac_nicks_ins_REV2	GCTGAGGGAAGAGCCCTTTGAGTGAGCTGATACC
Sp6_NEEL_FOR	GCTTGATTTAGGTGACACTATAGACTTGTTC
mNG_BbvCI_REV2	CAGCTATGACCATGCTGAGGGATTACGCCAAGC

Table 6.8. Chapter 3 PCR Primer Sequences

PCR	DNA template	For primer	Rev Primer
pTac_mNeonGreen_BbvCI+BspQI BB	pTac1-mNeonGreen	pTac_nicks_ins_FOR1	pTac_nicks_ins_REV1
pTac_mNeonGreen_BbvCI+BspQI insert	pTac1-mNeonGreen	pTac_nicks_ins_FOR2	pTac_nicks_ins_REV2
Sp6_linear_DNA	mNG_BbvCI/Sp6_mNG	Sp6_NEEL_FOR	mNG_BbvCI_REV2

Table 6.9. Chapter 3 PCR reaction compositions

PCR	Tm/ ºC	Extension time/ mins
pTac_mNeonGreen_BbvCI+BspQI BB	58	1
pTac_mNeonGreen_BbvCI+BspQI insert	61	1
Sp6_linear_DNA	66	0.5

Table 6.10. Chapter 3 PCR reaction conditions

Plasmid	Backbone	Insert
pTac_mNG_nicks	pTac_mNeonGreen_BbvCI+BspQI BB	pTac_mNeonGreen_BbvCI+BspQI insert

Table 6.11. Chapter 3 homologous recombination reactions

Name	Sequence
pTac_5 amines	CXXGACAATTAATCATCGGCTCGXAXAAXGTGTG
pTac_7 amines	CXXGACAATXAATCATCGGCXCGXAXAAXGTGTG
SP6_5 amines	GCTTGATXXAGGXGACACXAXAGCTTGTTC
SP6_7 amines	GCTTGATXXAGGXGACACXAXAGCXTGXTC
Tac_2amines1	GTGCXAGCCCCGCTTGACAATTAATCATCGGCTCGTATAATGXGTG
Tac_2amines2	GTGCTAGCCCCGCTTGACAAXTAATCAXCGGCTCGTATAATGTGTG

Table 6.12. Primers used to cage the tac and Sp6 promoters

X denoted the amino-C6-dT modifications. Underlined sequences represented the -35 and -10 regions for the tac promoter and the Sp6 promoter.

Promoter	Number of modifications	Collection period/ mins	Increments/ mins	Fractions pooled
Tac	5	21.3-22.3	0.10	1-5
Tac	7	23.0-24.0	0.10	3-5
Tac	2 (outside -35 and -10 regions)	13.8-14.8	0.10	2-6
Tac	2 (in between -35 and -10 regions)	14.0-15.0	0.10	1-5
Sp6	5	21.0-22.0	0.10	1-6
Sp6	7	23.3-24.3	0.10	1-5

Table 6.13. Purification of tac and Sp6 PCB oligos

6.5. Chapter 4 Methods

6.5.1. Binding WTS to Bt-Cy5

Reactions were performed in 1.5 mL Protein LoBind Tubes and consisted of WTS (20 µL, 1 µg/µL), biotin-Cy5 (1 or 3 molar equivalents relative to WTS), Tris (final

concentration = 20 µL, 1 M, pH 7.4 or 8.0), and made up to 1 mL with MilliQ water. The reactions were incubated on the thermomixer at 26 ºC overnight.

6.5.2. Binding Bt-Tf and Bt-BSA to WTS-Cy5

Reactions were performed in 1.5 mL Protein LoBind Tubes and consisted of: WTS-Cy5 (100 µL, 20 ng/µL), Bt-Tf or Bt-BSA (at 0.1, 0.5, 1, or 3 molar equivalents relative to WTS-Cy5), and made up to 120 µL with MilliQ water. The reactions were incubated on the thermomixer at 26 ºC overnight. The reactions were lyophilised to dryness the next day.

6.5.3 Native PAGE

The 6% acrylamide separating gel solution consisted of acrylamide/bis-acrylamide (2 mL, 30%/0.8% w/v), Tris-HCl (1.973 mL, 1.5 M, pH 8.8), APS (100 µL, 10% w/v), TEMED (10 µL) and made up to 10 mL with MilliQ water. The solution was inverted several times to mix and 8 mL was quickly loaded onto the glass plates for gel casting. Isopropanol was used to fill the space to create an even gel layer, and the solution was allowed to set for 30 mins.

The stacking gel solution was prepared whilst the separating solution polymerised, and consisted of Tris-HCl (1.068 mL, 1.5 M, pH 8.8), acrylamide/bis-acrylamide (0.67 mL, 30%/0.8% w/v), APS (50 µL, 10%), TEMED (5 µL) and made up to 5 mL with MilliQ water. APS and TEMED were added right before use. Once the separating gel solution had set, the isopropanol was poured away, and the stacking gel solution was applied on top, and allowed to set for another 30 mins. Proteins (1 µg) were mixed with homemade 2x loading dye (5 µL) and made up to 10 µL with MilliQ water. The 2X loading dye consisted of glycerol (1 mL, 50% v/v), xylene-cyanol (50 µL, 2% w/v), bromophenol blue (100 µL, 1% w/v), Orange G (100 µL, 1% w/v), EDTA (12.5 µL, 500

mM, pH 8.0). The gel was run in a buffer consisting of Tris-HCl (25 mM final conc.) and glycine (192 mM final conc.) at 200 V for 0.5-1 h. The gel was stained overnight in QuickBlue Protein Stain. The gel was de-stained in MilliQ water for 1 h and imaged.

6.5.4. HPLC Purification of mNG 3-PCBs Oligo

This method had not been optimsied and was initially used to collect mNG 3-PCBs oligo only. Subsequent PCB oligos purification was performed using the method from Section 6.2.14.

Lines B/C/D were: MilliQ water, acetonitrile, and ammonium acetate buffer (100 mM, pH 7.5). Line D was kept constant at 10%. The flow rate was 4.0 mL/min. Solvent percentages refer to line C (organic phase). Separations were UV-monitored at 260, 280, 310, and 360 nm.

The PCB oligo (100 µL, 50 µM) was injected onto a Discovery C-18 column pre-equilibrated with 5% for ~20 mins. A 5-35% gradient over 40 mins was used to separate the oligo. The organic phase was increased to 90% at 40.01 mins and maintained until 47 mins to wash the column. The organic phase was returned to 5% at 47.01 mins and kept until 52 mins to re-equilibrate the column.

6.5.5. Maintaining HEK293T and HeLa-GFP Cells

Full media was prepared from Fluorobrite™ DMEM containing FBS (10% v/v), Non-Essential Amino Acids (final concentration = 1X), and GlutaMAX™ (final concentration = 1X). HEK293T and HeLa-GFP cells were cultured similarly.

When cells reached >80% confluency (every 3-4 days), the media was removed. Cells were washed with PBS (10 mL) which was then removed. Trypsin (2 mL) was added, and cells were incubated at 37 ºC, 5% CO_2 (0.5-1 min for HEK293T, 2-3 mins for

HeLa-GFP, or until the majority of cells were detached). Full media (8 mL) was added and the solution was transferred to a 15-mL Falcon tube and centrifuged at 1,000 rpm for 3 mins. The supernatant was carefully removed. The pellets were resuspended in 10 mL full media and cells were split in a 1:5 – 1:10 ratio with a total volume of 10 mL, depending on cell density, and need. Cells were incubated at 37 ºC, 5% CO_2.

6.5.6. Plating Cells

If cells were to be used for RT-qPCR, the plates were coated with Poly-D-Lysine. Poly-D-Lysine (50 µL) was added to each well. The plates were incubated at r.t. for 5 mins. Poly-D-Lysine was removed and the wells were washed twice with PBS (100 µL each time).

After cell resuspension (Section 6.5.5), cells were counted on the Countess II FL cell counter (Thermo Fisher Scientific) by mixing 10 µL of the cell stock solution with Tryptan Blue (10 µL). Two 10 µL measurements were made and the average was taken and used to calculate the stock volume required for cell splitting solution.

Each well was seeded with 5.5 * 10^4 cells (180 µL) using a Distriman pipette (Gilson) to ensure consistent volumes. Wells receiving UV were seeded on a separate plate from those not receiving UV. The cells were incubated at 37 ºC, 5% CO_2 overnight.

6.5.7. Transfection with Lipofectamine 3000

Templates were prepared according to Sections 6.2.4-6.2.12. Transfection was performed according to the manufacturer's protocol. Briefly, a single Lipofectamine 3000 MM tube was prepared, volumes given were for single reactions and consisted of Opti-MEM (5 µL), and Lipofectamine 3000 (0.2 µL). The contents were mixed by pipetting.

DNA MM was also prepared for each condition. Volumes given were for single reactions and consisted of: Opti-MEM (5 µL), DNA (100 ng), P3000 (0.20 µL, added last), and the reactions were made up to 7.20 µL with MilliQ water.

Lipofectamine 3000 MM (5.20 µL per reaction) was added to DNA MM, and mixed well by pipetting. The reactions were incubated at r.t. for 15 mins to form lipoplexes. The complexes (12.40 µL per well) were added to the wells carefully to avoid introducing bubbles and disturbing the cells. The plates were mixed by gently rocking in circular motions. The plates were incubated at 37 ºC, 5% CO_2. Plates were irradiated after 4 h according to Section 6.2.18.

6.5.8. Epifluorescence Microscopy

Wells were imaged with a Leica DMi8 inverted epifluorescence microscope using a 10x objective lens. Cells were imaged using the brightfield, GFP, and TxRed filter settings. Images were analysed with ImageJ.

6.5.9. Transfection with B-Tf/BSA

Resuspended dried constructs in 20 µL MilliQ water. Streptavidin, Bt-Tf/BSA, or the constructs (1 µg) were added to the wells carefully to avoid introducing bubbles and disturbing the cells. The plates were mixed by gently rocking in circular motions. The plates were incubated at 37 ºC, 5% CO_2. After 4 h, plates were imaged according to Section 6.5.8.

6.5.10. Cell Viability Assay

Cell viability assay was performed according to literature[173] with minor emendations. CellTiter-Glo® Buffer and CellTiter-Glo® Substrate were equilibrated to r.t. CellTiter-Glo® Reagent was prepared by transferring 10 mL of the Buffer to the substrate and

mixed with a stripette and inverting. The mixture was split into 2 mL aliquots and stored in tinfoil at -20 ºC.

Cells were seeded as in Section 6.5.6, except 6.0 x 10^4 cells in 100 µL were seeded per well. UV conditions were applied as in Section 6.2.18. The plates were re-incubated at 37 ºC, 5% CO_2 overnight. The CellTiter-Glo® Reagent (100 µL) was added to each well and the samples were mixed on the thermomixer at 20 ºC with gentle shaking (600 rpm) for 2 mins to induce cell lysis. The plates were incubated at r.t. for 10 mins to stabilise the luminescence signal. Chemiluminescence was measured on the plate reader. The plate was shaken at 654 rpm for 10 s and luminescence was measured with luminescence scan mode: 559-561 nm. Values at 560 nm were used for analysis.

6.5.11. RNA Extraction

RNA extraction was performed according to the manufacturer's guidelines. Initially, a working solution of Buffer RLT containing beta-mercaptoethanol (βME) was prepared (10 µL βME for every 1 mL Buffer RLT). 4 volumes of 99.8% ethanol were also added to Buffer RPE.

The following was performed in the tissue culture hood. The media in each well was removed. The wells were washed with PBS Buffer (100 µL), and PBS was removed. Trypsin (100 µL) was added and the cells were incubated at 37 ºC until the majority of cells were detached. Full media (100 µL) was added and the contents were transferred to 1.5 mL Eppendorf tubes. The rest of the procedure was performed on a sterile bench.

The tubes were centrifuged at 300 g for 5 mins. The supernatant was removed carefully to not disturb the pellets. Buffer RLT (350 µL) was added to each tube and

vortexed vigorously to disrupt and homogenise the cells. Ethanol (70%, 1 volume) was added to each reaction and mixed well by pipetting. The solutions were transferred to RNeasy spin columns and centrifuged at 8,000 g for 15 s, the filtrate was discarded. Buffer RW1 (700 µL) was added to the columns and centrifuged at 8,000 g for 15 s, the filtrate was discarded. Buffer RPE (500 µL) was added and centrifuged at 8,000 g for 15 s, the filtrate was discarded. This step was repeated but centrifuged for 2 mins to completely discard Buffer RPE. The columns were transferred to a new 2 mL collection tube and centrifuged at max speed for 1 min to remove residual Buffer RPE. Columns were transferred to 1.5 mL collection tubes. RNA was eluted with RNase-free water (50 µL) by centrifugation at 8,000 g for 1 min. RNA was ethanol-precipitated (Section 6.2.4) and concentrations were determined using the NanoDrop. 260/230 ratios were typically >1.90. RNAs were stored at -80 ºC.

6.5.12. cDNA Synthesis

Genomic DNA was removed from RNA prep by ezDNase digestion. Reactions were performed in PCR tubes and consisted of RNA templates (60 ng), ezDNase Reaction Buffer (0.6 µL, 10X), ezDNase (0.6 µL), and made up to 6 µL with MilliQ water. The reactions were mixed by pipetting, incubated on the thermal cycler at 37 ºC for 2 mins, and kept at 4 ºC.

To each reaction, SuperScript IV Vilo MM (2.40 µL, 5X) was added, and the reactions were made up to 12 µL with MilliQ water. The reactions were mixed by pipetting, and incubated on the thermal cycler as followed: annealed at 25 ºC for 10 mins, reverse transcribed at 50 ºC for 10 mins, and heat-deactivated at 85 ºC for 5 mins. The reactions were then stored at -80 ºC.

6.5.13. RT-qPCR

cDNA from Section 6.5.12 was used as the template for reverse transcription. Reactions were performed in PCR tubes and consisted of cDNA (2 µL), mNG_qPCR_FOR (0.4 µL, 10 µM), mNG_qPCR_REV (0.4 µL, 10 µM, Table 6.18), Fast SYBR Green MM (10 µL, 2X), and made up to 20 µL with MilliQ water. Similar reactions were used to amplify the GAPDH housekeeping gene and used GAPDH_FOR_Origene and GAPDH_REV_IDT as the forward and reverse primers, respectively. Reactions were incubated on the CFX96 real-Time System (C1000 Touch^{TM} Thermalcycler, Bio-Rad) with SYBR Green monitoring as follows: denaturation at 95 ºC for 20 s, 30 cycles of [95 ºC for 3 s, 60 ºC for 30 s]. The ΔΔCt method was used to calculate transcription fold-change. Data were plotted using Excel.

6.5.14. Fluorescence Plate Reader of mNG Plasmids

After 2 dpt, plates were assessed for mNG expression on the plate reader. Plates were shaken at 654 rpm for 10 s and fluorescence was measured with settings: λ_{ex} = 504 nm, λ_{em} = 517 nm, bandwidth = 5 nm. Data were plotted on Origin and background-substracted.

6.5.15. Summary of Primers, PCR, and Homologous Recombination Conditions, and HPLC Details in Chapter 4

Name	Sequence
mNG_nick_mutagenesis_FOR1	GCGTATTGGGCGAAGAGCCGCTTCCTCG
mNG_nick_mutagenesis_REV1	CGAGAGGCCTTGAATTCGAATCGATG
mNG_nick_mutagenesis_FOR2	CATCGATTCGAATTCAAGGCCTCTCG
mNG_nick_mutagenesis_REV2	CGAGGAAGCGGCTCTTCGCCCAATACGC
mNG_BbvCI_FOR1	GCTTGGCGTAATCCCTCAGCATGGTCATAGCTG
mNG_BbvCI_REV2	CAGCTATGACCATGCTGAGGGATTACGCCAAGC

Table 6.14. Chapter 4 PCR Primer Sequences

PCR	DNA template	For primer	Rev Primer
mNG_inv_BB	CMV_mNG	mNG_nick_mutagenesis_FOR1	mNG_nick_mutagenesis_REV1
mNG_inv_insert	CMV_mNG	mNG_nick_mutagenesis_FOR2	mNG_nick_mutagenesis_REV2
mNG_BbvCI_BB	CMV_mNG	mNG_BbvCI_FOR1	mNG_nick_mutagenesis_REV1
mNG_BbvCI_insert	CMV_mNG	mNG_BbvCI_REV2	mNG_nick_mutagenesis_FOR2

Table 6.15. Chapter 4 PCR reaction compositions

PCR	Tm/ ºC	Extension time/ mins
mNG_inv_BB	62	2.5
mNG_inv_insert	60	0.5
mNG_BbvCI_BB	68	2.5
mNG_BbvCI_insert	68	0.5

Table 6.16. Chapter 4 PCR reaction conditions

Plasmid	Backbone	Insert
mNG_inv	mNG_inv_BB	mNG_inv_insert
mNG_BbvCI	mNG_BbvCI_BB	mNG_BbvCI_insert

Table 6.17. Chapter 4 homologous recombination reactions

Name	Sequence
mNG_qPCR_FOR	CGTGTTCCGTAAGACGGAGC
mNG_RT-qPCR_REV	CCTTGTACAGCTCGTCCATGC
GAPDH_FOR_Origene	GTCTCCTCTGACTTCAACAGCG
GAPDH_REV_IDT	TCCACCACCCTGTTGCTGTA

Table 6.18. Chapter 4 RT-qPCR Primer Sequences

Name	Sequence
mNG_CMV_3amine_FOR	**GAGGTC**XAXAXAAG**CAATGCT**CG

mNG_8amines_31b	CCXAAXG**GGAGGXC**XAXAXAAG**CAAXGCX**CG
mNG_9amines_46b	**GGXC**XAXAXAAG**CAAXGCX**CGTTTAGGGAACCG**CCAT XCX**GCCXGG

Table 6.19. Primers used to cage the CMV promoter

Underlined sequence represented the TATA box recognition sequence. Bold sequences immediately upstream and downstream of the TATA box represented the BRE^{u} and BRE^{d}, respectively. The bold sequence further downstream represented the Inr recognition sequence. X represented amino-C6-dT modifications.

Promoter	Number of modifications	Collection period/ mins	Increments/ mins	Fractions pooled
CMV	3	27.0-28.6	0.20	4-6
CMV	8	24.0-25.5	0.15	7-9
CMV	9	24.0-25.5	0.15	3-5

Table 6.20. Purification of CMV PCB oligos

Bibliography

1. Gunge, N. Yeast Dna Plasmids. *Annual Review of Microbiology* **37**, 253–276 (1983).
2. Billane, K., Harrison, E., Cameron, D. & Brockhurst, M. A. Why do plasmids manipulate the expression of bacterial phenotypes? *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**, 20200461 (2021).
3. del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Díaz-Orejas, R. Replication and Control of Circular Bacterial Plasmids. *Microbiology and Molecular Biology Reviews* **62**, 434–464 (1998).
4. Helinski, D. R. A Brief History of Plasmids. *EcoSal Plus* **10**, eESP-0028-2021 (2022).
5. Khan, S. *et al.* Role of Recombinant DNA Technology to Improve Life. *International Journal of Genomics* **2016**, e2405954 (2016).
6. Dussoix, D. & Arber, W. Host specificity of DNA produced by Escherichia coli: II. Control over acceptance of DNA from infecting phage λ. *Journal of Molecular Biology* **5**, 37–49 (1962).
7. Meselson, M. & Yuan, R. DNA Restriction Enzyme from E. coli. *Nature* **217**, 1110–1114 (1968).
8. Lederberg, S. & Meselson, M. Degradation of non-replicating bacteriophage DNA in non-accepting cells. *Journal of Molecular Biology* **8**, 623–628 (1964).
9. Zimmerman, S. B., Little, J. W., Oshinsky, C. K. & Gellert, M. Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide. *Proceedings of the National Academy of Sciences* **57**, 1841–1848 (1967).

10. Cohen, S. N., Chang, A. C. Y. & Hsu, L. Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of Escherichia coli by R-Factor DNA*. *Proceedings of the National Academy of Sciences* **69**, 2110–2114 (1972).

11. Mertz, J. E. & Davis, R. W. Cleavage of DNA by R1 Restriction Endonuclease Generates Cohesive Ends. *Proceedings of the National Academy of Sciences* **69**, 3370–3374 (1972).

12. Jackson, D. A., Symons, R. H. & Berg, P. Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of Escherichia coli. *Proceedings of the National Academy of Sciences* **69**, 2904–2909 (1972).

13. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proceedings of the National Academy of Sciences* **70**, 3240–3244 (1973).

14. Dower, W. J. Electroporation of Bacteria: A General Approach to Genetic Transformation. in *Genetic Engineering: Principles and Methods* (ed. Setlow, J. K.) 275–295 (Springer US, 1990). doi:10.1007/978-1-4613-0641-2_14.

15. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).

16. Gibson, D. G. *et al.* Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science* **319**, 1215–1220 (2008).

17. Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52–56 (2010).

18. Gibson, D. G., Smith, H. O., Hutchison, C. A., Venter, J. C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat Methods* **7**, 901–903 (2010).

19. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE* **4**, e5553 (2009).

20. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* **3**, e3647 (2008).

21. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE* **6**, e16765 (2011).

22. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).

23. Emtage, J. S. *et al.* Synthesis of calf prochymosin (prorennin) in Escherichia coli. *Proceedings of the National Academy of Sciences* **80**, 3671–3675 (1983).

24. Johnson, M. E. & Lucey, J. A. Major Technological Advances and Trends in Cheese. *Journal of Dairy Science* **89**, 1174–1178 (2006).

25. Ye, X. *et al.* Engineering the Provitamin A (β-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. *Science* **287**, 303–305 (2000).

26. Paine, J. A. *et al.* Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat Biotechnol* **23**, 482–487 (2005).

27. Des Soye, B. J., Davidson, S. R., Weinstock, M. T., Gibson, D. G. & Jewett, M. C. Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens. *ACS Synth. Biol.* **7**, 2245–2255 (2018).

28. Tripathi, N. K. & Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology* **7**, (2019).

29. Manzur, M. J. *et al.* Production of recombinant enzymes of wide use for research. *Electron. J. Biotechnol.* **9**, 0–0 (2006).

30. Tuckey, C., Asahara, H., Zhou, Y. & Chong, S. Protein Synthesis Using a Reconstituted Cell-Free System. *Current Protocols in Molecular Biology* **108**, 16.31.1-16.31.22 (2014).

31. Martemyanov, K. A., Shirokov, V. A., Kurnasov, O. V., Gudkov, A. T. & Spirin, A. S. Cell-Free Production of Biologically Active Polypeptides: Application to the Synthesis of Antibacterial Peptide Cecropin. *Protein Expression and Purification* **21**, 456–461 (2001).

32. Kim, D.-M. & Swartz, J. R. Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of Escherichia coli. *Biotechnology and Bioengineering* **85**, 122–129 (2004).

33. Yin, G. & Swartz, J. R. Enhancing multiple disulfide bonded protein folding in a cell-free system. *Biotechnology and Bioengineering* **86**, 188–195 (2004).

34. Kai, L., Dötsch, V., Kaldenhoff, R. & Bernhard, F. Artificial Environments for the Co-Translational Stabilization of Cell-Free Expressed Proteins. *PLOS ONE* **8**, e56637 (2013).

35. Hartman, M. C. T., Josephson, K., Lin, C.-W. & Szostak, J. W. An Expanded Set of Amino Acid Analogs for the Ribosomal Translation of Unnatural Peptides. *PLOS ONE* **2**, e972 (2007).

36. Stark, J. C. *et al.* On-demand biomanufacturing of protective conjugate vaccines. *Science Advances* **7**, eabe9444 (2021).

37. Warfel, K. F. *et al.* A Low-Cost, Thermostable, Cell-Free Protein Synthesis Platform for On-Demand Production of Conjugate Vaccines. *ACS Synth. Biol.* **12**, 95–107 (2023).

38. Gregorio, N. E., Levine, M. Z. & Oza, J. P. A User's Guide to Cell-Free Protein Synthesis. *Methods and Protocols* **2**, 24 (2019).

39. Dudley, Q. M., Karim, A. S. & Jewett, M. C. Cell-free metabolic engineering: Biomanufacturing beyond the cell. *Biotechnology Journal* **10**, 69–82 (2015).

40. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat Biotechnol* **19**, 751–755 (2001).

41. Wang, H. H. *et al.* Multiplexed in Vivo His-Tagging of Enzyme Pathways for in Vitro Single-Pot Multienzyme Catalysis. *ACS Synth. Biol.* **1**, 43–52 (2012).

42. Villarreal, F. *et al.* Synthetic microbial consortia enable rapid assembly of pure translation machinery. *Nat Chem Biol* **14**, 29–35 (2018).

43. Shepherd, T. R. *et al.* De novo design and synthesis of a 30-cistron translation-factor module. *Nucleic Acids Research* **45**, 10895–10905 (2017).

44. Lavickova, B. & Maerkl, S. J. A Simple, Robust, and Low-Cost Method To Produce the PURE Cell-Free System. *ACS Synth. Biol.* **8**, 455–462 (2019).

45. Woof, J. M. & Burton, D. R. Human antibody–Fc receptor interactions illuminated by crystal structures. *Nat Rev Immunol* **4**, 89–99 (2004).

46. Roth, K. D. R. *et al.* Developing Recombinant Antibodies by Phage Display Against Infectious Diseases and Toxins for Diagnostics and Therapy. *Frontiers in Cellular and Infection Microbiology* **11**, (2021).

47. Hammers, C. M. & Stanley, J. R. Antibody Phage Display: Technique and Applications. *Journal of Investigative Dermatology* **134**, 1–5 (2014).

48. Peterson, E., Owens, S. M. & Henry, R. L. Monoclonal antibody form and function: Manufacturing the right antibodies for treating drug abuse. *AAPS J* **8**, 43 (2006).

49. Salazar, A. *et al. Allergen-Based Diagnostic: Novel and Old Methodologies with New Approaches*. *Allergen* (IntechOpen, 2017). doi:10.5772/intechopen.69276.

50. Kramer, R. A. *et al.* The human antibody repertoire specific for rabies virus glycoprotein as selected from immune libraries. *European Journal of Immunology* **35**, 2131–2145 (2005).

51. Gu, T. *et al.* Identification of Binding Epitope for Anti-Rabies Virus Glycoprotein Single-Chain Fv Fragment FV57. *Protein & Peptide Letters* **18**, 1099–1106.

52. Yuan, R. *et al.* Preparation and diagnostic use of a novel recombinant single-chain antibody against rabies virus glycoprotein. *Appl Microbiol Biotechnol* **98**, 1547–1555 (2014).

53. Baeshen, N. A. *et al.* Cell factories for insulin production. *Microbial Cell Factories* **13**, 141 (2014).

54. Kitabchi, A. E., Umpierrez, G. E., Miles, J. M. & Fisher, J. N. Hyperglycemic Crises in Adult Patients With Diabetes. *Diabetes Care* **32**, 1335–1343 (2009).

55. Commissioner, O. of the. 100 Years of Insulin. *FDA* (2022).

56. Goeddel, D. V. *et al.* Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences* **76**, 106–110 (1979).

57. Blundell, T., Dodson, G., Hodgkin, D. & Mercola, D. Insulin: The Structure in the Crystal and its Reflection in Chemistry and Biology by. in *Advances in Protein Chemistry* (eds. Anfinsen, C. B., Edsall, J. T. & Richards, F. M.) vol. 26 279–402 (Academic Press, 1972).

58. Walsh, G. Therapeutic insulins and their large-scale manufacture. *Appl Microbiol Biotechnol* **67**, 151–159 (2005).

59. Liu, M. A. A Comparison of Plasmid DNA and mRNA as Vaccine Technologies. *Vaccines* **7**, 37 (2019).

60. Shafaati, M. *et al.* A brief review on DNA vaccines in the era of COVID-19. *Future Virology* **17**, 49–66 (2022).

61. Ye, T., Zhong, Z., García-Sastre, A., Schotsaert, M. & De Geest, B. G. Current Status of COVID-19 (Pre)Clinical Vaccine Development. *Angewandte Chemie International Edition* **59**, 18885–18897 (2020).

62. Wolff, J. A. *et al.* Direct Gene Transfer into Mouse Muscle in Vivo. *Science* **247**, 1465–1468 (1990).

63. Ulmer, J. B. *et al.* Heterologous Protection Against Influenza by Injection of DNA Encoding a Viral Protein. *Science* **259**, 1745–1749 (1993).

64. Martinon, F. *et al.* Induction of virus-specific cytotoxic T lymphocytes in vivo by liposome-entrapped mRNA. *European Journal of Immunology* **23**, 1719–1722 (1993).

65. Durymanov, M. & Reineke, J. Non-viral Delivery of Nucleic Acids: Insight Into Mechanisms of Overcoming Intracellular Barriers. *Frontiers in Pharmacology* **9**, (2018).

66. Nishikawa, M. & Huang, L. Nonviral Vectors in the New Millennium: Delivery Barriers in Gene Transfer. *Human Gene Therapy* **12**, 861–870 (2001).

67. Gaudinski, M. R. *et al.* Safety, tolerability, and immunogenicity of two Zika virus DNA vaccine candidates in healthy adults: randomised, open-label, phase 1 clinical trials. *The Lancet* **391**, 552–562 (2018).

68. Alberer, M. *et al.* Safety and immunogenicity of a mRNA rabies vaccine in healthy adults: an open-label, non-randomised, prospective, first-in-human phase 1 clinical trial. *The Lancet* **390**, 1511–1520 (2017).

69. Houseley, J. & Tollervey, D. The Many Pathways of RNA Degradation. *Cell* **136**, 763–776 (2009).

70. Cherng, J.-Y. *et al.* Effect of DNA topology on the transfection efficiency of poly((2-dimethylamino)ethyl methacrylate)–plasmid complexes. *Journal of Controlled Release* **60**, 343–353 (1999).

71. Remaut, K., Sanders, N. N., Fayazpour, F., Demeester, J. & De Smedt, S. C. Influence of plasmid DNA topology on the transfection properties of DOTAP/DOPE lipoplexes. *Journal of Controlled Release* **115**, 335–343 (2006).

72. Schmid, A. Considerations for Producing mRNA Vaccines for Clinical Trials. in *RNA Vaccines: Methods and Protocols* (eds. Kramps, T. & Elbers, K.) 237–251 (Springer, 2017). doi:10.1007/978-1-4939-6481-9_15.

73. Reautschnig, P., Vogel, P. & Stafforst, T. The notorious R.N.A. in the spotlight - drug or target for the treatment of disease. *RNA Biology* **14**, 651–668 (2017).

74. Doener, F. *et al.* RNA-based adjuvant CV8102 enhances the immunogenicity of a licensed rabies vaccine in a first-in-human trial. *Vaccine* **37**, 1819–1826 (2019).

75. Sahin, U., Karikó, K. & Türeci, Ö. mRNA-based therapeutics — developing a new class of drugs. *Nat Rev Drug Discov* **13**, 759–780 (2014).

76. Karikó, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA Recognition by Toll-like Receptors: The Impact of Nucleoside Modification and the Evolutionary Origin of RNA. *Immunity* **23**, 165–175 (2005).

77. Karikó, K. *et al.* Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. *Molecular Therapy* **16**, 1833–1840 (2008).

78. Lechardeur, D. *et al.* Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. *Gene Ther* **6**, 482–497 (1999).

79. Lu, S. Heterologous prime–boost vaccination. *Current Opinion in Immunology* **21**, 346–351 (2009).

80. Wolff, J. A., Ludtke, J. J., Acsadi, G., Williams, P. & Jani, A. Long-term persistence of plasmid DNA and foreign gone expression in mouse muscle. *Human Molecular Genetics* **1**, 363–369 (1992).

81. Nichols, W. W., Ledwith, B. J., Manam, S. V. & Troilo, P. J. Potential DNA Vaccine Integration into Host Cell Genome. *Annals of the New York Academy of Sciences* **772**, 30–39 (1995).

82. Sheets, R. L. *et al.* Biodistribution of DNA plasmid vaccines against HIV-1, Ebola, Severe Acute Respiratory Syndrome, or West Nile virus is similar, without integration, despite differing plasmid backbones or gene inserts. *Toxicol Sci* **91**, 610–619 (2006).

83. Ledwith, B. J. *et al.* Plasmid DNA Vaccines: Investigation of Integration into Host Cellular DNA following Intramuscular Injection in Mice. *INT* **43**, 258–272 (2000).

84. Cichutek, K. DNA Vaccines: Development, Standardization and Regulation. *INT* **43**, 331–338 (2000).

85. Wang, Z. *et al.* Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Ther* **11**, 711–721 (2004).

86. Kutzler, M. A. & Weiner, D. B. DNA vaccines: ready for prime time? *Nat Rev Genet* **9**, 776–788 (2008).

87. Klinman, D. M., Klaschik, S., Tross, D., Shirota, H. & Steinhagen, F. FDA guidance on prophylactic DNA vaccines: Analysis and recommendations. *Vaccine* **28**, 2801–2805 (2010).

88. Babiuk, S., Babiuk, L. A. & van Drunen Littel-van den Hurk, S. Editorial: DNA Vaccination: A Simple Concept with Challenges Regarding Implementation. *International Reviews of Immunology* **25**, 51–81 (2006).

89. Vaine, M., Wang, S., Hackett, A., Arthos, J. & Lu, S. Antibody responses elicited through homologous or heterologous prime-boost DNA and protein vaccinations differ in functional activity and avidity. *Vaccine* **28**, 2999–3007 (2010).

90. Eo, S. K., Lee, S., Chun, S. & Rouse, B. T. Modulation of Immunity against Herpes Simplex Virus Infection via Mucosal Genetic Transfer of Plasmid DNA Encoding Chemokines. *Journal of Virology* **75**, 569–578 (2001).

91. Uchijima, M., Yoshida, A., Nagata, T. & Koide, Y. Optimization of Codon Usage of Plasmid DNA Vaccine Is Required for the Effective MHC Class I-Restricted T Cell Responses Against an Intracellular Bacterium1. *The Journal of Immunology* **161**, 5594–5599 (1998).

92. Powell, K. DNA vaccines—back in the saddle again? *Nature Biotechnology* **22**, 799–801 (2004).

93. Sheridan, C. First COVID-19 DNA vaccine approved, others in hot pursuit. *Nature Biotechnology* **39**, 1479–1482 (2021).

94. Mallapaty, S. India's DNA COVID vaccine is a world first – more are coming. *Nature* **597**, 161–162 (2021).

95. Jorritsma, S. H. T., Gowans, E. J., Grubor-Bauk, B. & Wijesundara, D. K. Delivery methods to increase cellular uptake and immunogenicity of DNA vaccines. *Vaccine* **34**, 5488–5494 (2016).

96. Graham, B. S. *et al.* DNA Vaccine Delivered by a Needle-Free Injection Device Improves Potency of Priming for Antibody and CD8+ T-Cell Responses after rAd5 Boost in a Randomized Clinical Trial. *PLOS ONE* **8**, e59340 (2013).

97. Dean, H. J., Haynes, J. & Schmaljohn, C. The role of particle-mediated DNA vaccines in biodefense preparedness. *Advanced Drug Delivery Reviews* **57**, 1315–1342 (2005).

98. Fuller, D. H., Loudon, P. & Schmaljohn, C. Preclinical and clinical progress of particle-mediated DNA vaccines for infectious diseases. *Methods* **40**, 86–97 (2006).

99. Prazeres, D. M. F. & Monteiro, G. A. Plasmid Biopharmaceuticals. *Microbiology Spectrum* **2**, 2.6.02 (2014).

100. Kim, T. J. *et al.* Clearance of persistent HPV infection and cervical lesion by therapeutic DNA vaccine in CIN3 patients. *Nat Commun* **5**, 5317 (2014).

101. Gottlieb, P., Utz, P. J., Robinson, W. & Steinman, L. Clinical optimization of antigen specific modulation of type 1 diabetes with the plasmid DNA platform. *Clinical Immunology* **149**, 297–306 (2013).

102. Gary, E. N. & Weiner, D. B. DNA vaccines: prime time is now. *Current Opinion in Immunology* **65**, 21–27 (2020).

103. Ledgerwood, J. E. *et al.* A West Nile Virus DNA Vaccine Utilizing a Modified Promoter Induces Neutralizing Antibody in Younger and Older Healthy Adults in a Phase I Clinical Trial. *The Journal of Infectious Diseases* **203**, 1396–1404 (2011).

104. Sarwar, U. N. *et al.* Safety and immunogenicity of DNA vaccines encoding Ebolavirus and Marburgvirus wild-type glycoproteins in a phase I clinical trial. *J Infect Dis* **211**, 549–557 (2015).

105. Alton, E. *et al.* Cationic lipid-mediated CFTR gene transfer to the lungs and nose of patients with cystic fibrosis: a double-blind placebo-controlled trial. *The Lancet* **353**, 947–954 (1999).

106. Croze, F. & Prud'homme, G. J. Gene therapy of streptozotocin-induced diabetes by intramuscular delivery of modified preproinsulin genes. *The Journal of Gene Medicine* **5**, 425–437 (2003).

107. Marqus, S., Pirogova, E. & Piva, T. J. Evaluation of the use of therapeutic peptides for cancer treatment. *Journal of Biomedical Science* **24**, 21 (2017).

108. Pahle, J. & Walther, W. Bacterial Toxins for Oncoleaking Suicidal Cancer Gene Therapy. in *Current Strategies in Cancer Gene Therapy* (ed. Walther, W.) 95–110 (Springer International Publishing, 2016). doi:10.1007/978-3-319-42934-2_7.

109. Gómez-Navarro, J., Arafat, W. & Xiang, J. Gene therapy for carcinoma of the breast: Pro-apoptotic gene therapy. *Breast Cancer Research* **2**, 32 (2000).

110. Ardiani, A. *et al.* Enzymes To Die For: Exploiting Nucleotide Metabolizing Enzymes for Cancer Gene Therapy. *Current Gene Therapy* **12**, 77–91.

111. Zeitelhofer, M. *et al.* High-efficiency transfection of short hairpin RNAs-encoding plasmids into primary hippocampal neurons. *Journal of Neuroscience Research* **87**, 289–300 (2009).

112. Kaykas, A. & Moon, R. T. A plasmid-based system for expressing small interfering RNA libraries in mammalian cells. *BMC Cell Biology* **5**, 16 (2004).

113. Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol* **11**, 373–384 (2010).

114. Takeuchi, O. & Akira, S. Pattern Recognition Receptors and Inflammation. *Cell* **140**, 805–820 (2010).

115. Reyes-Sandoval, A. & Ertl, H. C. J. CpG Methylation of a Plasmid Vector Results in Extended Transgene Product Expression by Circumventing Induction of Immune Responses. *Molecular Therapy* **9**, 249–261 (2004).

116. Lesueur, L. L., Mir, L. M. & André, F. M. Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro. *Molecular Therapy - Nucleic Acids* **5**, (2016).

117. Hornstein, B. D., Roman, D., Arévalo-Soliz, L. M., Engevik, M. A. & Zechiedrich, L. Effects of Circular DNA Length on Transfection Efficiency by Electroporation into HeLa Cells. *PLOS ONE* **11**, e0167537 (2016).

118. Kreiss, P. *et al.* Plasmid DNA size does not affect the physicochemical properties of lipoplexes but modulates gene transfer efficiency. *Nucleic Acids Research* **27**, 3792–3798 (1999).

119. Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N. & Wuertz, S. Studying plasmid horizontal transfer in situ: a critical review. *Nat Rev Microbiol* **3**, 700–710 (2005).

120. Oliveira, P. H. & Mairhofer, J. Marker-free plasmids for biotechnological applications - implications and perspectives. *Trends Biotechnol* **31**, 539–547 (2013).

121. Lu, J., Zhang, F., Xu, S., Fire, A. Z. & Kay, M. A. The Extragenic Spacer Length Between the 5′ and 3′ Ends of the Transgene Expression Cassette Affects Transgene Silencing From Plasmid-based Vectors. *Molecular Therapy* **20**, 2111–2119 (2012).

122. Darquet, A.-M., Cameron, B., Wils, P., Scherman, D. & Crouzet, J. A new DNA vehicle for nonviral gene delivery: supercoiled minicircle. *Gene Ther* **4**, 1341–1349 (1997).

123. Mayrhofer, P., Blaesen, M., Schleef, M. & Jechlinger, W. Minicircle-DNA production by site specific recombination and protein–DNA interaction chromatography. *The Journal of Gene Medicine* **10**, 1253–1269 (2008).

124. Kay, M. A., He, C.-Y. & Chen, Z.-Y. A robust system for production of minicircle DNA vectors. *Nat Biotechnol* **28**, 1287–1289 (2010).

125. Azzoni, A. R., Ribeiro, S. C., Monteiro, G. A. & Prazeres, D. M. F. The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. *The Journal of Gene Medicine* **9**, 392–402 (2007).

126. Ribeiro, S. C., Monteiro, G. A. & Prazeres, D. M. F. The role of polyadenylation signal secondary structures on the resistance of plasmid vectors to nucleases. *The Journal of Gene Medicine* **6**, 565–573 (2004).

127. Walther, W., Stein, U., Siegel, R., Fichtner, I. & Schlag, P. M. Use of the nuclease inhibitor aurintricarboxylic acid (ATA) for improved non-viral intratumoral in vivo gene transfer by jet-injection. *The Journal of Gene Medicine* **7**, 477–485 (2005).

128. Ross, G. F. *et al.* Enhanced reporter gene expression in cells transfected in the presence of DMI-2, an acid nuclease inhibitor. *Gene Ther* **5**, 1244–1250 (1998).

129. Oliveira, P. H., Prather, K. J., Prazeres, D. M. F. & Monteiro, G. A. Structural instability of plasmid biopharmaceuticals: challenges and implications. *Trends in Biotechnology* **27**, 503–511 (2009).

130. Arabi, F., Mansouri, V. & Ahmadbeigi, N. Gene therapy clinical trials, where do we go? An overview. *Biomedicine & Pharmacotherapy* **153**, 113324 (2022).

131. Deev, R. *et al.* Results of 5-year follow-up study in patients with peripheral artery disease treated with PL-VEGF165 for intermittent claudication. *Therapeutic Advances in Cardiovascular Disease* **12**, 237–246 (2018).

132. Ylä-Herttuala, S. Gene Therapy of Critical Limb Ischemia Enters Clinical Use. *Molecular Therapy* **27**, 2053 (2019).

133. Suda, H., Murakami, A., Kaga, T., Tomioka, H. & Morishita, R. Beperminogene perplasmid for the treatment of critical limb ischemia. *Expert Review of Cardiovascular Therapy* **12**, 1145–1156 (2014).

134. Norgren, L. *et al.* Inter-Society Consensus for the Management of Peripheral Arterial Disease (TASC II). *Journal of Vascular Surgery* **45**, S5–S67 (2007).

135. Yamada, M., Nagasaki, S. C., Ozawa, T. & Imayoshi, I. Light-mediated control of Gene expression in mammalian cells. *Neuroscience Research* **152**, 66–77 (2020).

136. Das, A. T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Current Gene Therapy* **16**, 156–167.

137. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences* **89**, 5547–5551 (1992).

138. Gossen, M. *et al.* Transcriptional Activation by Tetracyclines in Mammalian Cells. *Science* **268**, 1766–1769 (1995).

139. Ramos, J. L. *et al.* The TetR Family of Transcriptional Repressors. *Microbiology and Molecular Biology Reviews* **69**, 326–356 (2005).

140. Kallunki, T., Barisic, M., Jäättelä, M. & Liu, B. How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells* **8**, 796 (2019).

141. Cambridge, S. B. *et al.* Doxycycline-dependent photoactivated gene expression in eukaryotic systems. *Nat Methods* **6**, 527–531 (2009).

142. Kamiya, Y. *et al.* Synthetic Gene Involving Azobenzene-Tethered T7 Promoter for the Photocontrol of Gene Expression by Visible Light. *ACS Synth. Biol.* **4**, 365–370 (2015).

143. Venancio-Marques, A. *et al.* Modification-Free Photocontrol of β-Lactam Conversion with Spatiotemporal Resolution. *ACS Synth. Biol.* **1**, 526–531 (2012).

144. Estévez-Torres, A. *et al.* Sequence-independent and reversible photocontrol of transcription/expression systems using a photosensitive nucleic acid binder. *Proceedings of the National Academy of Sciences* **106**, 12219–12223 (2009).

145. Hartmann, D., Smith, J. M., Mazzotti, G., Chowdhry, R. & Booth, M. J. Controlling gene expression with light: a multidisciplinary endeavour. *Biochemical Society Transactions* **48**, 1645–1659 (2020).

146. Hughes, R. M. A compendium of chemical and genetic approaches to light-regulated gene transcription. *Critical Reviews in Biochemistry and Molecular Biology* **53**, 453–474 (2018).

147. Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. A light-switchable gene promoter system. *Nat Biotechnol* **20**, 1041–1044 (2002).

148. Müller, K. *et al.* A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. *Nucleic Acids Research* **41**, e77 (2013).

149. Möglich, A., Ayers, R. A. & Moffat, K. Design and Signaling Mechanism of Light-Regulated Histidine Kinases. *Journal of Molecular Biology* **385**, 1433–1444 (2009).

150. Ohlendorf, R., Vidavski, R. R., Eldar, A., Moffat, K. & Möglich, A. From Dusk till Dawn: One-Plasmid Systems for Light-Regulated Gene Expression. *Journal of Molecular Biology* **416**, 534–542 (2012).

151. Ong, N. T., Olson, E. J. & Tabor, J. J. Engineering an E. coli Near-Infrared Light Sensor. *ACS Synth. Biol.* **7**, 240–248 (2018).

152. Yazawa, M., Sadaghiani, A. M., Hsueh, B. & Dolmetsch, R. E. Induction of protein-protein interactions in live cells using light. *Nat Biotechnol* **27**, 941–945 (2009).

153. Taslimi, A. *et al.* Optimized second-generation CRY2–CIB dimerizers and photoactivatable Cre recombinase. *Nat Chem Biol* **12**, 425–430 (2016).

154. Niopek, D. *et al.* Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat Commun* **5**, 4404 (2014).

155. Yumerefendi, H. *et al.* Control of Protein Activity and Cell Fate Specification via Light-Mediated Nuclear Translocation. *PLOS ONE* **10**, e0128443 (2015).

156. Yumerefendi, H. *et al.* Light-induced nuclear export reveals rapid dynamics of epigenetic modifications. *Nat Chem Biol* **12**, 399–401 (2016).

157. Niopek, D., Wehler, P., Roensch, J., Eils, R. & Di Ventura, B. Optogenetic control of nuclear protein export. *Nat Commun* **7**, 10624 (2016).

158. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).

159. Tabor, J. J., Levskaya, A. & Voigt, C. A. Multichromatic Control of Gene Expression in Escherichia coli. *Journal of Molecular Biology* **405**, 315–324 (2011).

160. Levskaya, A. *et al.* Engineering Escherichia coli to see light. *Nature* **438**, 441–442 (2005).

161. Kennedy, M. J. *et al.* Rapid blue-light–mediated induction of protein interactions in living cells. *Nat Methods* **7**, 973–975 (2010).

162. Jayaraman, P. *et al.* Blue light-mediated transcriptional activation and repression of gene expression in bacteria. *Nucleic Acids Research* **44**, 6994–7005 (2016).

163. Motta-Mena, L. B. *et al.* An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nat Chem Biol* **10**, 196–202 (2014).

164. Ye, H., Baba, M. D.-E., Peng, R.-W. & Fussenegger, M. A Synthetic Optogenetic Transcription Device Enhances Blood-Glucose Homeostasis in Mice. *Science* **332**, 1565–1568 (2011).

165. Schmidl, S. R., Sheth, R. U., Wu, A. & Tabor, J. J. Refactoring and Optimization of Light-Switchable Escherichia coli Two-Component Systems. *ACS Synth. Biol.* **3**, 820–831 (2014).

166. Monroe, W. T., McQuain, M. M., Chang, M. S., Alexander, J. S. & Haselton, F. R. Targeting Expression with Light Using Caged DNA*. *Journal of Biological Chemistry* **274**, 20895–20900 (1999).

167. Schroeder, A. *et al.* Remotely Activated Protein-Producing Nanoparticles. *Nano Lett.* **12**, 2685–2689 (2012).

168. Jayakumar, M. K. G., Idris, N. M. & Zhang, Y. Remote activation of biomolecules in deep tissues using near-infrared-to-UV upconversion nanotransducers. *Proceedings of the National Academy of Sciences* **109**, 8483–8488 (2012).

169. Frangioni, J. V. In vivo near-infrared fluorescence imaging. *Current Opinion in Chemical Biology* **7**, 626–634 (2003).

170. Polesskaya, O. *et al.* Optogenetic regulation of transcription. *BMC Neuroscience* **19**, 12 (2018).

171. Ando, H., Furuta, T., Tsien, R. Y. & Okamoto, H. Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nat Genet* **28**, 317–325 (2001).

172. Yamaguchi, S., Chen, Y., Nakajima, S., Furuta, T. & Nagamune, T. Light-activated gene expression from site-specific caged DNA with a biotinylated photolabile protection group. *Chem. Commun.* **46**, 2244–2246 (2010).

173. Hemphill, J., Govan, J., Uprety, R., Tsang, M. & Deiters, A. Site-Specific Promoter Caging Enables Optochemical Gene Activation in Cells and Animals. *J. Am. Chem. Soc.* **136**, 7152–7158 (2014).

174. Ogasawara, S. Transcription Driven by Reversible Photocontrol of Hyperstable G-Quadruplexes. *ACS Synth. Biol.* **7**, 2507–2513 (2018).

175. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proceedings of the National Academy of Sciences* **82**, 488–492 (1985).

176. Huang, R., Fang, P. & Kay, B. K. Improvements to the Kunkel mutagenesis protocol for constructing primary and secondary phage-display libraries. *Methods* **58**, 10–17 (2012).

177. Booth, M. J., Schild, V. R., Graham, A. D., Olof, S. N. & Bayley, H. Light-activated communication in synthetic tissues. *Science Advances* **2**, e1600056 (2016).

178. Ah, L., M, D. & Fs, P. Increased cloning efficiency by temperature-cycle ligation. *Nucleic acids research* **24**, (1996).

179. Karata, K., Vidal, A. E. & Woodgate, R. Construction of a circular single-stranded DNA template containing a defined lesion. *DNA Repair* **8**, 852–856 (2009).

180. Handa, P. & Varshney, U. Rapid and reliable site directed mutagenesis using Kunkel's approach. *Indian J Biochem Biophys* **35**, 63–66 (1998).

181. Tonikian, R., Zhang, Y., Boone, C. & Sidhu, S. S. Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat Protoc* **2**, 1368–1386 (2007).

182. Bebenek, K. & Kunkel, T. A. The use of native T7 DNA polymerase for site-directed mutagenesis. *Nucleic Acids Research* **17**, 5408 (1989).

183. Tabor, S. & Richardson, C. C. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proceedings of the National Academy of Sciences* **84**, 4767–4771 (1987).

184. Green, M. R. & Sambrook, J. Isolation of DNA Fragments from Polyacrylamide Gels by the Crush and Soak Method. *Cold Spring Harb Protoc* **2019**, pdb.prot100479 (2019).

185. Johnston, C., Polard, P. & Claverys, J.-P. The DpnI/DpnII pneumococcal system, defense against foreign attack without compromising genetic exchange. *Mobile Genetic Elements* **3**, e25582 (2013).

186. Hermann, A. & Jeltsch, A. Methylation Sensitivity of Restriction Enzymes Interacting with GATC Sites. *BioTechniques* **34**, 924–930 (2003).

187. Browning, D. F. & Busby, S. J. W. The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**, 57–65 (2004).

188. Hawley, D. K. & McClure, W. R. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res* **11**, 2237–2255 (1983).

189. Ishihama, A. Functional Modulation of Escherichia Coli RNA Polymerase. *Annual Review of Microbiology* **54**, 499–518 (2000).

190. de Boer, H. A., Comstock, L. J. & Vasser, M. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences* **80**, 21–25 (1983).

191. Amann, E., Ochs, B. & Abel, K.-J. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. *Gene* **69**, 301–315 (1988).

192. Shilling, P. J. *et al.* Improved designs for pET expression plasmids increase protein production yield in Escherichia coli. *Commun Biol* **3**, 1–8 (2020).

193. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research* **15**, 8783–8798 (1987).

194. Chapman, K. A., Gunderson, S. I., Anello, M., Wells, R. D. & Burgess, R. R. Bacteriophage T7 late promoters with point mutations: quantitative footprinting and in vivo expression. *Nucleic Acids Research* **16**, 4511–4524 (1988).

195. Martin, C. T. & Coleman, J. E. Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry* **26**, 2690–2696 (1987).

196. Ikeda, R. A., Ligman, C. M. & Warshamana, S. T7 promoter contacts essential for promoter activity in vivo. *Nucleic Acids Research* **20**, 2517–2524 (1992).

197. Rong, M., He, B., McAllister, W. T. & Durbin, R. K. Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences* **95**, 515–519 (1998).

198. Shin, I., Kim, J., Cantor, C. R. & Kang, C. Effects of saturation mutagenesis of the phage SP6 promoter on transcription activity, presented by activity logos. *Proceedings of the National Academy of Sciences* **97**, 3890–3895 (2000).

199. Melton, D. A. *et al.* Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Research* **12**, 7035–7056 (1984).

200. Sagawa, H., Ohshima, A. & Kato, I. A tightly regulated expression system in Escherichia coli with SP6 RNA polymerase. *Gene* **168**, 37–41 (1996).

201. Luo, Y.-R. *Comprehensive Handbook of Chemical Bond Energies*. (CRC Press, 2007). doi:10.1201/9781420007282.

202. Thomas, D. D., Liu, X., Kantrow, S. P. & Lancaster, J. R. The biological lifetime of nitric oxide: Implications for the perivascular dynamics of NO and O2. *Proceedings of the National Academy of Sciences* **98**, 355–360 (2001).

203. Quinlivan, E. P. & Gregory, J. F. DNA digestion to deoxyribonucleoside: A simplified one-step procedure. *Analytical Biochemistry* **373**, 383–385 (2008).

204. Crain, P. F. [42] Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. in *Methods in Enzymology* vol. 193 782–790 (Academic Press, 1990).

205. Il'ichev, Y. V., Schwörer, M. A. & Wirz, J. Photochemical Reaction Mechanisms of 2-Nitrobenzyl Compounds: Methyl Ethers and Caged ATP. *J. Am. Chem. Soc.* **126**, 4581–4595 (2004).

206. Salih, T., Ahlford, A., Nilsson, M., Plichta, Z. & Horák, D. Streptavidin-modified monodispersed magnetic poly(2-hydroxyethyl methacrylate) microspheres as solid support in DNA-based molecular protocols. *Materials Science and Engineering: C* **61**, 362–367 (2016).

207. Caldwell, S. T. *et al.* Protein-mediated dethreading of a biotin-functionalised pseudorotaxane. *Org. Biomol. Chem.* **12**, 511–516 (2013).

208. Smith, J. M., Hartmann, D. & Booth, M. J. Engineering cellular communication between light-activated synthetic cells and bacteria. 2022.07.22.500923 Preprint at https://doi.org/10.1101/2022.07.22.500923 (2022).

209. Qin, W. *et al.* Expansion of CRISPR/Cas9 genome targeting sites in zebrafish by Csy4-based RNA processing. *Cell Res* **25**, 1074–1077 (2015).

210. Singh, S. R. & Murali, A. pH modulates the role of SP6 RNA polymerase in transcription process: an in silico study. *Journal of Biomolecular Structure and Dynamics* **0**, 1–18 (2023).

211. Schenborn, E. T. & Mierendorf, R. C., Jr. A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucleic Acids Research* **13**, 6223–6236 (1985).

212. Ngoc, L. V., Wang, Y.-L., Kassavetis, G. A. & Kadonaga, J. T. The punctilious RNA polymerase II core promoter. *Genes Dev.* **31**, 1289–1301 (2017).

213. Petrenko, N., Jin, Y., Dong, L., Wong, K. H. & Struhl, K. Requirements for RNA polymerase II preinitiation complex formation in vivo. *eLife* **8**, e43654 (2019).

214. Kaufmann, J. & Smale, S. T. Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.* **8**, 821–829 (1994).

215. Deng, W. & Roberts, S. G. E. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev.* **19**, 2418–2423 (2005).

216. Bucher, P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *Journal of Molecular Biology* **212**, 563–578 (1990).

217. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. & Smale, S. T. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol* **14**, 116–127 (1994).

218. Evans, R., Fairley, J. A. & Roberts, S. G. E. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes Dev.* **15**, 2945–2949 (2001).

219. Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. L. & Gao, G. Viral vector platforms within the gene therapy landscape. *Sig Transduct Target Ther* **6**, 1–24 (2021).

220. Raper, S. E. *et al.* Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Molecular Genetics and Metabolism* **80**, 148–158 (2003).

221. Hacein-Bey-Abina, S. *et al.* A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency. *N Engl J Med* **348**, 255–256 (2003).

222. Burnett, J. R. & Hooper, A. J. Alipogene tiparvovec, an adeno-associated virus encoding the Ser(447)X variant of the human lipoprotein lipase gene for the treatment of patients with lipoprotein lipase deficiency. *Curr Opin Mol Ther* **11**, 681–691 (2009).

223. Naddaf, M. Researchers welcome $3.5-million haemophilia gene therapy — but questions remain. *Nature* **612**, 388–389 (2022).

224. Collins, L. T., Ponnazhagan, S. & Curiel, D. T. Synthetic Biology Design as a Paradigm Shift toward Manufacturing Affordable Adeno-Associated Virus Gene Therapies. *ACS Synth. Biol.* **12**, 17–26 (2023).

225. Ludwig, S. L. *et al.* Prevalence of Antibodies to Adenovirus Serotypes 4 and 7 among Unimmunized US Army Trainees: Results of a Retrospective Nationwide Seroprevalence Survey. *The Journal of Infectious Diseases* **178**, 1776–1778 (1998).

226. Xiang, Z. *et al.* Chimpanzee Adenovirus Antibodies in Humans, Sub-Saharan Africa - Volume 12, Number 10—October 2006 - Emerging Infectious Diseases journal - CDC. doi:10.3201/eid1210.060078.

227. Wu, Z., Yang, H. & Colosi, P. Effect of Genome Size on AAV Vector Packaging. *Molecular Therapy* **18**, 80–86 (2010).

228. Mitchell, M. J. *et al.* Engineering precision nanoparticles for drug delivery. *Nat Rev Drug Discov* **20**, 101–124 (2021).

229. Sercombe, L. *et al.* Advances and Challenges of Liposome Assisted Drug Delivery. *Frontiers in Pharmacology* **6**, (2015).

230. Moghimi, S. M. & Szebeni, J. Stealth liposomes and long circulating nanoparticles: critical issues in pharmacokinetics, opsonization and protein-binding properties. *Progress in Lipid Research* **42**, 463–478 (2003).

231. Chrai, S. S., Murari, R. & Ahmad, I. Liposomes (a Review): Part Two, Drug Delivery Systems. *BioPharm International* **15**, (2002).

232. Forssen, E. & Willis, M. Ligand-targeted liposomes. *Advanced Drug Delivery Reviews* **29**, 249–271 (1998).

233. Kristen, A. V. *et al.* Patisiran, an RNAi therapeutic for the treatment of hereditary transthyretin-mediated amyloidosis. *Neurodegenerative Disease Management* **9**, 5–23 (2019).

234. Akinc, A. *et al.* The Onpattro story and the clinical translation of nanomedicines containing nucleic acid-based drugs. *Nat. Nanotechnol.* **14**, 1084–1087 (2019).

235. Hafez, I. M., Maurer, N. & Cullis, P. R. On the mechanism whereby cationic lipids promote intracellular delivery of polynucleic acids. *Gene Ther* **8**, 1188–1196 (2001).

236. Akinc, A. *et al.* Targeted Delivery of RNAi Therapeutics With Endogenous and Exogenous Ligand-Based Mechanisms. *Molecular Therapy* **18**, 1357–1364 (2010).

237. Demongeot, J. & Fougère, C. mRNA COVID-19 Vaccines—Facts and Hypotheses on Fragmentation and Encapsulation. *Vaccines* **11**, 40 (2023).

238. Dams, E. T. M. *et al.* Accelerated Blood Clearance and Altered Biodistribution of Repeated Injections of Sterically Stabilized Liposomes. *J Pharmacol Exp Ther* **292**, 1071–1079 (2000).

239. Ishida, T. *et al.* Accelerated clearance of a second injection of PEGylated liposomes in mice. *International Journal of Pharmaceutics* **255**, 167–174 (2003).

240. Ishida, T. *et al.* Injection of PEGylated liposomes in rats elicits PEG-specific IgM, which is responsible for rapid elimination of a second dose of PEGylated liposomes. *Journal of Controlled Release* **112**, 15–25 (2006).

241. McSweeney, M. D. *et al.* A minimal physiologically based pharmacokinetic model that predicts anti-PEG IgG-mediated clearance of PEGylated drugs in human and mouse. *Journal of Controlled Release* **284**, 171–178 (2018).

242. Wang, X., Ishida, T. & Kiwada, H. Anti-PEG IgM elicited by injection of liposomes is involved in the enhanced blood clearance of a subsequent dose of PEGylated liposomes. *Journal of Controlled Release* **119**, 236–244 (2007).

243. Szebeni, J. Complement activation-related pseudoallergy: A new class of drug-induced acute immune toxicity. *Toxicology* **216**, 106–121 (2005).

244. Dokka, S., Toledo, D., Shi, X., Castranova, V. & Rojanasakul, Y. Oxygen radical-mediated pulmonary toxicity induced by some cationic liposomes. *Pharm Res* **17**, 521–525 (2000).

245. Kedmi, R., Ben-Arie, N. & Peer, D. The systemic toxicity of positively charged lipid nanoparticles and the role of Toll-like receptor 4 in immune activation. *Biomaterials* **31**, 6867–6875 (2010).

246. Sedic, M. *et al.* Safety Evaluation of Lipid Nanoparticle–Formulated Modified mRNA in the Sprague-Dawley Rat and Cynomolgus Monkey. *Vet Pathol* **55**, 341–354 (2018).

247. Hou, X., Zaks, T., Langer, R. & Dong, Y. Lipid nanoparticles for mRNA delivery. *Nat Rev Mater* **6**, 1078–1094 (2021).

248. Blanco, E., Shen, H. & Ferrari, M. Principles of nanoparticle design for overcoming biological barriers to drug delivery. *Nat Biotechnol* **33**, 941–951 (2015).

249. Kulkarni, J. A. *et al.* Fusion-dependent formation of lipid nanoparticles containing macromolecular payloads. *Nanoscale* **11**, 9023–9031 (2019).

250. Leung, A. K. K., Tam, Y. Y. C., Chen, S., Hafez, I. M. & Cullis, P. R. Microfluidic Mixing: A General Method for Encapsulating Macromolecules in Lipid Nanoparticle Systems. *J. Phys. Chem. B* **119**, 8698–8706 (2015).

251. Simões, S. *et al.* Mechanisms of gene transfer mediated by lipoplexes associated with targeting ligands or pH-sensitive peptides. *Gene Ther* **6**, 1798–1807 (1999).

252. Dufès, C., Al Robaian, M. & Somani, S. Transferrin and the transferrin receptor for the targeted delivery of therapeutic agents to the brain and cancer cells. *Therapeutic Delivery* **4**, 629–640 (2013).

253. Daniels, T. R. *et al.* The transferrin receptor and the targeted delivery of therapeutic agents against cancer. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 291–317 (2012).

254. Sato, Y. *et al.* In vivo gene delivery to tumor cells by transferrin-streptavidin-DNA conjugate. *The FASEB Journal* **14**, 2108–2018 (2000).

255. Oishi, T. *et al.* Efficacy of HSV-TK/GCV system suicide gene therapy using SHED expressing modified HSV-TK against lung cancer brain metastases. *Molecular Therapy - Methods & Clinical Development* **26**, 253–265 (2022).

256. Merlot, A. M., Kalinowski, D. S. & Richardson, D. R. Unraveling the mysteries of serum albumin—more than just a serum protein. *Frontiers in Physiology* **5**, (2014).

257. Francis, G. L. Albumin and mammalian cell culture: implications for biotechnology applications. *Cytotechnology* **62**, 1–16 (2010).

258. Conner, S. D. & Schmid, S. L. Regulated portals of entry into the cell. *Nature* **422**, 37–44 (2003).

259. Hassanin, I. & Elzoghby, A. Albumin-based nanoparticles: a promising strategy to overcome cancer drug resistance. *Cancer Drug Resistance* **3**, 930–946 (2020).

260. Malhotra, A. & Mittal, B. R. SiRNA gene therapy using albumin as a carrier. *Pharmacogenetics and Genomics* **24**, 582 (2014).

261. Gelamo, E. L. & Tabak, M. Spectroscopic studies on the interaction of bovine (BSA) and human (HSA) serum albumins with ionic surfactants. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **56**, 2255–2271 (2000).

262. Guan, G. *et al.* An Effective Cationic Human Serum Albumin-Based Gene-Delivery Carrier Containing the Nuclear Localization Signal. *Pharmaceutics* **11**, 608 (2019).

263. Sineshchekov, O. A., Jung, K.-H. & Spudich, J. L. Two rhodopsins mediate phototaxis to low- and high-intensity light in Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences* **99**, 8689–8694 (2002).

264. Rogers, S. G. & Weiss, B. [26] Exonuclease III of Escherichia coli K-12, an AP endonuclease. in *Methods in Enzymology* vol. 65 201–211 (Academic Press, 1980).

265. Tang, X., Maegawa, S., Weinberg, E. S. & Dmochowski, I. J. Regulating Gene Expression in Zebrafish Embryos Using Light-Activated, Negatively Charged Peptide Nucleic Acids. *J. Am. Chem. Soc.* **129**, 11000–11001 (2007).

266. Kay, B. K., Thai, S. & Volgina, V. V. High-Throughput Biotinylation of Proteins. in *High Throughput Protein Expression and Purification: Methods and Protocols* (ed. Doyle, S. A.) 185–198 (Humana Press, 2009). doi:10.1007/978-1-59745-196-3_13.

267. Fairhead, M., Krndija, D., Lowe, E. D. & Howarth, M. Plug-and-Play Pairing via Defined Divalent Streptavidins. *Journal of Molecular Biology* **426**, 199–214 (2014).

268. Zhang, K., Zhu, X., Jia, F., Auyeung, E. & Mirkin, C. A. Temperature-Activated Nucleic Acid Nanostructures. *J. Am. Chem. Soc.* **135**, 14102–14105 (2013).

269. Hartmann, D., Chowdhry, R., Smith, J. M. & Booth, M. J. Orthogonal Light-Activated DNA for Patterned Biocomputing within Synthetic Cells. *J. Am. Chem. Soc.* **145**, 9471–9480 (2023).

Appendix

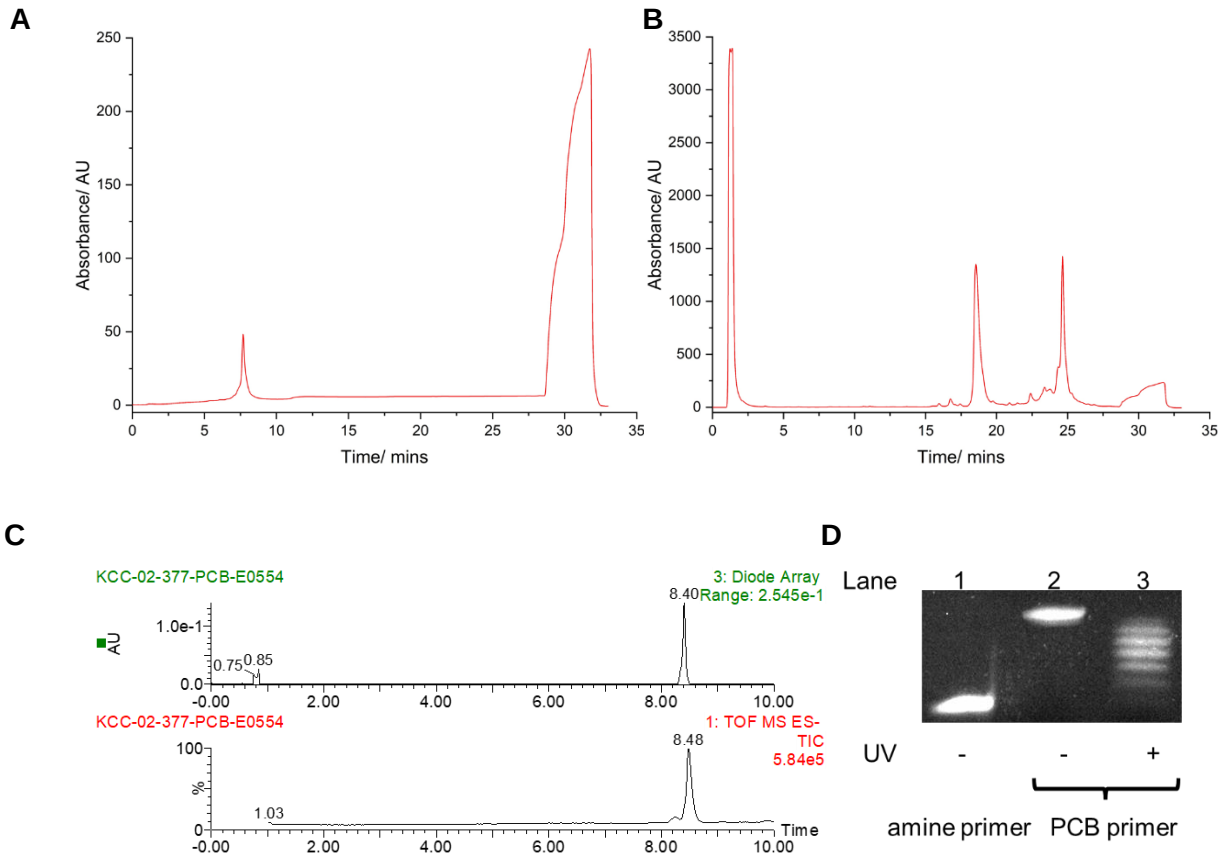


Figure A1. Purification and validation of T7 7-PCBs oligo

A. HPLC chromatogram of the amine primer starting material. B. HPLC chromatogram of the PCB primer product. C. LC and UV monitoring chromatograms of the PCB oligo gave the major product with m = 15398, corresponding to the 7-PCBs oligo. The adjacent side peak had m = 14672 which was the 6-PCBs oligo that co-eluted during HPLC purification. D. Denaturing PAGE of PCB oligo and its uncaging showing multiple bands.

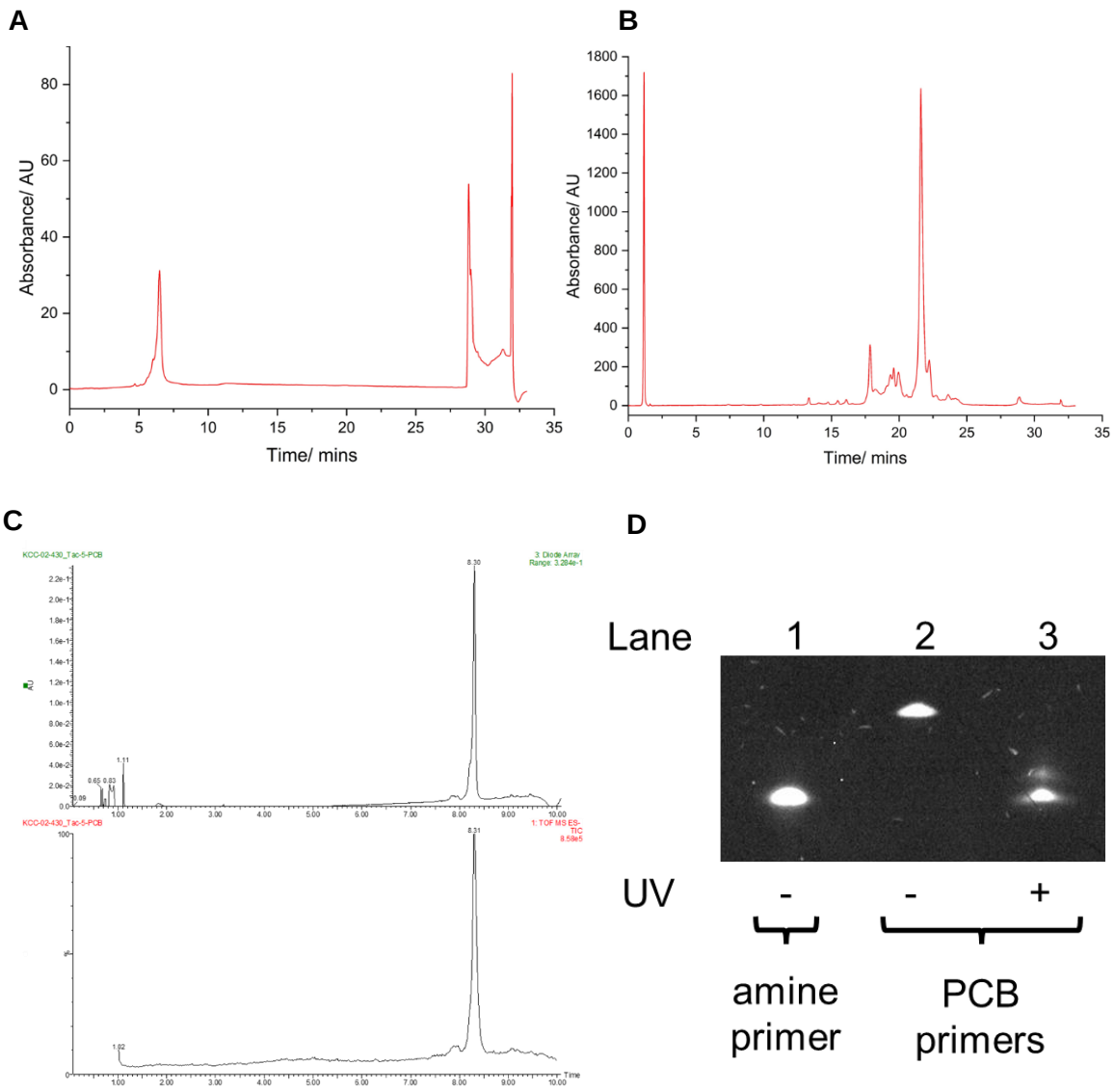


Figure A2. Purification and validation of tac 5-PCBs oligo

A. HPLC chromatogram of the amine primer starting material. B. HPLC chromatogram of the PCB primer product. C. LC and UV monitoring chromatograms of the PCB oligo gave the major product with m = 14846, corresponding to the 5-PCBs oligo. D. Denaturing PAGE of PCB oligo and its uncaging.

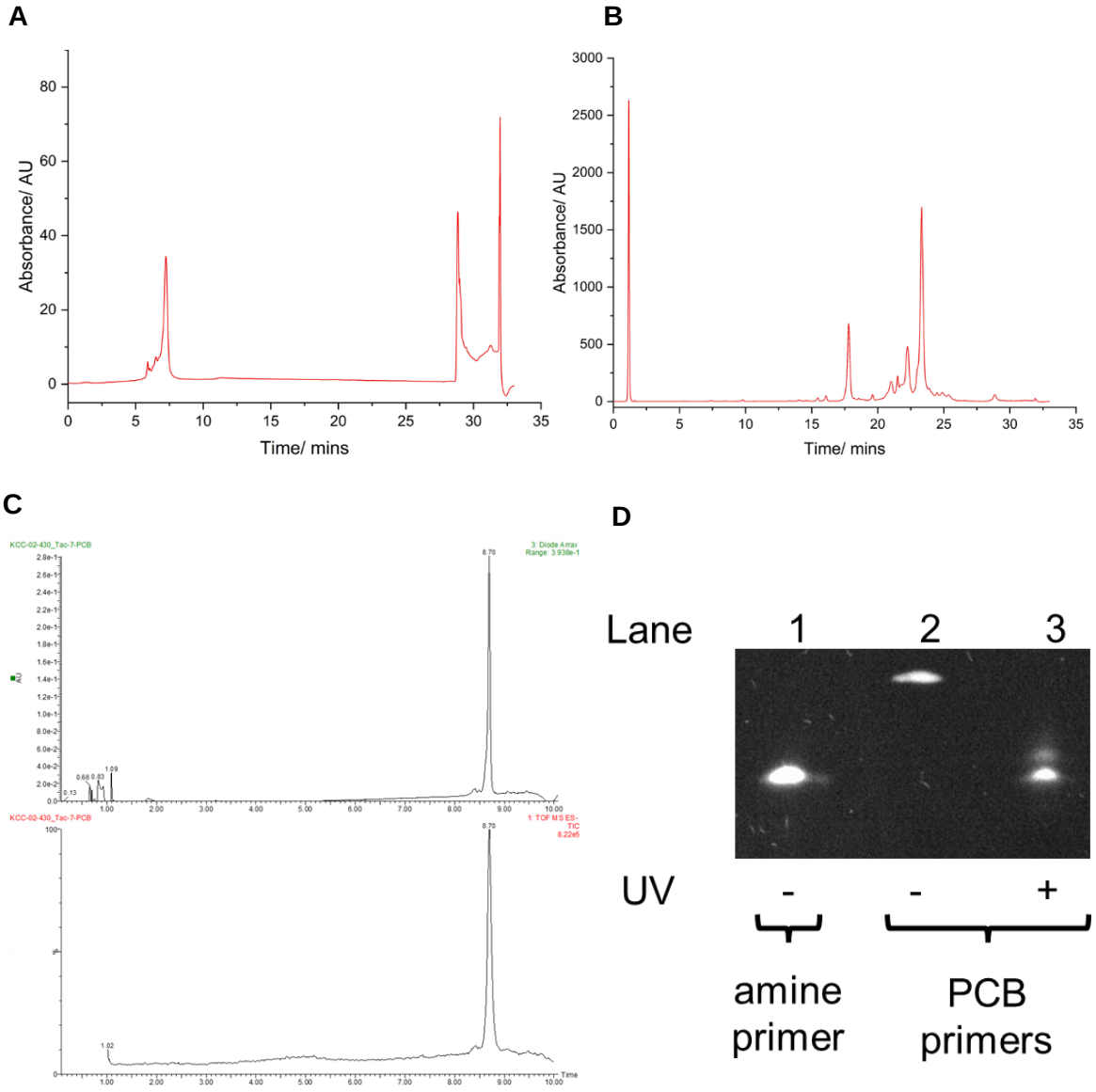


Figure A3. Purification and validation of tac 7-PCBs oligo

A. HPLC chromatogram of the amine primer starting material. B. HPLC chromatogram of the PCB primer product. C. LC and UV monitoring chromatograms of the PCB oligo gave the major product with m = 16605, corresponding to the 7-PCBs oligo. D. Denaturing PAGE of PCB oligo and its uncaging.

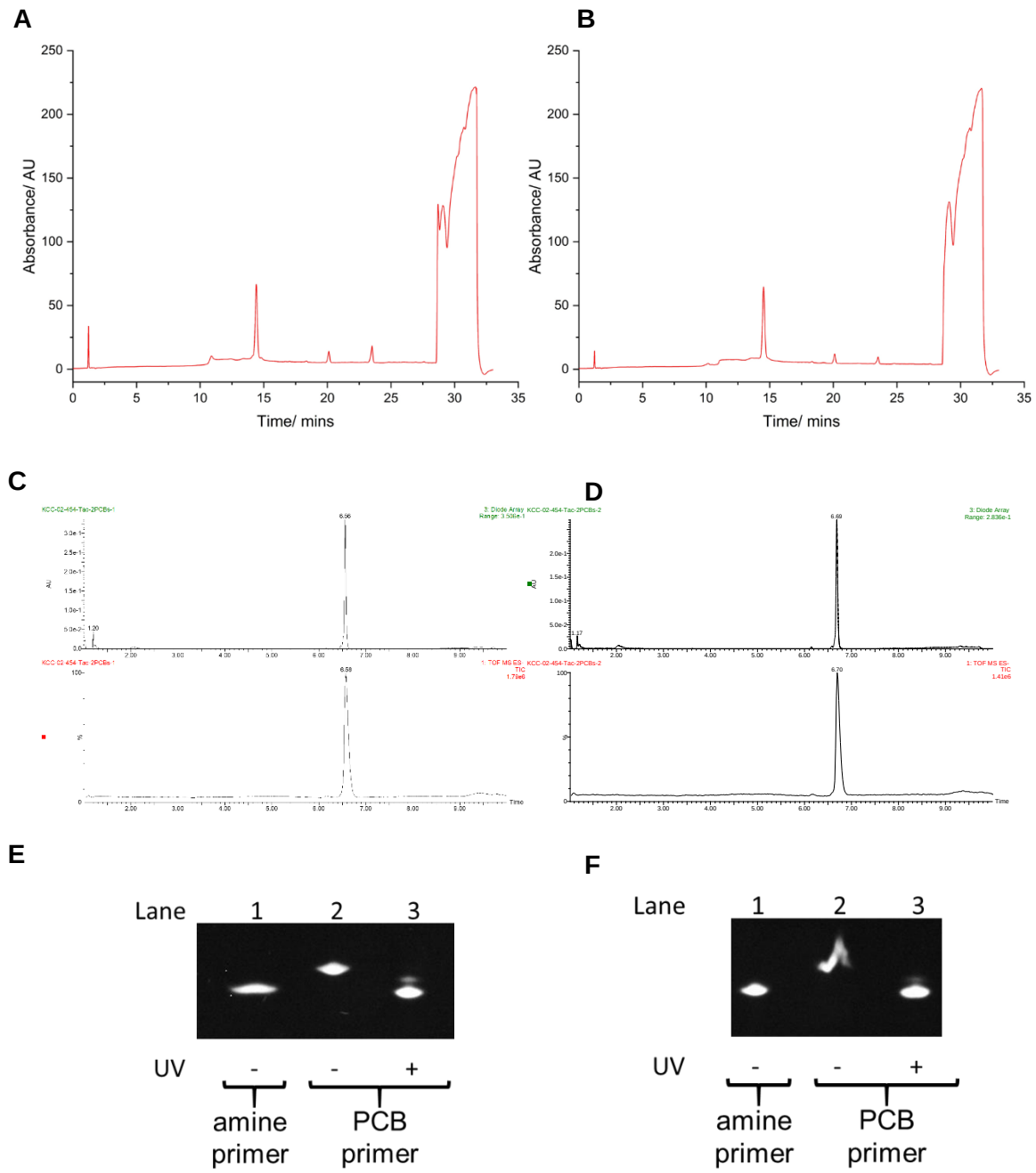


Figure A4. Purification and validation of tac 2-PCBs oligos

A-B. HPLC chromatograms of the PCB primer products. C-D. LC and UV monitoring chromatograms of the PCB oligos gave the major product with m = 15891, corresponding to the 2-PCBs oligos. E-F. Denaturing PAGE of PCB oligos and their uncaging. A/C/E corresponded to Tac-2amines1; B/D/F correspond to Tac-2amines2.

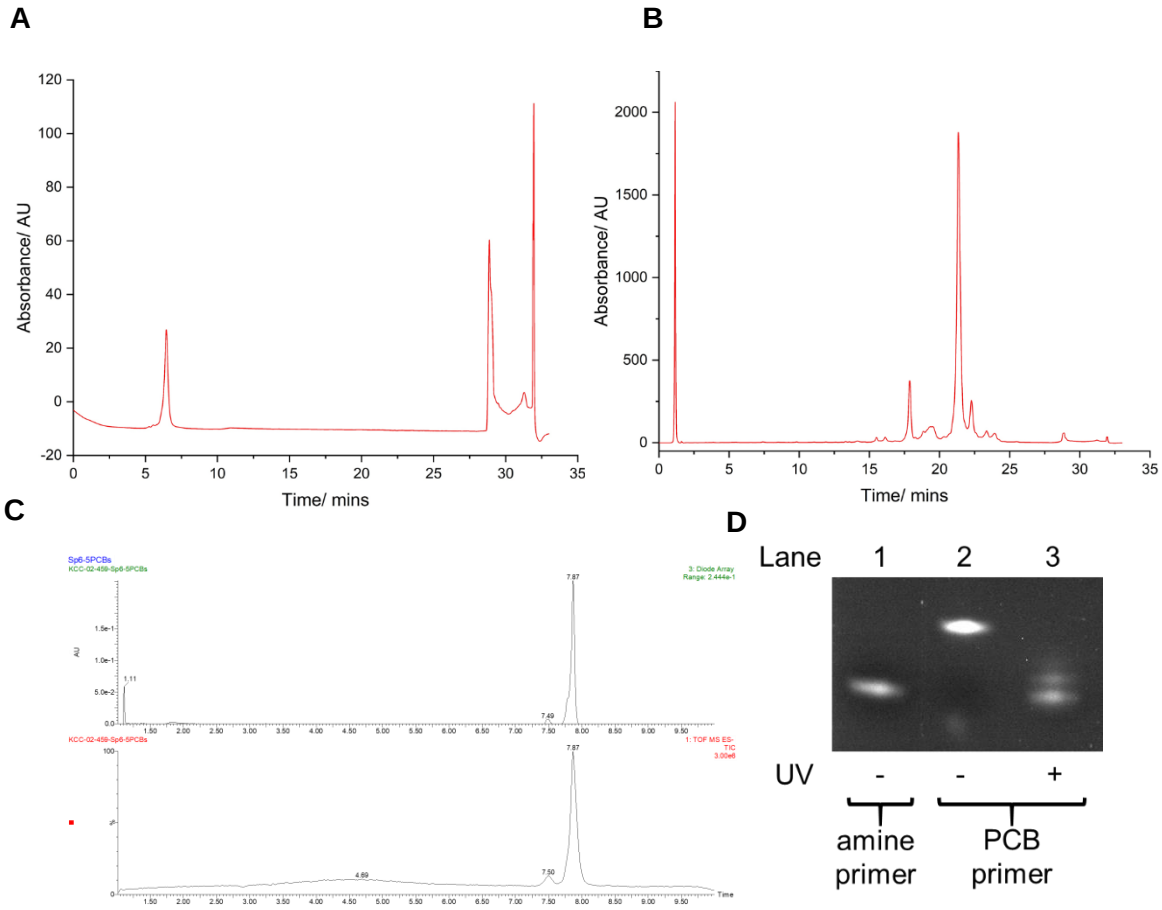


Figure A5. Purification and validation of Sp6 5-PCBs oligo

A. HPLC chromatogram of the amine primer starting material. B. HPLC chromatogram of the PCB primer product. C. LC and UV monitoring chromatograms of the PCB oligo gave the major product with m = 13618, corresponding to the 5-PCBs oligo. The minor adjacent peak had m = 12893, corresponding to the 4 PCBs oligo (5% relative abundance) D. Denaturing PAGE of PCB oligo and its uncaging.

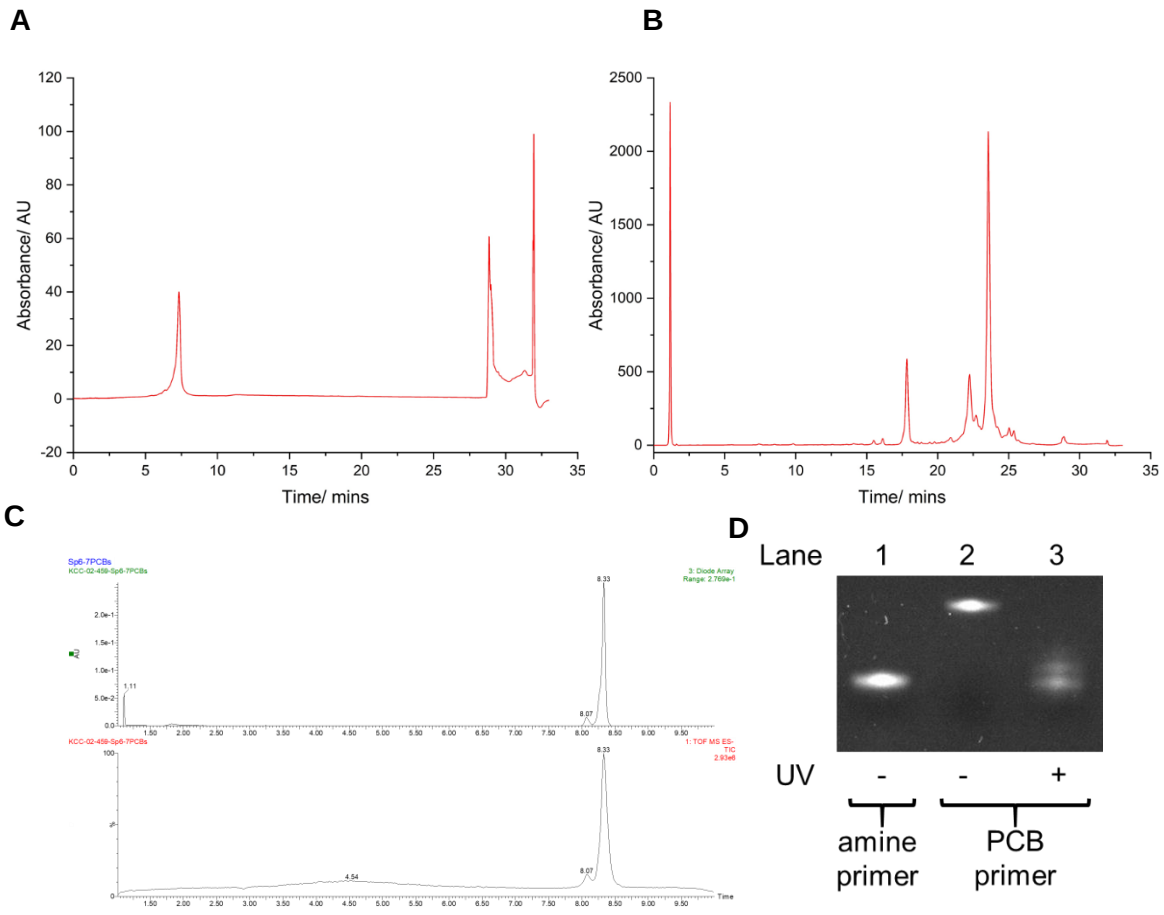


Figure A6. Purification and validation of Sp6 7-PCBs oligo

A. HPLC chromatogram of the amine primer starting material. B. HPLC chromatogram of the PCB primer product. C. LC and UV monitoring chromatograms of the PCB oligo gave the major product with m = 15378, corresponding to the 7-PCBs oligo. The minor adjacent peak had m = 14653, corresponding to the 6 PCBs oligo (6% relative abundance) D. Denaturing PAGE of PCB oligo and its uncaging.

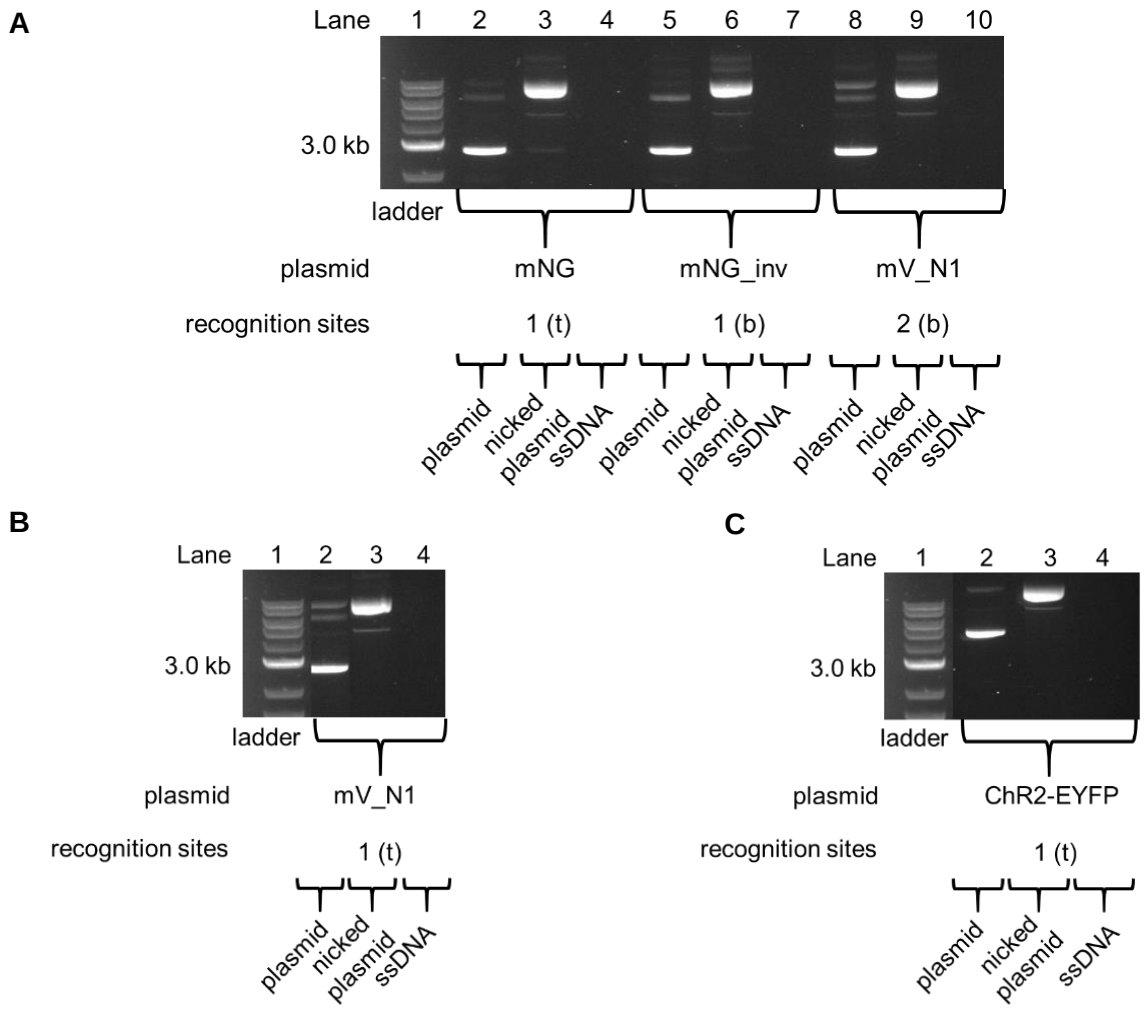


Figure A7. Investigating nickase promiscuity when the recognition sites were on the same strand

Plasmids either contained 1 or 2 recognition sites for the nickase being tested. In the latter case, both sites were on the same strand. Hence, after the exonuclease reaction, across all cases, there should be a band for ssDNA. A. Nt.BspQI B. Nb.BsrDI C. Nb.BsmI. (t) and (b) corresponded to the sense and antisense strands, respectively.

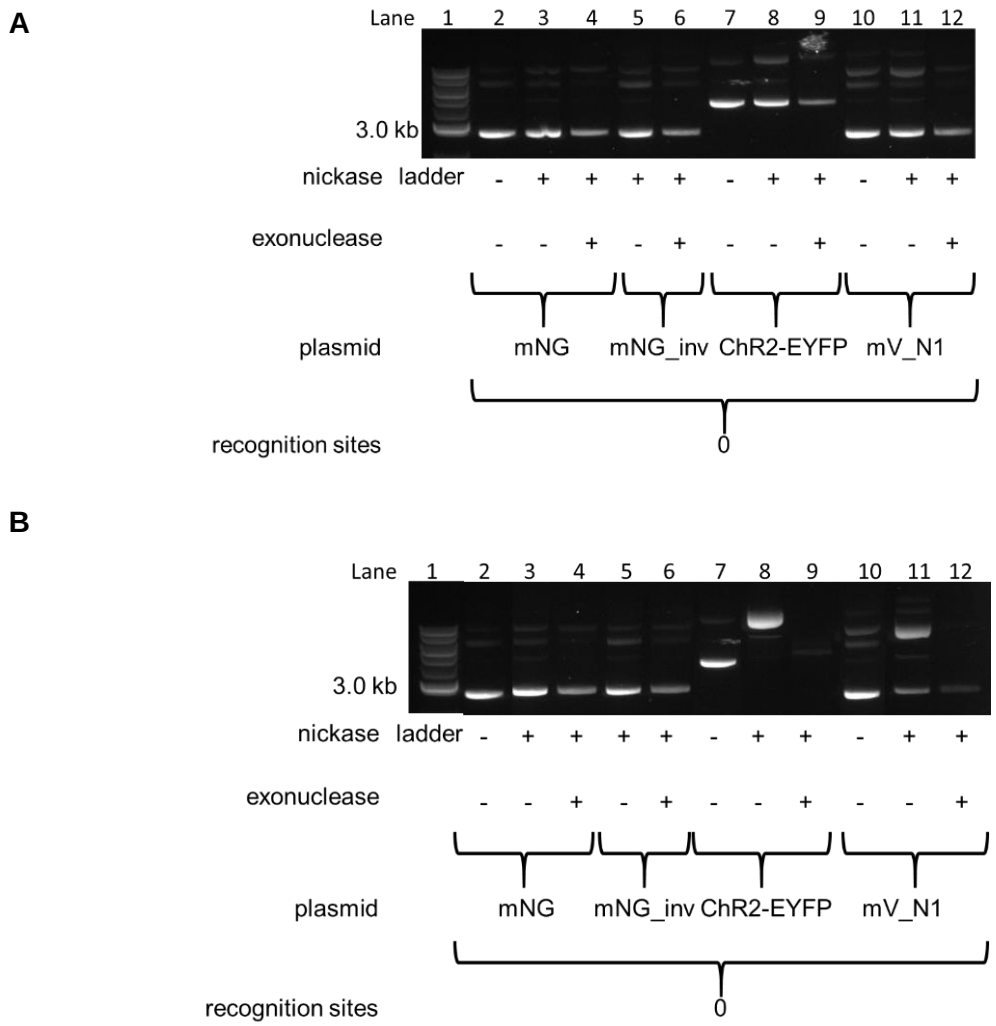


Figure A8. Investigating nickase promiscuity when there were no recognition sites

Plasmids did not contain a recognition site for the nickase being tested. Hence, the plasmids should not be nicked and bands should appear at the same position as the native supercoiled plasmid across all conditions. A. Nb.BbvCI B. Nt.BbvCI.

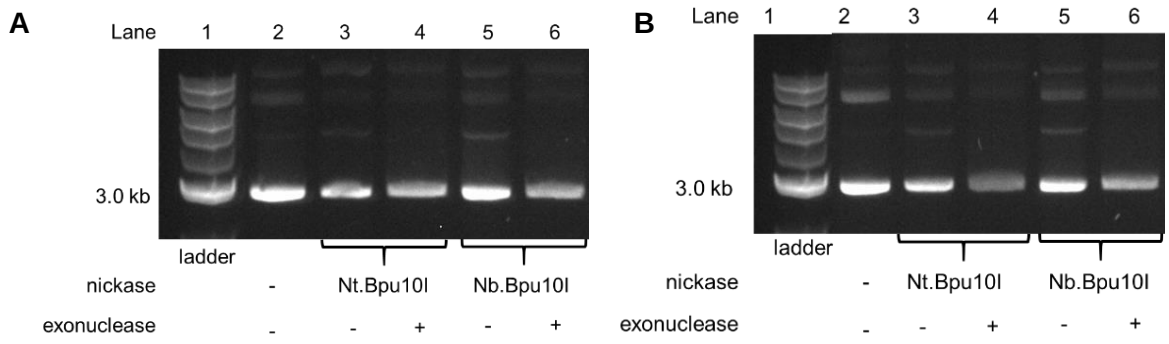


Figure A9. Testing Nt- and Nb.Bpu10I specificity

A. mNG and B. mNG_inv were not nicked by either Nt.- or Nb.Bpu10I as these plasmids did not contain the BbvCI/Bpu10I recognition sites.

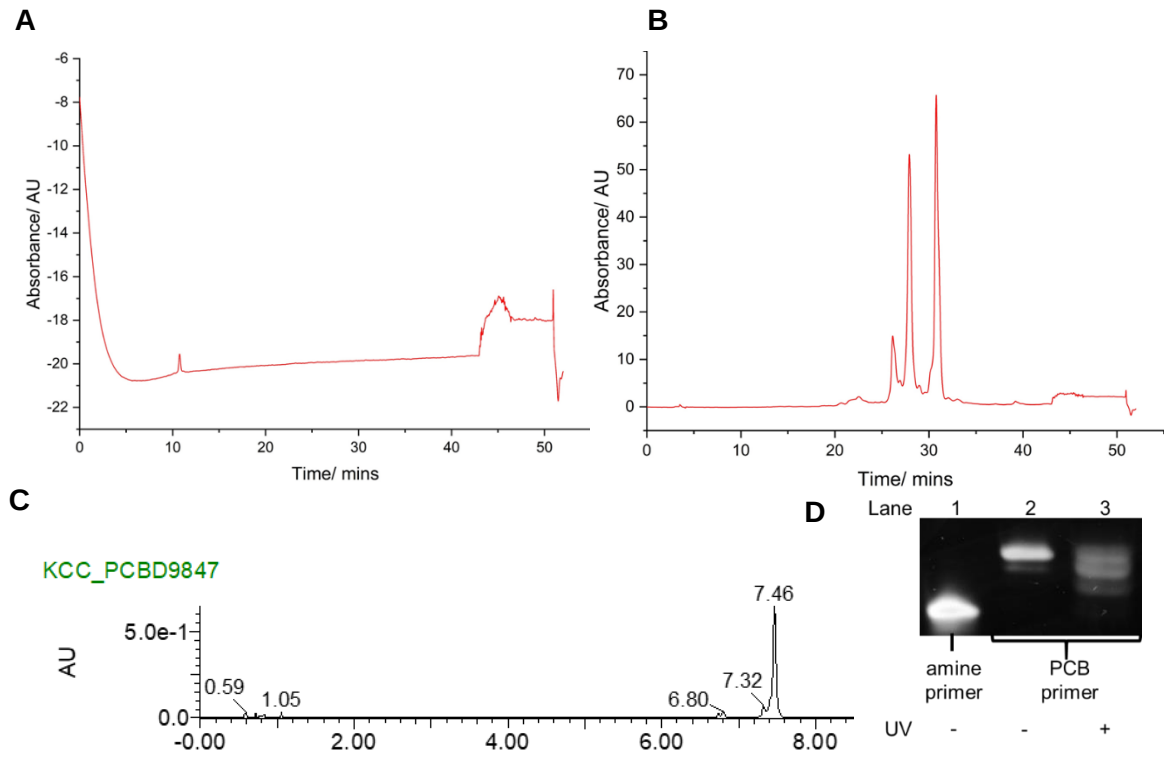


Figure A10. Purification and validation of CMV 3-PCBs oligo

A. HPLC chromatogram of the amine primer starting material. B. HPLC chromatogram of the PCB primer product. C. UV monitoring chromatograms of the PCB oligo gave the major product with m = 9728, corresponding to the 3-PCBs oligo. The minor adjacent peak had m = 9002, corresponding to the 2 PCBs oligo (8% relative abundance) D. Denaturing PAGE of PCB oligo and its uncaging.

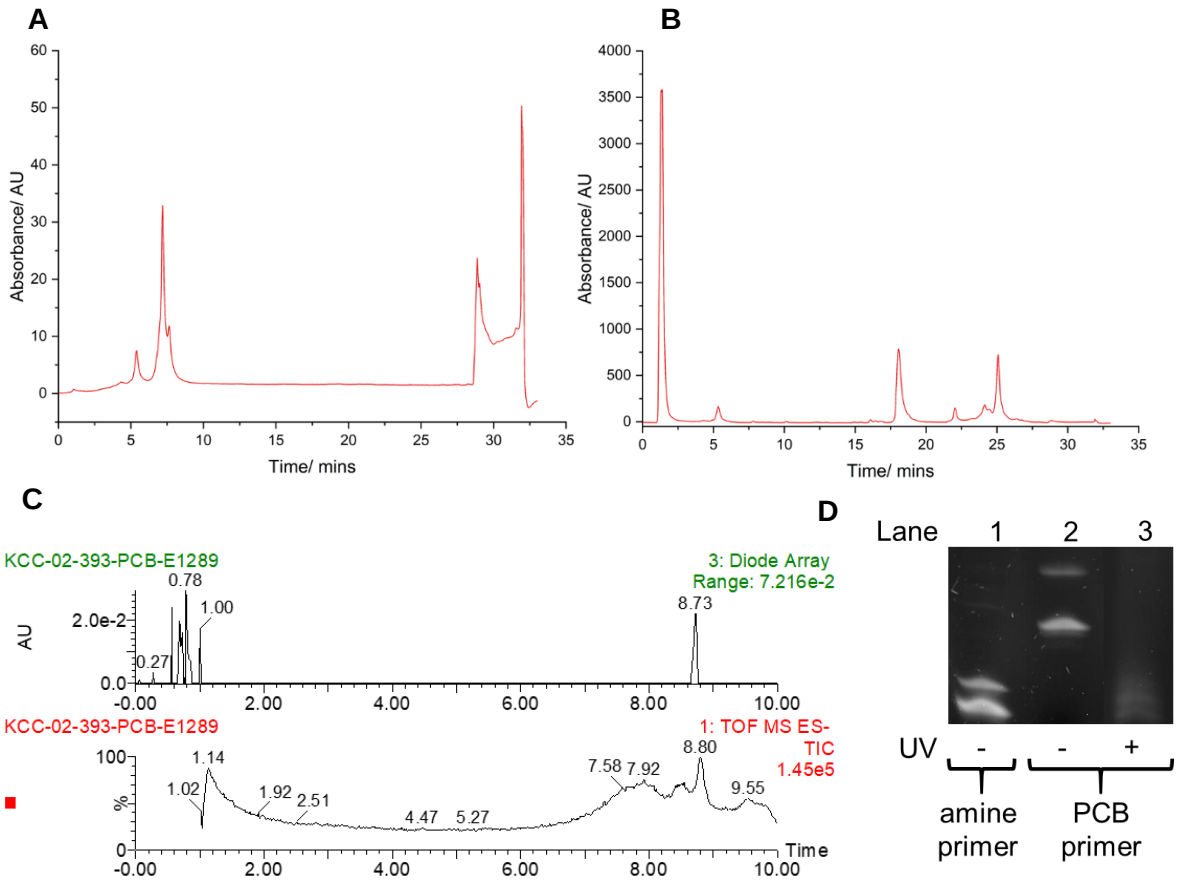


Figure A11. Purification and validation of CMV 8-PCBs oligo

A. HPLC chromatogram of the amine primer starting material. B. HPLC chromatogram of the PCB primer product. C. LC and UV monitoring chromatograms of the PCBs oligo. Deconvolution of the major signal at 8.80 mins did not give mass for the PCB product D. Denaturing PAGE of PCB oligo showed 2 bands for the starting material and product.

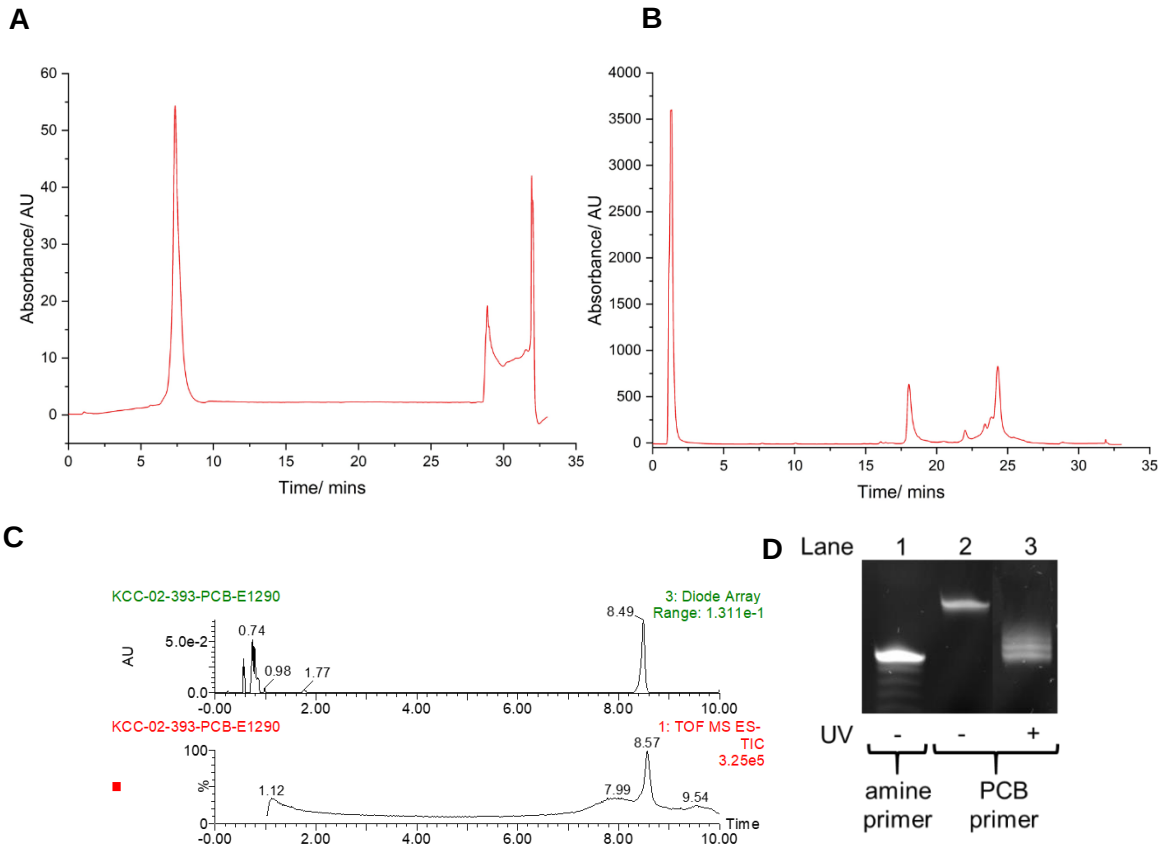


Figure A12. Purification and validation of CMV 9-PCBs oligo

A. HPLC chromatogram of the amine primer starting material. B. HPLC chromatogram of the PCB primer product. C. LC and UV monitoring chromatograms of the PCB oligo gave the major product with m = 22076, corresponding to the 9-PCBs oligo. D. Denaturing PAGE of PCB oligo and its uncaging.

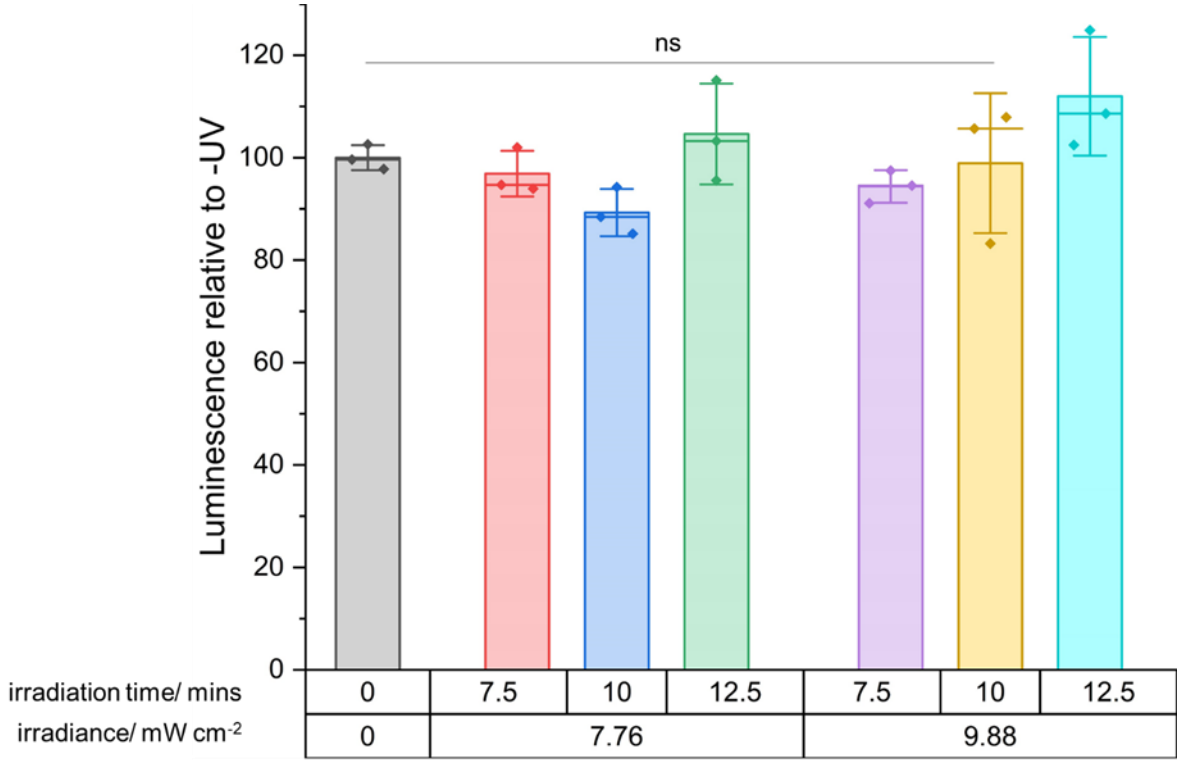


Figure A13. Cell viability assay with optimised UV settings for transfection

Cell viability assay showed the UV conditions employed, 9.88 mW cm^{-2} for 10 mins, had little damage to cells. Error bars showed standard deviations of technical triplicates (*n* = 3). Two-tailed, unpaired Student's t-test was performed. ns = non-significant (P > 0.05).

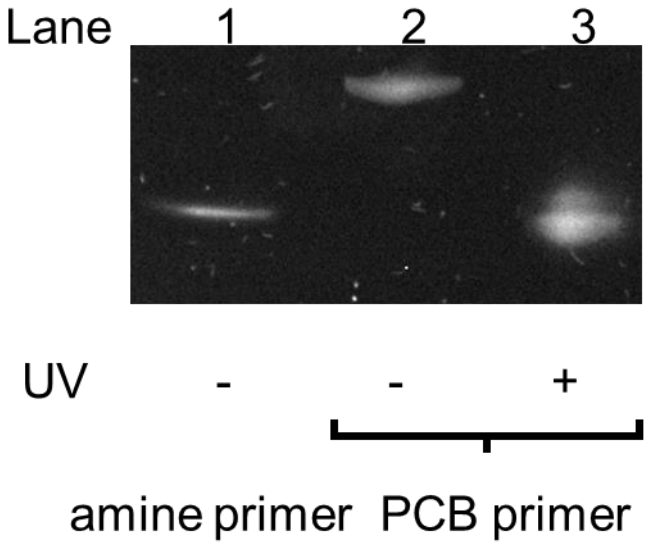


Figure A14. Cleavage of the CMV 9-PCBs oligo with optimized UV settings

CMV 9-PCBs oligo was irradiated with 9.88 mW cm^{-2} for 10 mins and analysed on denaturing PAGE.

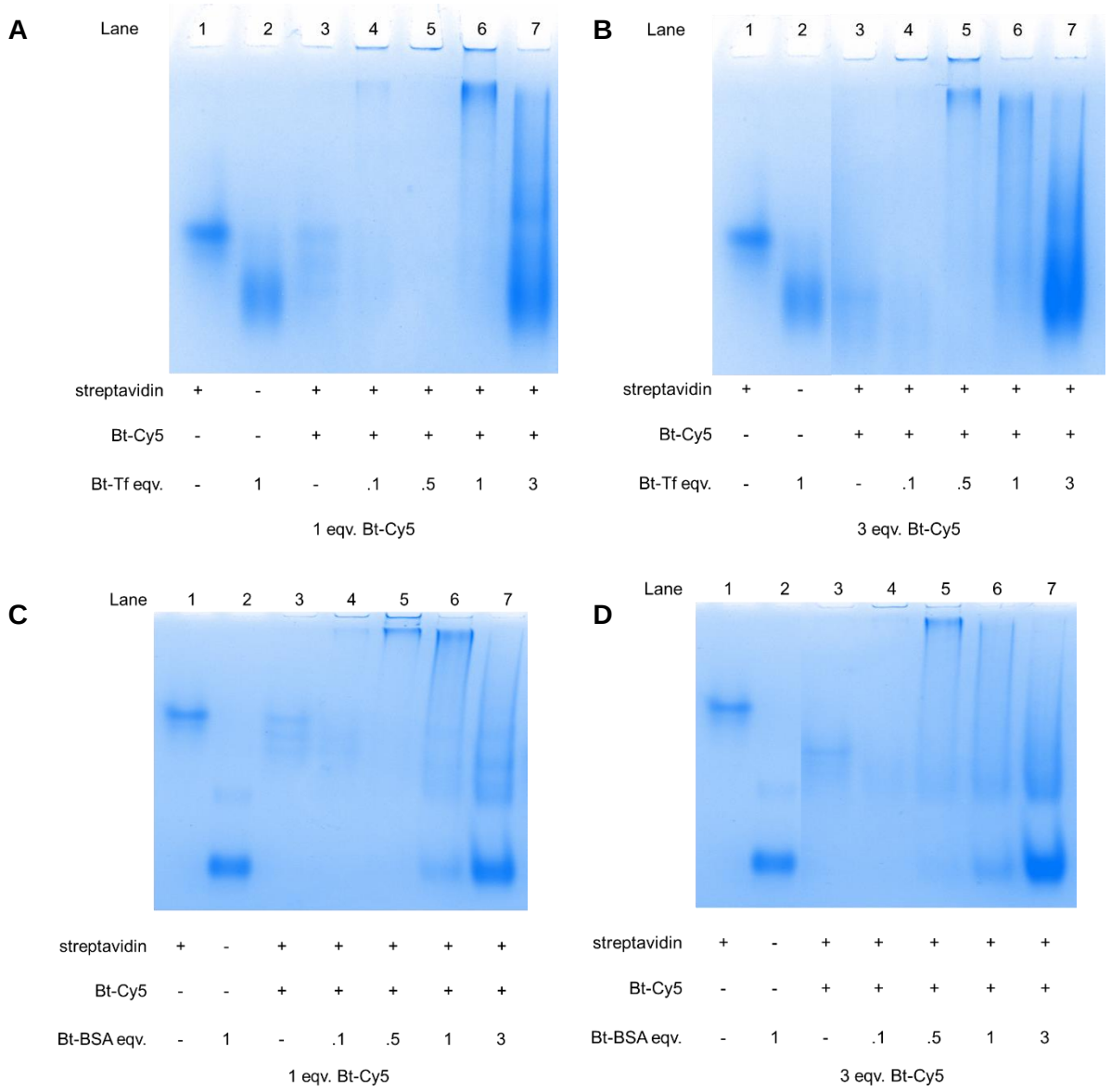


Figure A15. Lowering Bt-Tf/BSA and increasing Bt-Cy5 equivalents

WTS was bound to either 1 or 3 equivalents of Bt-Cy5 at high dilution. Subsequently, different equivalents of either Bt-Tf and Bt-BSA were directly added to the WTS-Cy5 solution at high volume. A-B. Transferrin constructs at (A) 1 eqv. or (B). 3 eqv. Bt-Cy5. C-D. BSA constructs at (C) 1 eqv. or (D) 3 eqv. Bt-Cy5.